

kaechele@sip.medizin.uni-ulm.de

letters from the SSCP net

Society for Scientific
Clinical Psychology



*in memoriam A E Meyer
dem unentwegt
neugierigen
e-mail kundigen
www fan*

RCT Studies

pro and con

letters from the sscpnet 1/95

From: DonaldK737@aol.com To: psycho-pharm@netcom.com July 7, 1995 Subject: Re: "Neuro-behavioral Disorder"

martin seligman says 1/7/95:

> "A more telling example could not have been produced by the other side. I think Ivan's facts are simply inaccurate here and they illustrate the problem: There is a large outcome literature comparing the antidepressants and the anti-anxiety agents to the Clark-Barlow-Beck cognitive therapy of panic. The literature shows that the drugs relieve panic someplace in the 50-60% range depending on how you quantify relief. Recurrence is close to 100% when the drugs stop. CT produces 85-90% relief in fewer than 12 sessions with virtually zero recurrence. To call panic a "neuro" anything is to be back in about 1986."<

It is being back in 1950 to cite inadequately controlled trials as if they were definitive unfortunately it is possible to pick samples of patients that meet that investigator's criteria for panic disorder that have a remarkable propensity for panic remission . In Marks' recent article in Brit J PPsychiat. 100% !! of the patients on placebo became panic free. Therefore any comparison of medication and psychotherapy requires a pill placebo arm to demonstrate that this is a medication responsive sample and also to evaluate placebo effects. There is only one- count it one- such published study. Black et al Arch Gen Psychiat. in it fluvoxamine was markedly superior to only fair placebo response(about 30%) and cognitive therapy was barely distinguishable from placebo.(my colleagues,gorman &papp are currently collaborating with barlow, shear, woods, to properly replicate this.) further there are studies (Shear, Ost) where non cognitive therapies do as well as CT. So panic studies are repeating the sad history of depression studies where similar poorly controlled studies touted the advantages of CT.

To repeat: any comparison of medication and psychotherapy requires a pill placebo arm to demonstrate that this is a medication responsive sample and also to evaluate placebo effects. There is only one- count it one- such published study in the depression area -the Elkin et al NIMH study. Again cognitive therapy is only marginally different from placebo and hardly differs from interpersonal therapy and much inferior to medication in the more depressed. As further evidence for the nonspecific nature of psychotherapy CT does differ from IPT as follows in depression: if you have severe dysfunctional attitudes you do better on IPT not CT. conversely if you have severe interpersonal difficulties CT is better than interpersonal rx¹.(IPT) This makes perfect sense from Jerome Frank's antidemoralization theory of psychotherapy, which should be a success experience. Boring in precisely where the patient is most impaired is pro-demoralization. Further the implication is that psychorx is not even treating the primary illness but a secondary complication (demoralization) .

¹rx = therapy

To any one interested I would be glad to send references and reprints.
Donald F. Klein M.D.

From seligman@cattell.psych.upenn.edu Sun Jan 8 04:28:01 1995 To: sscpnet@bailey.psych.nwu.edu
--

Something significant just happened on the psych-pharm@netcom.com net I had responded to Ivan Goldberg's assertions that panic was a biological disorder with the summary of the Clark-Beck-Barlow work below. To my surprise, who should appear to challenge my statements but Don Klein himself. He asserts below that

- 1) These panic studies are poorly controlled
- 2) That CT barely works better than placebo for depression and for panic
- 3) That drugs work for severe depression much better than CT or IPT
- 4) That IPT is specific to social problems and CT to cognitive patients.
- 5) And generally, between the lines, the psychotherapy for real DISORDERS is a crock.

Now I view Klein's speaking so publically as a positive development. You probably know the important role he has played over the years, often behind the scenes, in getting psychiatry and the general public and NIMH to view psychological problems as disorders and disorders as biological in nature. I actually respect his contribution enormously, and I attribute the "decade of the brain" etc in part to his charisma and unceasing efforts to biologize psychological problems. But I think each of the assertions above is flawed, varying from overstated to false. Since I am more a reader and summarizer of the literatures referred to in 1-5 above, I want to defer my response to many of you on SSCPNET who are the primary contributors. Would you be good enough to take on--chapter and verse--any of the five assertions above?

I will then get them posted under your name, or perhaps organize them into one long response, on the psycho-pharm net. Or you may want to post directly there (but you have to be a member). At any rate, I think the doors are now open to a public debate over these core disputes between psychology and biological psychiatry. The truth (someplace in between no doubt) may even emerge.

From seligman@cattell.psych.upenn.edu Sun Jan 8 19:38:52 1995 To: psycho-pharm@netcom.com

Here are comments concerning the importance or nonimportance of a pill placebo control in panic, and on the notion of selecting for drug responsive patients:

Richard McNally speaks:

Forwarded message: Subject: comments on Don Klein's posting

Dear Dr. Seligman:

I am responding to your invitation to comment on Don Klein's recent posting.

1) Don is surely correct in stating that it would have been desirable for David Clark to have included a pill placebo group in his panic treatment study; this would have enabled David to assess for placebo response. Don also noted that it would have enabled David to determine whether his patients were an imipramine-responsive sample. That is, had imipramine failed to produce a significantly better response than pill placebo, one might then question the representativeness of the patients. The notion of "imipramine-responsiveness" is trickier than it first appears, and it introduces potential problems of circular reasoning. For example, let's say an investigator finds that cognitive treatment produces an 80% panic-remission rate, imipramine produces a 50% panic-remission rate, and pill placebo produces a 30% panic-remission rate.

Let's also assume that the difference between cognitive therapy and the other two conditions are significant, whereas imipramine does not differ significantly from placebo. What do we conclude? One might explain the (relatively) poor showing by invoking the concept of imipramine-nonresponsiveness. Although this is certainly a possible explanation, it would be desirable to have some independent means of establishing this other than the results one is trying to explain. [There are interesting possibilities. For example, I would not be surprised that folks who panic in response to yohimbine are more imipramine-responsive than those who do not.]

But let's suppose our investigator finds that cognitive therapy produces significantly more panic remission (say, 90%) than does imipramine (say, 70%) which, in turn, does better than placebo (say 30%). What do we now conclude? Do we say that cognitive therapy is more effective than imipramine which is more effective than placebo? That's one interpretation.

Another is to assert that we drew our patients from a RELATIVELY imipramine-nonresponsive population, and infer this from the failure of imipramine to do at least as well as cognitive therapy. Indeed, the proportions of imipramine-responsive patients in any give sample will likely vary; it won't be an all-or-nothing thing. The upshot of all this is that we can apply the concept of (relative) imipramine-

nonresponsiveness any time that we discover that imipramine performs less well than cognitive therapy.

Therefore, even if David had included a pill placebo and discovered that imipramine outperformed it, one could still complain that his patients were relative relatively imipramine-responsive, and infer this (in circular fashion?) from their failure to do as well as the cognitive therapy patients. This argument about imipramine-responsiveness, of course, applies equally to behavior therapy, cognitive therapy, or whatnot. For example, Black et al. reported a less effective cognitive therapy result than David Clark. What do we conclude? Did Black draw his patients from a (relatively) cognitive-therapy nonresponsive population? [Black did not include a psychosocial attention placebo group.] Do we infer this, ipso facto, because the University of Iowa patients did not do as well as the University of Oxford ones?

2) Don makes the important observation of high placebo remission rates in certain studies (e.g., Marks's), and suggests that Isaac might have had an unusual population. While that is certainly possible, another reason why might have to do with the window for assessing panic-free status. It is common for behavior therapist and psychopharmacologists alike to select, say, a two-week window posttreatment to track panic attacks. If a person has no attacks during these 14 days, the researcher classifies the patient as panic-free. Given that panic attacks do not occur on a regular basis (e.g., someone might have a volley of attacks during a several day period, yet be panic-free for a month or more before the next attack), it is possible that some fairly ill people might appear panic-free.

3) Don also makes the important point regarding other approaches to treating panic. As a behavior therapist, I was surprised as the next guy at Kathy Shear's impressive results. She plans to replicate these intriguing findings, but at this point I think that her results suggest that there may very well be many "roads to Rome" in the treatment of panic disorder.

Rich McNally

From njacob@u.washington.edu
Sun Jan 8 18:19:40 1995
To: "Martin E. P. Seligman" <seligman@cattell.psych.upenn.edu>

You may recall my remarks of a few weeks ago regarding TDCRP, which was an attempt to set the record straight on the misinterpretations of how CT performed in the TDCRP study. They spoke directly to the erroneous "depression assertions" made by Don Klein, and you could forward those directly to this network. In case you did not save those messages, I will forward them to you again.

From: Neil Jacobson <njacob@u.washington.edu>
Sun Jan 8 19:42:48 1995
To: section3 <:sscpnet@bailey.psych.nwu.edu>
Subject: science, politics, and the Treatment of Depression Collaborative Research Program

I have always been alarmed by science, when it is conducted by rumor. Yet, the unpublished results from the Treatment of Depression Collaborative Research Program (TDCRP), as well as misrepresentations and misinterpretations of the findings, have led to some fairly disastrous consequences for scuttlebut regarding the treatment of depression.

The immediate impetus for my remarks comes from an article published by Michael Thase in "The Behavior Therapist". It is a good article, where Thase seeks to salvage cognitive behavioral treatments for depression in light of the TDCRP findings. But the article should never have needed to be written.

The TDCRP found that there were no differences between pharmacotherapy, CBT, and Interpersonal Psychotherapy (IPT). Based on a priori hypotheses, there is no evidence from this study that CBT was inferior to either IPT or Imipramine plus clinical management. As a post hoc, hypothesis generating tactic, the authors (Irene Elkin and colleagues, in a 1989 article published in the Archives of General Psychiatry), did a median split on severity). I repeat: there were no PREDICTED interactions between severity and outcome. Based on these post hoc analyses, there were still no differences between the two psychotherapies, or between the drug and psychotherapy conditions, for either mild or severe depressives. However, only Imipramine and IPT outperformed placebo for this post hoc derived "severe" subgroup. Differences between CBT and placebo, while favoring CBT, were not statistically significant.

On this basis, people in the field have suggested that CBT may be inappropriate for severe depression, and this suggestion actually made it into

a recently published set of treatment guidelines for major depression prepared by the American Psychiatric Association.

There are so many problems with this level of discourse that I hardly know where to start. First of all, even after the post hoc median split on severity measures, there were no statistically significant differences between CBT and either IPT or drugs. Therefore, there is no basis for the inference that CBT is less effective than either IPT or drugs with severe depressives.

Second, even if there had been significant interactions between severity and outcome, these interactions resulted from a fishing expedition. In other words, they were not based on a priori hypotheses, and therefore have to be viewed as hypotheses generating, exploratory analyses, which could only be tested on an independent sample.

I learned in my very first statistics course that you cannot generate and confirm a hypothesis from the same set of data. Many seem to have forgotten this lesson in elementary statistics.

Third, this is just one study, albeit an expensive one. Even if the findings had been predicted, and confirmed, they would await replication for confirmation. As it stands, there is no a priori finding to replicate, and the post hoc findings were null, which means there is hardly even a suggestive hypothesis generated by the post hoc analyses.

Finally, to the extent that there are mysteries in this provocative data set, the scientific community, and perhaps clients themselves, have suffered from the slow, plodding rate of publication. Taxpayers have been cheated by the unwarranted inferences drawn from the study, and clients might be the ultimate victims. The lack of timely publication has resulted in what I have come to call "science by rumor". Thase's article summarizes the gossip that substitutes for scientific inference. There is a mysterious site by treatment interaction which is rumored to shed light on the post hoc findings,

From seligman@cattell.psych.upenn.edu Mon Jan 9 01:00:45 1995 To: rjm@burrhus.harvard.edu Subject: Re: comments on Don Klein's posting

Rich McNally points out that for a sample imipramine responsivity may vary in degree. That is true but I don't think it is relevant given our current level of understanding. For instance the Klosko/Barlow trial showed CT much better than placebo but in that panic sample alprazolam could not be statistically distinguished from either CT or placebo. Putting aside other problems with that study these results cannot be used to affirm the relative value of CT vs placebo in a medication responsive sample since that sample characteristic was not demonstrated. Let's say- to quote McNally "But let's suppose our investigator finds that cognitive therapy produces significantly more panic remission (say, 90%) than does imipramine (say, 70%) which, in turn, does better than placebo (say 30%). What do we now conclude? Do we say that cognitive

therapy is more effective than imipramine which is more effective than placebo? That's one interpretation. Another is to assert that we drew our patients from a RELATIVELY imipramine-nonresponsive population, and infer this from the failure of imipramine to do at least as well as cognitive therapy. Indeed, the proportions of imipramine- responsive patients in any give sample will likely vary; it won't be an all-or-nothing thing. The upshot of all this is that we can apply the concept of (relative) imipramine-nonresponsiveness any time that we discover that imipramine performs less well than cognitive therapy. " the problem here is that clinical trials ,if properly done, should be on random samples of the population of concern thus allowing proper generalizations. needless to say this is never done. instead we rely on grab samples and unsystematic replication to make the case. because of this all conclusions from clinical trials should be in the format--this sample's contrasts generalize to a population that is hard to define.cold comfort, practically mitigated by replication. however if you can't demonstrate drug superiority to placebo in that sample you are debarred from claiming psychotherapy superiority to placebo in any drug responsive population. in mcnally's first example (like klosko/barlow)it is moot whether this is a relatively drug unresponsive sample.(i don't share his enthusiasm for yohimbine altho lactate or co2 might help).however (in McNally's second example) if ct was better than drug which was better than placebo(ideally in a contrast similar to what is modally reported), and this was widely replicated , I would agree that there must be a fairly substantial strata of drug responsive patients where ct is superior treatment. let's see those data.with regard to non-specificity note both shear and ost and someone i forget have shown equivalence of non-ct with ct. finally it is of no practical importance that black did not include a psychosocial placebo. if a psychotherapy cant beat pill placebo it cant be terrific.

From: Neil Jacobson <njacob@u.washington.edu Mon Jan 9 04:38:28 1995 To: DonaldK737@aol.com Subject: Re: more on TDCRP

On Sun, 8 Jan 1995 DonaldK737@aol.com wrote:

> we are currently analysing these public access tapes. So far no evidence of site interaction.(sites distinguished but not identified). I entirely agree that these results have been too long delayed. <
don klein

I hope my other points did not get lost. TDCRP provides NO evidence that drugs work better than either psychotherapy for the more severe subgroup. This is elementary statistics. The Elkin paper is clear that this was a post hoc

analysis, and even then the groups did not differ significantly. Even if they had, they would have simply been hypotheses for further research. And even if they had been predicted in advance, they would need to be replicated. they were not replicated by Thase and Simons at WPIC.

From DonaldK737@aol.com Mon Jan 9 04:38:48 1995 To: seligman@cattell.psych.upenn.edu Subject: evaluating psychotherapy/pharmacorx
--

I appreciate Marty Seligman forwarding my critique and look forward to a rational discussion. However Marty misread my statement I did not say "That IPT is specific to social problems and CT to cognitive patients" but quite the reverse for more severe states. counterintuitive but good Jerome Frank sense.
Donald Klein

From rjm@burrhus.harvard.edu Mon Jan 9 16:40:11 1995 To: sscpnet@bailey.psych.nwu.edu Subject: comment on Klein's comment
--

Dear Don:

I'm glad we agree that at least in principle one might demonstrate that CBT is more effective than drug in a drug-responsive sample, as exemplified by my hypothetical case involving panic-remission rates of 90% (CBT), 70% (drug), and 30% (pill placebo). If my memory serves me correctly, Dave Barlow's multicenter trial includes the relevant control groups to investigate this possibility. Thus, the data bearing on this issue ought to be available soon. I am still not entirely comfortable about explaining a disappointing drug versus placebo outcome solely by assuming that the sample must somehow be unrepresentative. That is, it would be helpful to have independent evidence of unrepresentativeness other than the very data unrepresentativeness is designed to explain.

It would be especially interesting, for example, if one could distinguish between responders versus nonresponders in their response to biological challenges such as lactate infusion. What should we make of Klosko et al.'s failure to obtain a drug versus placebo effect? Surprise at their failure to obtain an effect seems to presuppose that alprazolam is, indeed, a reliably effective antipanic agent. But the effectiveness of this compound was precisely what was at issue (among other things) in this early trial.

I agree that CBT's failure to outperform pill placebo in the Black et al. study hardly constitutes a ringing endorsement of CBT for panic disorder. But I can

imagine someone drawing the following conclusions: "Given that panic remission rates have been much higher in other studies (e.g., Clark et al., Barlow et al.) involving CBT, there must have been something strange about the University of Iowa patients or how they were treated. Perhaps they constituted a 'CBT-nonresponsive' sample or perhaps CBT was not conducted with the expertise typical of that at Oxford or Albany."
Rich McNally

From njacob@u.washington.edu Mon Jan 9 19:20:30 1995 To: "Richard J. McNally" <rjm@burrhus.harvard.edu> Subject: Re: comment on Klein's comment
--

You are too generous in your critique of the "drug responsiveness" argument. Actually, the argument is completely circular. An alternative interpretation is that the drug doesn't work better than the placebo, and that studies showing an effect were flawed by the fact that those who administered the pills were not really blind. Greenberg and colleagues have shown that in studies where old antidepressants are the control groups for studies of the new antidepressants, the old antidepressants typically fail to outperform placebos. Furthermore, in the NIMH study the patients were similar on severity criteria to studies which did show a drug effect, yet there was no drug effect except for post hoc analyses in TDCRP.

From zuckerma@chopin.udel.edu
Mon Jan 9 19:33:20 1995
To: "Richard J. McNally" <rjm@burrhus.harvard.edu>
Subject: Re: comment on Klein's comment

Dear Rich:

It might be useful to compare the responders vs. non-responders to each type of therapy to see if they could be distinguished by onset and symptom characteristics of their depressions. Of course this would be post-hoc, but interesting leads could be followed up by deliberate selection studies. Population differences may not be the whole story of inconsistent findings but they may be an important part of it.

Sincerely,
Marvin

From rjm@burrhus.harvard.edu
Mon Jan 9 21:45:25 1995
To: njacob@u.washington.edu
Subject: Re: Re: comment on Klein's comment

Actually, I was not being overly generous regarding the circularity issue, just merely polite. Your point about raters being nonblind is a good one. Thanks to side-effects, maintaining blindness can be tough, indeed.

Jurgen Margraf and his colleagues published some interesting data regarding panic disorder on this issue in JCCP in 1991.

Rich McNally

From njacob@u.washington.edu
Mon Jan 9 21:57:19 1995
To: "Richard J. McNally"

In my last comment, I did not realize that the debate was about panic. But the argument that a study comparing psychotherapy with pharmacotherapy is invalid unless the drug therapy works better than a placebo is not sound for any disorder, because it presupposes that drugs work better than placebo when in fact it may not.

There have been numerous failures to replicate drug vs. placebo comparisons across a variety of disorders. It is true that such a finding means that the sample was not drug responsive, but that does not mean it was not a clinically relevant, or even a severely impaired population. It simply speaks to the lack of robustness of the drug vs. placebo difference. The circularity of this argument has been lost on many. Imagine if psychotherapy researcher were to

argue that all drug studies are invalid unless they included a "psychotherapy responsive" sample, i.e., a sample where psychotherapy had been established to be more effective than some control condition. This would invalidate 99% of the drug studies that have thus far been completed.

From rjm@burrhus.harvard.edu Tue Jan 10 17:03:13 1995 Subject: drug trial methods

Dear Don:

Klosko and Barlow's failure to obtain a significant alprazolam versus placebo panic remission effect need not imply that their sample was drug-unresponsive. Indeed, their data are remarkably similar to those of Ballenger et al. whose study provided the chief basis for alprazolam's obtaining FDA approval. Panic remission rates in Klosko-Barlow were 50% (alprazolam) and 36% (placebo), whereas panic remission rates in Ballenger et al.'s study were 55% (alprazolam) and 32% (placebo). Ballenger et al.'s effect was significant, probably because their statistical power was greater than that of Klosko and Barlow. Therefore, Klosko and Barlow's alprazolam results are not dramatically different from those of other investigators. In fact, had Klosko and Barlow obtained similar results with a larger sample, their study would have fit my hypothetical case of behavior therapy outperforming drug which outperforms placebo in a "drug-responsive" sample. I agree that ensuring fidelity to psychotherapy protocols is more challenging than ensuring fidelity to pharmacotherapy protocols. But failure to conduct psychotherapy protocols optimally may be precisely WHY investigators have encountered difficulties distinguishing among psychotherapies. Therapists whose adherence to protocols varies widely are likely to produce substantial variance that obscures potential differences across therapies. Best regards,
Rich McNally

From: DonaldK737@aol.com
Tue Jan 10 05:51:06 1995
To: njacob@u.washington.edu
Subject: Re: comments on Don Klein's posting

I will shortly answer your questions. however I suggest that you read

Klein, DF & Ross, DC: Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. *Neuropsychopharmacology* 8(3):241-252, 1993. also-

Klein, DF: NIMH collaborative research on treatment of depression. *Arch Gen Psych* 47(7):682-684, 1990.

e mail is not ideal for extended scholarly discussion and an informed audience shortens useless discussion. it appears from your questions that these documents eluded your notice.

From: rjm@burrhus.harvard.edu
Tue Jan 10 16:01:10 1995
Subject: Re: Re: comments on Don Klein's posting
To Rich Mc Nally

Dear Rich,
you say "I am still not entirely comfortable about explaining a disappointing drug versus placebo outcome solely by assuming that the sample must somehow be unrepresentative. That is, it would be helpful to have independent evidence of unrepresentativeness other than the very data unrepresentativeness is designed to explain. It would be especially interesting, for example, if one could distinguish between responders and nonresponders in their response to biological challenges such as lactate infusion. What should we make of Klosko et al.'s failure to obtain a drug versus pill placebo effect? Surprise at their failure to obtain an effect seems to presuppose that alprazolam is, indeed, a reliably effective antipanic agent. But the effectiveness of this compound was precisely what was at issue (among other things) in this early trial.

I agree that CBT's failure to outperform pill placebo in the Black et al. study hardly constitutes a ringing endorsement of CBT for panic disorder. But I can imagine someone drawing conclusions similar to yours concerning drug unresponsiveness: "Given that panic remission rates have been much higher in other studies (e.g., Clark et al., Barlow et al.) involving CBT, there must have been something strange about the University of Iowa patients or how they were treated. Perhaps they were inadvertently drawn from a 'CBT-unresponsive' population or perhaps CBT was not delivered with the expertise

typical of that of Oxford or Albany." it isn't symmetrical. Black had a placebo control and Clark didn't. with regard to k/b i think we can take it as established that alprazolam works. even the conservative fda agrees. Therefore it is their trial and sample that is thrown into question. you make it sound like an infinite regress and it isn't practically or theoretically. what you refer to as dave barlow's multicenter trial was largely instigated by my colleague gorman with my critical support. we are interested in getting understandable data and showing the field the methodological necessities in this tough area. with regard to defective expertise in treatment delivery it is obviously true that delivering medication in a relatively constant setting is a lot less difficult than assuring psychotherapy equivalence. however this concern is somewhat mitigated by the difficulty in distinguishing the benefits of quite different appearing psychotherapies. in any case it is tangential to my main point that a three arm trial (pill placebo, medication (well done), psychotherapy (well done)) is a minimum design for meaningful inference.

cordially, don klein

From DonaldK737@aol.com Wed Jan 11 19:27:21 1995 To: psycho-pharm@netcom.com, psycho-therapy@netcom.com Subject: Re: Re: Re: comments on Don Klein's posting

Rich McNally says:

>Klosko and Barlow's failure to obtain a significant alprazolam versus placebo panic remission effect need not imply that the sample was drug-unresponsive. Indeed, their data are remarkably similar to those of Ballenger et al. whose study provided the chief basis for alprazolam's FDA approval. Panic remission rates in Klosko-Barlow were 50% (alprazolam) and 36% (placebo), whereas panic remission rates in Ballenger et al. were 55% (alprazolam) and 32% (placebo). Ballenger et al.'s effect was significant, probably because their statistical power was greater than that of Klosko and Barlow. Klosko and Barlow's alprazolam remission rate was not really much different from that of Ballenger et al. In fact, had Klosko and Barlow obtained similar results with a larger sample, their study would have fit my hypothetical case of behavior therapy outperforming drug which outperforms placebo in a "drug-responsive" sample. <

Dear Rich,
you miss my point. If you wish your study to affirm that you have demonstrated that a psychotherapy is as effective or better than a medication you have to internally calibrate your sample by demonstrating statistically that it is medication responsive, in that sample. Your point that they might be medication responsive is off since it is the experimenter that is making the affirmation of equivalence. It is also incorrect to say that barlow gets similar results to ballenger. the numbers may look similar but they are only sample

estimates. the question is to what population can you generalize to from this grab sample. since you haven't shown a statistically reliable difference between drug and placebo in the b/k sample. You can't affirmatively declare that you are entitled to generalize to the population of drug responsive patients. You explain the difference in the inferences possible from b/k vs ballenger by lack of power, which assumes similar sample composition but that is what you are supposed to be demonstrating. You may be right but you may be wrong and there is no way to find out from that sample. certainly if b/k had the same results in a larger sample. Your point would be well taken but as things stand this is simply an act of faith. That is why I'm insisting on demonstrated sample relevance via an affirmatively demonstrated drug placebo difference. After all do you believe that the early Beck study showing clear superiority of ct to imipramine would not have benefited from a placebo arm. in the Elkin study this proved crucial in establishing the sample generalizability. I don't think you get the importance of our practice of non-representative grab sampling as a source of confusion. cordially

don klein

From njacob@u.washington.edu Wed Jan 11 21:56:39 1995 To Donald Klein

Dear Don,

I did not understand your last message to me. First, you said you would respond to my comments briefly. I didn't see any subsequent response. Apparently, some did, but I never got it. Second, you assumed that I had not read an article of yours, which, had I read it, would make me clearly understand that your arguments were valid and mine were not. Well, I had read the article, and I read it again. I fail to see the connection between what you did/find in this study and the points I was making in the internet TDCRP remarks. I was making some points which involve elementary research design and the logic of inferential statistics: first, that if differences between treatment groups are not statistically significant, they should not be discussed as if they are; second, a post hoc analysis cannot be used to both generate and confirm a hypothesis; third, if the between group comparisons in the post hoc analysis are not significant, then the post hoc analysis cannot be considered to have even generated a hypothesis; fourth, even if a hypothesis had been a priori rather than post hoc, they could not be discussed as "truth" prior to replication; and fifth, when there are site by treatment interactions which correspond to or can be explained by the allegiance or theoretical orientation of the sites, results are not interpretable without knowing the nature of the interaction.

These are all basic, fundamental points about research design, and based on them: first, the TDCRP found no evidence for the superiority of imipramine over cognitive therapy; second, the post hoc analyses involving severity similarly failed to produce significant differences between CT and imipramine for the severe subgroup; there is no interaction between treatment and severity to replicate; even if there had been, it would be uninterpretable without replication, and there have been several failures to replicate; and fifth, the "site" problem makes everything one big mess.

Above and beyond these points, there is the point that to demand demonstration of a sample being drug responsive in order for its comparison with psychotherapy to be valid is circular. What if failure to find a difference between drug and placebo is simply a failure to replicate, an indication of the lack of robustness of the effect, not a commentary on the nature of the sample? As I'm sure you know, there is now good evidence that when tricyclics are not the experimental treatment, but rather a control treatment to test a new antidepressant, they often fail to outperform placebos. Thus, the effect in past studies could be largely the result of blindness not being maintained, rather than a true drug effect.

Finally, I disagree with you in that I believe that the internet is a perfectly delightful format for debating these points.

Neil S. Jacobson

From STRAUMANT@macc.wisc.edu Wed Jan 11 21:42:07 1995
--

Perhaps, like Brad Smith, I am missing something as well with regard to Don Klein's interpretation of the presence vs. absence of drug-vs.- placebo differences across treatment studies. On a theoretical note, I am concerned that drug-responsiveness is (at least implicitly) being proffered as a gold standard for outcome studies in certain disorders. I recognize that this logic may be an understandable corollary of a view of (say) depression or panic disorder in which a specific pathophysiology is regarded as the undeniable, necessary- and-sufficient core of the disorder. However, if one holds that view, then it is hard to imagine how ANY data concerning the efficacy of psychotherapy would be persuasive. I have some passing familiarity with Klein's impressive research utilizing the psychopharmacologic dissection approach, but I remain skeptical that the links from pathophysiology to symptoms to drug-responsiveness are as clear-cut and ironclad as some would hold.

From a statistical perspective, why not compare the relative likelihood of the two competing interpretations discussed by Rich McNally? If Don Klein could supply some empirically-derived estimate of what proportion of all

panic sufferers are drug responsive, simple arithmetic would lead to probability estimates by which the reader could compare likelihoods. That is, given a sample of size X, what is the likelihood that the sample was (for whatever reason) non- representative enough to have accounted for the alleged discrepant findings? This likelihood could be evaluated the same way that the statistical power of particular studies factors into our interpretation of their findings.

Tim Strauman

Dept. of Psychiatry University of Wisconsin-Madison

B6/247 Clinical Science Center

600 Highland Avenue Madison, WI 53792

Phone: (608)262-5219

E-Mail: STRAUMANT@MACC.WISC.EDU

From rjm@burrhus.harvard.edu

Wed Jan 11 19:55:49 1995

Subject: drug trials

Dear Don:

I do not find your argument about the Klosko-Barlow trial entirely persuasive. Had their panic remission rate for alprazolam been notably lower than that of Ballenger et al., your argument about sample unrepresentativeness (i.e., medication-unresponsiveness) would have greater force. But when Ballenger et al. achieve a 55% panic remission rate with alprazolam, and Klosko-Barlow achieve a 50% panic remission rate with the same drug, one must strongly suspect that the drug responsiveness of the two samples are more similar than different (especially when placebo remission rates were nearly identical as well: 36% versus 32%). The most plausible explanation for why nearly identical drug versus placebo differences were significant in Ballenger et al.'s study but not in Klosko-Barlow's study is that the statistical power of the first study was far greater than that of the second study. To be sure, my interpretation is an "act of faith," but you, too, have invoked power interpretations to account for other nonsignificant effects in our field (e.g., Mike Telch's 1985 imipramine/behavior therapy trial in Behaviour Research and Therapy). Low power is as at least as plausible explanation as medication-unresponsiveness for Klosko-Barlow's failure to obtain a significant drug versus placebo effect. Regarding assumptions of equivalence of sample composition across studies, the best way to do this (in a noncircular way) is to rely on standardized entry criteria for the trial rather than on response to treatment in the trial itself. Regarding generalizability, are researchers who demonstrate a significant drug versus placebo effect in a two-group study barred from generalizing their findings to the population of "psychotherapy-responsive" patients? Do researchers need first to calibrate their sample to ensure that it is psychotherapy- responsive in order to generalize to the broad range of panic disorder patients?

Best regards,
Rich McNally

From rjm@burrhus.harvard.edu Thu Jan 12 21:30:37 1995 Subject: points of agreement/disagreement To Donald Klein
--

Dear Don:

Yes, we agree on inferential ambiguity, but disagree about causes. You're betting on sample unrepresentativeness, and I'm betting on power problems. We (apparently) agree that extant data are too scarce (or ambiguous) to satisfy BOTH psychosocial and psychopharmacologic researchers regarding comparative treatment effectiveness in panic disorder. Fair enough. We also agree that assumptions of sample equivalence ought (ideally) be established by explicit entry criteria, but disagree about the potential hazards of circular reasoning regarding inferring sample nonequivalence from response to treatment itself. Have panic disorder entry criteria been too lenient, thereby producing wide treatment versus placebo differences across studies? That surely seems possible. Here's yet another possibility: phenomenologically similar attacks may arise from different underlying dysfunctions, and these may respond differently to different treatments.

For example, Dysfunction A may be more responsive than Dysfunction B to Drug X but not to Drug Y (or vice versa). But our diagnostic criteria however stringent they may be, may be insufficient to capture distinctions between phenomenologically similar attacks arising from different dysfunctions.

Best regards,
Rich McNally

From DonaldK@ Thu Jan 12 21:32:52 1995 Subject: understanding NIMH study
--

As part of our general discussion of proper experimental design for estimating the relative merits of pharmlrx and psyrx there have been a number of comments by njacob@u.washington.edu that require rebuttal. Since I was a member of this key study's Scientific Advisory Committee that developed the protocol I have a close understanding of the process..

n jacobson states:

>The TDCRP found that there were no differences between pharmacotherapy, CBT, and Interpersonal Psychotherapy (IPT). Based on a priori hypotheses, there is no evidence from this study that CBT was inferior to either IPT or Imipramine plus clinical management. As a post hoc, hypothesis generating tactic, the authors (Irene Elkin and colleagues, in a 1989 article published in the Archives of General Psychiatry), did a median split on severity). I repeat: there were no PREDICTED interactions between severity and outcome. Based on these post hoc analyses, there were still no differences between the two psychotherapies, or between the drug and psychotherapy conditions, for either mild or severe depressives. However, only Imipramine and IPT outperformed placebo for this post hoc derived "severe" subgroup. Differences between CBT and placebo, while favoring CBT, were not statistically significant.. On this basis, people in the field have suggested that CBT may be inappropriate for severe depression, and this suggestion actually made it into a recently published set of treatment guidelines for major depression prepared by the American Psychiatric Association. There are so many problems with this level of discourse that I hardly know where to start".<

First of all, even after the post hoc median split on severity measures, there were no statistically significant differences between CBT and either IPT or drugs. Therefore, there is no basis for the inference that CBT is less effective than either IPT or drugs with severe depressives. Second, even if there had been significant interactions between severity and outcome, these interactions resulted from a fishing expedition. In other words, they were not based on a priori hypotheses, and therefore have to be viewed as hypotheses generating, exploratory analyses, which could only be tested on an independent sample. I learned in my very first statistics course that you cannot generate and confirm a hypothesis from the same set of data. Many seem to have forgotten this lesson in elementary statistics. This, this is just one study, albeit an expensive one.

Even if the findings had been predicted, and confirmed, they would await replication for confirmation. As it stands, there is no a priori finding to replicate, and the post hoc findings were null, which means there is hardly even a suggestive hypothesis generated by the post hoc analyses."this definite statement is in manifold error. The SAC discussed, a priori ,the importance of severity for outcome contrasts several times. for anyone with any experience in rx evaluation it is blindingly obvious. further I personally corresponded with elkin several times about just this, prior to the analyses. finally elkin

DID do severity analyses in her original total sample in the form of analyses of covariance properly preceded by analyses for heterogeneity of slope. It may not be elementary statistics but such analyses are severity by outcome analyses. So why did they find so little? as we point out

(Klein, DF & Ross, DC: Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. *Neuropsychopharmacology* 8(3):241-252, 1993.also see---Klein, DF: NIMH collaborative research on treatment of depression. *Arch Gen Psych* 47(7):682-684, 1990.)

They made several errors that may not be discussed in an elementary statistics course. Most egregious they fostered type 2 error by inappropriately stringent standards for analysis of heterogeneity of slope compounded by a contra-indicated bonferroni correction and unduly conservative 4 group analyses. These erroneous decisions by the NIMH staff were never discussed with the SAC . Note that with their low power approach only imipramin was superior to placebo. so imipramin was once more reconfirmed. The problem was that none of the other pairwise contrasts made significance by their techniques.

Correcting these errors, properly analysing for ancova model invalidating (but meaningful) heterogeneity of slope, and following up with the indicated johnson-neyman analyses on the unsplit sample, Ross and I found clear evidence for imipramine superiority to both ipt and ct as well as placebo, particularly in the more severe cases.further (from other reports) imipramine effects were more rapid and cheaper.this is not post hoc analysis. However as some consolation we did find some spotty evidence that the psychotherapies were statistically superior to placebo. Even here there is a hook.comparing ipt to cbt was fruitless by Elkin's low power methods.surprisingly on the Beck BDI scores, among the completers and those who had at least 3.5 weeks of rx, cbt was clearly inferior to ipt in patients who started with a bdi=> 30,the usual threshold for severity.in other words in just those patients where cbt would seem most indicated it did the worst among the psychotherapies. remarkably sotsky et al using quite different analyses found the symmetrical finding with regard to interpersonal difficulties.patients with high interpersonal problems did worse on interpersonal therapy than cbt! This hardly supports specificity of psychotherapeutic mechanism, but is expectable from Jerome Franks theory that the job of psychotherapy is to counter demoralization rather than to treat the illness. Focussed psychotherapies rub the patients nose in just the area where the patient has consistently failed, so this is unlikely to be a success experience. Non directive therapy may have its points, as Shear showed.

Jacobson continues: "this is just one study, albeit an expensive one. Even if the findings had been predicted, and confirmed, they would await replication for confirmation". The lack of replication attempts is a real issue but you do your best with what you have. The inference to me is that recommending CBT for severe depression is unwarranted. The recent AHCPR guidelines suggest a

switch in treatment if nothing happens in 6 weeks. Does cognitive theory suggest that this is possible? Is there any properly controlled data? I don't think so. Since the Elkin publication we have hardly been swamped with properly complex studies. I know of only two, both partially instigated by my Therapeutics department at NYSPI, in the areas of social phobia and panic. If we want such studies it will require a programmatic change in NIMH thinking and more self criticism in the psychotherapy community. For instance N Jacobson says: "What about previous studies by David Clark, Dave Barlow, and others? What about Beck's data?" well, what about it? Are we engaged in dropping names or talking design and data? Note that almost none of these studies were appropriately controlled and do not allow comparative rx inferences. Try being specific. N Jacobson also states: "An alternative interpretation is that the drug doesn't work better than the placebo, and that studies showing an effect were flawed by the fact that those who administered the pills were not really blind. Greenberg and colleagues have shown that in studies where old antidepressants are the control groups for studies of the new antidepressants, the old antidepressants typically fail to outperform placebos".

The Greenberg review is seriously flawed. If anyone wants a rebuttal write me. The idea that antidepressants really don't work is flabbergasting. See any of John Davis' reviews. If you maintain, in the face of common sense, that side effects breaking the blind and bias are doing the job of getting drugs approved by the FDA, it should be realized that 9 out of 10 drugs that go into clinical trials bomb out despite their side effects. If bias and piercing the double blind are sufficient why aren't they all approved? Pardon the length of this communication but science demands close critical reasoning and it's hard to be short.

Donald Klein

From njacob@u.washington.edu Thu Jan 12 21:37:11 1995 Subject: Re: understanding nimh study To Donald Klein@

Dear Don,

See? The internet is an excellent forum for debate. No time for a detailed rejoinder now, but if you read the Elkin et al. article, it says at least twice that the severity analyses were post hoc and exploratory. I prefer to base my inferences on the literature, rather than by the oral/rabbinical tradition, or what I like to call "science by rumor".

neil

From njacob@u.washington.edu

Fri Jan 13 00:01:23 1995 Subject: Re: understanding nimh study I To sscpnet@
--

More on TDCRP.

First, I remain confused about the discrepancy between the way the Elkin paper was written and the description of events as you have described them. There is a major difference between a priori hypotheses with specific predictions and post hoc findings. However, whether the findings were a priori or post hoc, and whether your analyses with greater statistical power are more correct than those of the original Elkin et al. paper, one study is not enough basis for making a treatment recommendation to primary care physicians.

In fact, there have been other studies, coming out of laboratories such as Steve Hollon's and the collaboration between Michael Thase and Ann Simons and WPIC suggesting that CBT suggesting that there is no interaction between severity and treatment response (comparing psychotherapy to pharmaco-therapy).

There are at least two meta-analyses, one published by Keith Dobson in JCCP and the other by Robinson and Berman in Psychological Bulletin: neither suggest an advantage for pharmacotherapy over psychotherapy, and Dobson's actually suggests an advantage for cognitive therapy.

So, the literature is ambiguous, and the argument that the only valid comparisons between drugs and psychotherapy must include a placebo is logically circular. Others on the internet have already spoken to this issue quite compellingly.

In short, a study can't be discounted because it doesn't have a drug placebo, any more than a drug study can be discounted because it doesn't have a psychotherapy placebo.

Furthermore, in the 18 month follow-up of TDCRP, the percentage of people who stayed in treatment, recovered, and stayed recovered was 19% for imipramine and 20% for placebo: hardly a ringing endorsement for tricyclics! CT, I believe had a 32% response rate, not significantly better, but as Dr. Klein has argued, perhaps that was because of inadequate power.

I am not so much arguing the CBT is wonderful as I am that antidepressants are not all that they are cracked up to be. I actually agree with Jerome Frank's point about psychotherapy, and agree with Klein's point about lack of specificity of psychotherapy effects. But all of these arguments can also be made about pharmacotherapy, which includes clinical management (which I consider to be a form of psychotherapy).

Finally, the trials by David Clark and David Barlow, not to mention the work of the Penn State group (Larry Michelson and Tom Borkovec for panic [with or without agoraphobia] and GAD respectively) suggests great promise for CBT as a treatment of panic. To dismiss these findings because there was no drug placebo is quite problematic. There was great rigor and care taken in each of these trials, and they are compelling, even to a skeptic like me.

From njacob@u.washington.edu Fri Jan 13 00:06:13 1995 Subject: Re: points of agreement/disagreement

On Thu, 12 Jan 1995, Richard J. McNally wrote:
<see above>

Here is a well-stated example of the circularity problem. If I had read it prior to my previous response, I would have referred to it.

From DonaldK737@aol.com Fri Jan 13 16:37:18 1995 Subject: Re: points of agreement/disagreement
--

Rich McNally says: >SEE ABOVE<

dear Rich,

I think your summary is incomplete. My major point is that there is a way to resolve these disputes experimentally in the form of 3 armed (pill placebo, pharmlrx, psyrx) or 4 armed (pill placebo, pharmlrx, psyrx, psyplacebo) controlled replicated trials. That's what I would like some agreement about. also I think you are being inconsistent. You say that you are "betting on power problems" but then go on to give an excellent example of how sample disparity may well occur despite apparently similar diagnostic criteria. In line with this the Brandon / Briggs study (B J Psychiat) showed that among panic patients those with dyspneic panics responded better to imipramine than alprazolam and that the converse was true among non dyspneic panickers.

cordially
don klein

From bsmith+@pitt.edu Fri Jan 13 16:46:31 1995 Subject: Severity ATI and Power
--

I find Don Klein's responses to Neil Jacobson regarding the so called "severity by treatment ATI" in the Collaborative Study on Depression to be rather unsatisfying. First of all, if the ATI was anticipated, why does Elkin et al report it as post-hoc? Secondly, how many ATIs were looked for. If the severity by treatment interaction HYPOTHESIS is lodged in the context of discovery (i.e., exploration) and its importance is diminished by each successive post hoc ATI search (e.g., therapist characteristics by outcome ATI hypotheses, site by treatment ATI hypotheses, SES by treatment ATI hypotheses, etc.) How many of these ATIs were searched for? Are ALL of the failed explorations reported anywhere? (almost certainly not). My concern is that, if somebody did 20 explorations one would be expected to be statistically significant due to chance alone ($p = .05$ and all that jazz). Shouldn't we be looking at experiment-wise error of the post-hoc analysis? Reporting the failures would help with this point.

In sum, I completely agree with Neil Jacobson that the ATI hypotheses reported MUST be presented within the context of discovery, NOT, as it has been, in the context of justification (i.e., as a confirmed result). Finally, I am dismayed by the discussions of statistical power. The focus should be on clinically meaningful improvement. Give me enough subjects and I'll find any effect. Sheer power is boring and potentially misleading. What I want to know is (1) did the subjects improve in a clinically meaningful manner, and (2) is the severity split a meaningful one or some arbitrary statistical cut (e.g., done at the median)? I have brought up these points in previous discussions, but it seems to have been forgotten in the current round. Please excuse the repetition.

Brad Smith WPIC

From rjm@burrhus.harvard.edu Fri Jan 13 18:05:35 1995 To Donald Klein Subject: Re: Re: points of agreement/disagreement
--

Dear Don:

Yes, my summary was incomplete: we both agree on what constitutes appropriate experimental designs for adjudicating these matters experimentally. I do not think I am being inconsistent, but I could very well be unclear regarding sample heterogeneity. In fact, I had the Brandon/Briggs study in mind when I conjectured that diagnostic similarity might mask biological heterogeneity. My concern is inferring heterogeneity SOLELY from response to treatment in the trial itself. But there are ways of avoiding the pitfalls of circularity even within the same data set, and you cite one example. Although Brandon/Briggs included patients who all qualified for the same diagnosis (i.e., phenomenologic similarity, broadly defined), those who responded poorly to imipramine ALSO differed in one other crucial

characteristic: dyspnea was not a prominent symptom of their panic attacks. Because there are independent, theoretical reasons for suspecting that dyspneic panics might be especially responsive to imipramine (and nondyspneic panics relatively less so), I think one is on firmer grounds conjecturing sample unrepresentativeness by in the imipramine nonresponders by noting that the cardinal symptom of dyspnea was missing from their panic profile rather than merely assuming sample unrepresentativeness by noting their response to treatment alone. That is, one can distinguish between treatment responders and nonresponders in a theoretically motivated, noncircular way other than just inferring this from their response to treatment alone. This is what I was driving at earlier when I mentioned that response to biological challenge tests (yohibine, lactate) might also be useful in distinguishing heterogeneity.

Best regards, Rich

From njacob@u.washington.edu Fri Jan 13 19:15:42 1995 Subject: Re: TDCRP
--

you say "See my last reply. Because that it is what it says in the Elkin et al. article. The words "exploratory" are used repeatedly, especially in the discussion section."

Why do you take Elkins confused words seriously if, as you said, you had read my article? My view is Elkin bungled the analysis and didnt understand the implications of what she was doing. To be precise she did ancova and analysed (badly) for heterogeneity of slope. Therefore de facto she was a priori analysing for severity - outcome interactions even if her confused presentation obscures this and she did not refer to past specific discussions on this point. also her arbitrary (not median) split of the sample was exploratory since no basis had been established for this, as you emphasize. But it was also unnecessary and obfuscatory as my analysis clearly shows. Action speaks louder than words. We spell that out in our article with Ross. please refresh your memory.

From njacob@u.washington.edu Fri Jan 13 19:17:34 1995 Subject: Re: drug trials
--

Another thing that I forgot to mention yesterday is that, according to rumor, the interaction between severity and treatment response in TDCRP held at only one site. Hence the site by treatment (by severity?) interaction. Thus,

there may have been two failures to replicate that interaction even within TDCRP. Perhaps we will never know the truth about these interactions, and whether they are predictable from the theoretical biases of the PI's and the staffs at the sites, but TDCRP was actually three studies, and the results were not consistent from one site to another. This weakens the argument for drawing inferences on the appropriateness of CBT for severe depressives even further.

From njacob@u.washington.edu Fri Jan 13 20:02:40 1995 To Donald Klein Subject: TDCRP

Don,

In reading over your responses to my internet messages, I noticed that you didn't respond to one major point. You claim that recommendations can be made to clinicians based on TDCRP since it is the only well-controlled study that we have comparing CBT to imipramine etc.

If you are correct, than what recommendations should we make in light of the fact that the response rate at the 18 month follow-up was 19% for imipramine, 20% for placebo, and 32% for CBT? Yesterday, I said it was "hardly a ringing endorsement for imipramine". Only a minority of people who receive this medicine stay in treatment, get better, and stay better. In fact, they are just as likely if not more likely to stay in treatment, get better, and stay better with placebo, and perhaps MORE likely with CBT.

Now that you have acknowledged that this is the only well-controlled study comparing CBT to imipramine, on that basis shouldn't we be advising primary care physicians that, on the average, across the entire range of major depression, CBT is more likely to be helpful than imipramine?

I would never make this argument, by the way, because I don't believe in making treatment recommendations to clinicians based on one study, especially one where there was not consistent replication across sites. To do so would be using science as a rhetorical device. I fear that is what you are doing. The appropriately cautious, scientifically justifiable, conclusion to draw based on TDCRP is that none of the treatments are very effective, and that clinicians should tell their clients that there is far less than a 50-50 chance that either tricyclics, IPT, or CBT will lead to lasting recovery.

Neil

From DonaldK737@aol.com Sat Jan 14 16:10:33 1995 Subject: Severity ATI and Power
--

To repeat, that Elkin misstated what she was doing is of no consequence. Doing ancovas and analysing for heterogeneity of slope is apriori analysis on entire sample with no cut and that was done.unfortunately done poorly but nevertheless done.action speaks louder than words. You explain why this was done. Unfortunately it was done poorly. Proper analysis i.e. analysing for heterogeneity of slope followed by johnson neyman requires no split samples. (check klein/ross article). Criticise that rather than ignore Elkins actions and stick with admittedly confused wording to maintain that this elementary consideration in any treatment analysis was somehow neglected.

From DonaldK737@aol.com Sat Jan 14 15:33:05 1995 Subject: Re: Severity ATI and Power
--

brad smith says:

First of all, if the ATI was anticiapted, why does Elkin et al report it as post-hoc? Secondly, how many ATIs were looked for. If the severity by treatment interaction HYPOTHESIS is lodged in the context of discovery (i.e., exploration) and its importance is diminished by each successive post hoc ATI search (e.g., therapist characteristicsby outcome ATI hypotheses, site by treatment ATI hypotheses, SES by treatment ATI hypotheses, etc.) How many of these ATIs were searched for? Are ALL of the failed explorations reported anywhere? (almost certainly not). My concern is that, if somebody did 20 explorations one would be expected to be statistically significant due to chance alone ($p = .05$ and all that jazz). Shouldn't we be looking at experiment-wise error of the post-hoc analysis? Reporting the failures would help with this point. In sum, I completely agree with Neil Jacobson that the ATI hypotheses reported MUST be presented within the context of discovery, NOT, as it has been, in the context of justification (ie., as a confirmed result). Finally, I am dismayed by the discussions of statistical power. The focus should be on clinically meaningful improvement. Give me enough subjects and I'll find any effect. Sheer power is boring and potentially misleading. What I want to know is (1) did the subjects improve in a clinically meaningful manner, and (2) is the severity split a meaninful one or some arbitrary statistical cut (e.g., done at the median)? I have brought up these points in previous discussions, but it seems to have been forgotten in the current round. Please excuse the repetition.

why Elkin did not verbally state the glaringly obvious and absolutely standard psychopharmacological hypothesis of ATI is beyond me. it is not like this hadnt been pointed out prior to the analysis.what elkin describes as post hoc was the arbitrary subdivision of the sample and this was both post hoc and unnecessary. but action speaks louder than words and her (poorly carried out) analysis for ancova heterogeneity of slope on the unsplit sample speaks loudly

enough i dont think you have read the klein/ross article either because it contains a discussion of your other points .in particular it critiques the unthinking application of family wise alpha rates in multi treatment situations where each contrast is meaningful and there is no overall inference to be made that depends on the validity of each pairwise contrast.we also cite appropriate statistical references indicating that this is not a maverick idea but standard statistical thinking. please note none of this has to do with split samples. is this a discovery? Particularly given the grab sample problem replication is extremely desirable, however fobbing off the largest best controlled study in an area that cries out for placebo controls seems tendentious at best. The study has been around a long time now. Where are the replications from those who affirm the benefits of CBT or IPT ? the burden of proof is on them. I couldnt agree more with regard to meaningful clinical improvement. in our current analysis of the public release tapes we will do this.
don klein

From DonaldK737@aol.com Sat Jan 14 15:32:59 1995 To: njacob@u.washington.edu Subject: Re: drug trials
--

NJacobson says another thing that I forgot to mention yesterday is that, according to rumor, the interaction between severity and treatment response in TDCRP held at only one site. Hence the site by treatment (by severity?) interaction. Thus, there may have been two failures to replicate that interaction even within TDCRP. Perhaps we will never know the truth about these interactions, and whether they are predictable from the theoretical biases of the PI's and the staffs at the sites, but TDCRP was actually three studies, and the results were not consistent from one site to another. This weakens the argument for drawing inferences on the appropriateness of CBT for severe depressives even further.

I thought you were against science by rumor.this is a new low in affirming the null hypothesis. To repeat - none of our analyses show a site interaction. see klein/ross . None of our current analyses show a site interaction.

don klein

From DonaldK737@aol.com Sat Jan 14 17:00:47 1995 Subject: Re: Re: points of agreement/disagreement
--

dear rich
we are in agreement re desirability of evidence over inference
don klein

From I.Pitchford@sheffield.ac.uk Sun Jan 15 01:10:31 1995 To: "Robert C. Carson" <rcarson@acpub.duke.edu>, Subject: Re: Drugs vs. Psychosocial Priority:

Dear Bob,
you wrote: Fellow members of SSCP:
>I hope someone is saving for posterity the current discussions among Jacobson,
Klein, McNally, Smith, et al. on the relative merits of pharmacologic and
psychosocial interventions, especially in depression. (I now am, but I got a late start).
<

REPLY: I am saving all of the responses posted here and on several of the
InterPsych lists. With the permission of the authors these will be collated and
converted into HTML format as a permanent archive of the debate.
They far exceed anything in the standard literature, some of which is so
obviously contaminated by politico/economic considerations as to be worse
than unintelligible.

Dr. Klein, in my judgment, represents the biological psychiatry viewpoint in
the most knowledgable and expert way I have ever seen, which is not wholly
surprising in light of his distinguished career as a visionary leader in the area.
My personal opinion is that it still loses, despite the spirited and expert
defense, but I'm not here trying to join the argument--only to express my
delight and amazement about what shows up on this net! Thank you guys.
Bob Carson REPLY: I endorse the foregoing wholeheartedly. Best wishes

Ian Pitchford Secretary, InterPsych Board of Directors Mr Ian Pitchford Department of
Biomedical Science University of Sheffield Western Bank SHEFFIELD, S10 2TN United
Kingdom Tel: +44 742 780319

From: k-howard@nwu.edu (Ken Howard) Tue, 17 Jan 1995 09:54:51 -0600 To: sscpnet@bailey.psych.nwu.edu Subject: The Don and Neil Show
--

Don Klein and our own lovely Neil Jacobson have come to the conclusion
that some patients under some conditions are better treated by one procedure
rather than another. In the grand tradition of Lee Sechrest (who has been
suspiciously quiet) let me share an experience.

Yesterday, I had an appointment with a dermatologist. After some lab tests and physical examination, he informed me that symptoms such as mine can be caused by many things (i.e., independent variables) and that most probably my symptoms were caused by some combination of these things. He then outlined a series of treatments that varied in their intrusiveness and side effects (including a treatment that he claimed was as good as a placebo!). But, since my symptoms had significantly abated in the 24 hours preceding the appointment, he suggested that I just ignore them and they would probably go away. Wish me luck.

Ken Howard Department of Psychology Northwestern University Evanston, IL 60208-2710 Phone: (708) 491-7373 Fax: (708) 491-7859

From k-howard@nwu.edu Tue Jan 17 18:10:19 1995 To: sscpnet@bailey.psych.nwu.edu Subject: Freud-Jung Debate

Like many of you, I have been "listening" to the exchanges regarding comparative treatment efficacy. I have found them historically interesting, much in the sense that I enjoyed reading the letters that Freud and Jung exchanged regarding the primacy of the id vs. the primacy of the collective unconscious.

I was shocked into reality, however, by Bob Carson's sense that there was something of current interest in the exchange, something that is so important that it should be archived. It occurred to me that Bob had been trained by Ben Underwood (may he rest in peace), who had been analyzed by Sir Ronald Fisher, and so was stuck with the outdated methodology of the randomized clinical trial. Lest younger, more open, minds be swayed let me remind clinical researchers of some limitations of this approach.

- | |
|---|
| <ol style="list-style-type: none">1. Nature abhors main effects (personal communication).2. There is virtually always within cell variance that exceeds measurement error (or any kind of "random" error).3. Random assignment of relatively few (say 20 per cell) subjects virtually never equates groups with regard to possible confounds, except OVER THE LONG RUN (thus, the need for replication regardless of whether hypotheses are pre hoc or post hoc).4. Attrition virtually always compromises a real trial. Some MODERN references that can be helpful in comparative treatment research. |
|---|

see

1. Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.
2. Cohen, J. (1990). Things I have learned (so far). American Psychologist. 45, 1304-1312.

3. Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494-509.
4. Rogers, J., Howard, K.I., & Vessey, J. (1993). Using significance tests to establish equivalence between two experimental groups. *Psychological Bulletin*, 113, 553-565.

Ken Howard Department of Psychology Northwestern University Evanston, IL 60208-2710 Phone: (708) 491-7373 Fax: (708) 491-7859

From: Neil Jacobson <njacob@u.washington.edu> Tue, 17 Jan 1995 19:54:51 To: Ken Howard <k-howard@nwu.edu> Subject: Re: Freud-Jung Debate

It was flattering to be compared to Freud and Jung (although I don't know which one was Klein and which one was I).

Ken's eccentric opposition to randomized clinical trials, and faith in naturalistic studies of psychotherapy outcome is wellknown. I share his concerns about randomized trials, although I think they should form a part of the research enterprise, since as part of a whole, they do allow for internally valid cause-effect inferences.

Whatever their limitations, it is important that valid inferences be drawn from them, and that they not be used as rhetorical devices to undersell psychosocial treatments and oversell biological treatments. Clinical trials in general have not produced encouraging findings about the clinical significance of either psychotherapy or pharmacotherapy effects. However, they do provide more optimistic estimates of psychotherapy efficacy than naturalistic studies do, at least for child disorders (see classic article by John Weisz in *American Psychologist*). For adult disorders, I believe a recent meta-analysis found that clinical trials and naturalistic studies lead to roughly equivalent estimates of psychotherapy effect size.

From: "Robert C. Carson" <rcarson@acpub.duke.edu> Tue, 17 Jan 1995 To: Ken Howard <k-howard@nwu.edu> Subject: Re: Freud-Jung Debate
--

Friends, colleagues, countrymen,

I posted my recent posting not to praise the research that has been the subject of the Jacobson et al. et al. debate, but to (help) bury it--which is the direction in which I took the debate to be moving.

I was shocked at Ken Howard's false attribution that I have learned nothing since my late adolescence at the feet of Ben Underwood, although I shall acknowledge that, in one way or another, he did anticipate most of what I've learned since.

Bob Carson

From k-howard@nwu.edu Tue Jan 17 21:30:53 1995 To: sscpnet@bailey.psych.nwu.edu

My sense is that Freud was more biological than was Jung, but I can't remember which was taller. I am not against clinical trials, although I can see no evidence that they provide superior internal validity for treatment comparisons where attrition and the failure of randomization to equate groups have been rampant. What I am arguing for is a patient-centered rather than a treatment centered approach. This is part of the old specificity litany, but boils down to what is the best approach for this patient at this time, given what is realistically available.

Effect sizes have some information, but as we have pointed out any effect size smaller than about 3.00 means that some patients were better off with the inferior treatment. Who were those patients? How can we better help them?

Ken Howard Department of Psychology Northwestern University Evanston, IL 60208-2710 Phone: (708) 491-7373 Fax: (708) 491-7859

From: Neil Jacobson <njacob@u.washington.edu> Wed, 18 Jan 1995 To: "Robert C. Carson" <rcarson@acpub.duke.edu> Subject: Re: Freud-Jung Debate
--

Let us not bury TDCRP. There is much to be learned from that study. For one thing, it might be the best drug trial ever designed. It is also a fascinating forum for examining how scientific findings can be used as rhetorical devices. Finally, if we bury the TDCRP, we may never find out about that site by treatment interaction. And, worse, someone may try such a study again. I think one thing Bob, Ken, and I agree on is that multi-site large scale randomized trials are not the best way to advance knowledge in an age where resources are slim and getting slimmer.

From: "David H. Barlow" Mon Jan 23 22:24:19 1995 To: sscpnet@bailey.psych.nwu.EDU Subject: Drugs and Psychotherapy

Warning - Long Message

Having just returned from 10 days of travel, I have learned another important lesson. Leaving the Net for too long, can be dangerous to the integrity of your data and even your mental health. In any case, I am happy to belatedly join the discussion with my colleague and research collaborator, Don Klein, on the current status of psychosocial and drug discussions we have had before in other forums but many members of this Net may not have heard them.

In this communication I will limit my remarks to a few facts on the psychosocial treatment of panic disorder. There are now at least 14 (count 'em) controlled trials of the psycho- social treatment of panic disorder wherein a cognitive-behavioral treatment such as panic control therapy (PCT) originated in Albany or cognitive therapy as developed by David Clark and his colleagues has been compared to either a wait list control or some credible alternative treatment. Two of the studies (Margraff & Schneider, 1993; Telch et al., 1993) utilized only wait list comparison conditions, the others compared CBT to either pill placebo, credible psychosocial placebos, or alternative psychosocial treatments with some evidence for effectiveness in their own right. Taking all but 3 of these studies, the percent of patients free of panic based on intent to treat analysis in each study at points ranging from post treatment to 24 months follow-up ranged from 76% (Clark et al., 1994) to 100% (Cote et al., 1992). In addition, there are many clinical replication series (series of cases) showing similar panic free results, one of them from Don Klein's own clinic (Welkowitz et al., 1991). In one of the 3 studies where results were not as good (Craske et al., 1994), an abbreviated version of PCT consisting of only 4 sessions achieved 53% panic free compared to an abbreviated version of nonprescriptive therapy (a non directive therapy focusing on possible interpersonal triggers for panic) originated by Shear and colleagues where the comparable panic free figure was 8%. This leaves 2 studies where cognitive therapy faired very poorly. One was the study by Black et al. (1993). Using the same intent to treat analysis on which the above data are calculated, the percentage panic free after cognitive therapy was 32% compared to 20% with pill placebo and 68% with fluvoxamine. This poor result is due partly to the fact that fully 40% of the patients in the Black et al. study dropped out (based on personal communication from Black.... in the actual article they mistakingly report 36%). Why would CBT do so well based on reports from 4 different countries in a variety of settings including medical and psychosocial settings, and turn out to be positively toxic in Iowa with a large percentage of patients dropping out from what was a very brief treatment to begin with (8 sessions)? We do not yet know, although Black et al. reported they used "modified" cognitive therapy (with the modifications unclear) and there are real questions about the integrity of treatment since, according to Black et al., audiotapes from only 1 case were checked (there is

even some question about that) and these tapes were not subjected to normal treatment adherence procedures. But another more disturbing possibility is that it may be somewhat difficult to disseminate CBT to settings where expertise did not previously exist, although psychopharmacological clinicians such as those in Don Klein's setting seemed to have no trouble (Welkowitz et al., 1991). In the other study by Kathy Shear et al. (1994), the comparable intent to treat analysis at 6 months yielded 45% panic free for CBT and 45% panic free for "nonprescriptive" treatment (the same as in the Craske study above). Kathy reports using an early version of CBT and notes that "the CBT treatment adherence ratings were only moderately high and this may mean that CBT was not delivered optimally" (p. 400). Also, the first 3 sessions of CBT and sessions of PCT, but without an assessment at Session 3, we cannot know if any changes that did occur were the result of this common portion of treatment, or some other factors. In any case, the lack of a no treatment control condition and the high attrition rate make it very difficult to evaluate these results, which are more in line with psychosocial placebos in other studies. Another interesting issue is that these latter 2 studies were both published in the prestigious Archives of General Psychiatry despite the fact that, by general agreement, the conduct of the psychosocial portions of the clinical trials would not pass mustard for any decent psychosocial research journal; but that's another story. So what do we make of this? Obviously, this enlarged data base presents a very different picture than that painted by my colleague Don Klein. But there is also plenty to be self-critical about. Panic free status is a relatively poor measure of outcome. Recently, groups of biological and psychological investigators have gathered together to come up with better indices of overall functioning published in the Archives earlier this year (Shear, Maser et al., 1994). Secondly, my own view is that much as with medications, a number of well construed psychosocial treatments may turn out to be effective for panic disorder, including possibly Shear's NPT treatment when it is evaluated more rigorously. Third, comparisons of CBT with effective drugs are relatively limited. The recent study by Clark et al. (1994) comparing CBT and imipramine shows rough equivalence in terms of panic free status at 6 months, with results in the imipramine group then dropping off significantly after discontinuation of imipramine. On the other hand, results of CBT were sustained over the one year follow-up (as is true in every single study of the psychosocial treatment of panic disorder carried out thus far). In our own study (Klosko, Barlow et al., 1990), the percent panic free was 87% with PCT, 50% with Alprazolam, 36% with placebo, 33% with wait list control. It is no accident that these were almost exactly the same results as obtained in the much larger Upjohn Cross National study comparing Alprazolam and placebo, since our study was funded by Upjohn, and we used their protocol for administering the drugs. Why then was alprazolam significantly different from placebo in the large cross national study and not ours since the percentage panic free were about identical? The most reasonable and logical conclusion is that the lack of an alprazolam-placebo difference in our study is simply due to power rather than representativeness of the sample. We conclude this since the samples were very well defined (as

in the Upjohn study) and we have had noted experts come to our center and interview our panic patients, later reporting that these were "classic panic patients". I don't like to name drop, but the initials of one of these noted experts are DK. Finally, as many on the Net know, we are in the midst of a large collaborative study with our site and the Columbia site and two others that will examine carefully the separate and combined effects of these treatments and begin to answer the all important questions of patient treatment matching. All of us involved in the study, and I think I speak for my colleague Don Klein here, have confidence in the integrity of this trial although we should all remember that much like the TDRCP it is only one trial and some quirks may arise. Ultimately and inevitably there is going to be plenty of room for both psychosocial and drug treatments and, in many cases, their combination, in view of the enormous numbers of people suffering from emotional disorders.

David H. Barlow, Ph.D.
<MB399@ALBNYVMS.BITNET>

From PeterKramer@Brown.EDU Sat Jan 28 05:03:00 1995 To: psycho-pharm@netcom.com Subject: The Dodo Bird

Martin Seligman wrote [much omitted]:

>the evidence is quite overwhelming that CT and BT work better, by far, for depression, panic, OCD, sexual problems, and phobias than PP. By concession on the part of PP, not by test. >

There are a handful of very expensive long term PP versus other modalities in the literature. It is these studies that Luborsky's argument presumably rests on.

I consider the horse-race studies, pitting one psychotherapy against another, to be of limited utility. A good study will be multi-stage: manualizing treatment, training therapists, checking the standardization of their actions, selecting outcome measures, the testing the therapy, etc. For these therapies there is no dose-response curve and no demonstrated method of action. Now therapies are run against one another. By this point, the number of uncontrolled conditions and unknown forces is enormous. It is easily (and perhaps properly) argued in favor of the "less effective" treatment that it was given over the wrong time frame or that the training had a differential effect on therapists, etc., etc.

But more complex problems are possible. Let us say that a study is sophisticated enough to measure many variables suspected to affect outcome- to take one example, therapist confidence. Say that therapist confidence is equally lowered in BT and PP training. We still do not know whether the horse-race has been fair. Perhaps low confidence is no impediment in BT and

an enormous one in PP, or vice-versa. The biggest uncontrolled variable is spontaneity: perhaps some treatments are severely handicapped by being standardized at all. (Indeed, I reviewed one Norwegian study where the best results were achieved by one therapist whose cases had to be discarded because she was unable to conform to the protocol.)

My contention is that horse-race studies are effectively anecdotal, convincing in the way that clinical experience is, no more and no less. For years, I have argued, unsuccessfully, for naturalistic studies--what are respected therapists doing in their practices, and what are their outcomes? Process studies, a bit more popular, are also of interest. But outcome studies, however careful and scientomorphic, are lightweight in the advance of understanding of process.//

Peter D. Kramer

From seligman@cattell.psych.upenn.edu
Sat Jan 28 16:45:55 1995
To: psycho-pharm@netcom.com

Martin Seligman wrote [much omitted]:
>see above<

Peter Kramer writes:
>see above<

This may be true of some "horse-race" studies, which are not numerous, but calling outcome studies lightweight and suggesting that they are anecdotal, no more or less than clinical experience, is a vast overstatement.

Unlike anecdotes, the good outcome studies--and there are many-- usually have Large N's Random Assignment to treatment or control, reliable diagnosis, manualized treatment and training of therapists, objective assessment multiple measures of outcome, single blindness, control groups, often no treatment controlling for nonspecifics, occasionally placebo and other modality controlled long term follow-up.

If these are anecdotes, no better than clinical experience, our words have lost their meaning. It is also incorrect to claim that there is no dose-response relationship and no measure of process in such studies.

Not uncommonly relief is plotted against number of sessions: so for example in CT for unipolar depression much of the effect occurs in the first four weeks with less improvement per session over the next eight sessions--if memory serves--with diminished returns after twelve sessions. This is a dose-response

curve. Process is now commonly measured. This is what the entire mediator-moderator technology and literature is about.

marty seligman

From zuckerma@chopin.udel.edu
Fri Jan 27 21:26:19 1995

Where can I find a good description of CBT for panic disorder? What is the most comprehensive study suggesting that CBT for panic is as or more effective than drugs? Marvin Zuckerman

From gdaviso@rcf.usc.edu Sat
Jan 28 00:17:30 1995

Marv,
A good description is in second edition of Dave Barlow's Handbook of Psychological Disorders (Guilford Press). Of course his Graywind Publications puts out very detailed therapy manuals and videos.
Jerry Davison

from H_LEITEN@dewey.uvm.edu
Fri Jan 27 23:57:35 1995
Subject: Don Klein and CBT for panic disorders

Don Klein's response to Dianne Chambless seems to miss one of the central characteristics of people who suffer from panic disorder and who engage in at least some escape/avoidance behavior.

He writes

"what I find flabbergasting is the complete discordance between the etiological theory of CBT (antecedent ingrained fearful long-term catastrophizing attitudes that produce an overlearned chronic autonomic overreaction to barely perceptible endogenous stimuli yielding a vicious circle that regularly culminates in horrific panics, although a thousand experiences have already demonstrated that the panic is not dangerous)".

He also states

"It should also be pointed out that no evidence that I know of shows that attitude change precedes panic remission"

In fact, people with panic disorder, most of whom also engage in at least some escape/avoidance behavior, remain convinced that anxiety symptoms put them in danger of dying, or going crazy, or making a spectacle because they believe their escape or avoidance behavior saved them. What exposure based CBT does is persuade many patients that their predictions are wrong, that they are not realized even when the person does not escape or avoid anxiety provoking situations or sensations. Also, as the literature has long demonstrated, attitude change often occurs hand-in-hand or subsequent to behavioral change, which is not at all contrary to behavioral theory.

Harold Leitenberg Dept of Psychology University of Vermont Burlington, VT 05405 Phone: 802-656-2661 Fax:802-656-8783

From seligman@cattell.psych.upenn.edu Fri Jan 27 22:29:56 1995 Subject: The Dodo Bird

Dear Gene:

I am answering your query in two ways. First with my own view and secondly by cc'ing this to SSCPNET, which includes the world's leading psychotherapy outcome researchers--I will let them write you as well.

SSCPnet'ers--the background:

Gene Stone is a serious journalist doing a set of articles comparing therapies for Esquire, GQ, and Smart Money. His credentials are impeccable, by the way. He has sent me a letter about the Dodo Bird hypothesis (Miller and Luborsky, 1993):

M & L once again assert that all of their studies produce the conclusion that "nonsignificant differences predominate when different forms of psychotherapy are compared. psychodynamic therapy is in general no better or worse in its benefits than other therapies."

Gene goes on to ask "Can you help me by providing me with some data and writing that backs your argument?" My argument being that psychodynamic psychotherapy is not as effective in MANY circumstances as other modalities.

Here is my view:

The joker in the pack of the Dodo bird hypothesis is duration of treatment. Psychodynamic psychotherapy (PP henceforward) so claim its adherents requires many sessions to be effective. It has therefore not been tested in the

wealth of studies that compare various other modalities--all short term therapies--for a variety of disorders.

It would not make sense to compare 100 sessions of PP to 10-20 sessions of cognitive therapy or behavior therapy in its effectiveness. So it has been documented that

- 1) CT and IPT given over 10-20 sessions break up depression about as well as imipramine and better than placebo.
- 2) CT cures panic at least as well as medications and better than placebo. Note again--4-15 sessions.
- 3) BT (e.g. Masters and Johnson) breaks up impotence, frigidity, ejaculatory problems in 10-30 hours of treatment. These are problems of long-standing duration.
- 4) BT (implosion and desensitization) break up simple phobias in 10-20 sessions.
- 5) BT works at least as well as medications in OCD--again in about 20 sessions or less.

These five points comprise several hundred studies in the outcome literature--references in WYCC&WYC. Now in none, or almost none, of these studies is there a PP control. And that's because no one--Luborsky included-- would contend that PP would touch these disorders in 4-20 sessions. So the evidence is quite overwhelming that CT and BT work better, by far, for depression, panic, OCD, sexual problems, and phobias than PP. By concession on the part of PP, not by test.

There are a handful of very expensive long term PP versus other modalities in the literature. It is these studies that Luborsky's argument presumably rests on. These do not bear on the wealth of short term studies for OCD, panic, depression, phobia, and sexual problems that CT and BT break up well inside the envelope of time that PP takes to begin working.

I would ask SSCPnetters to chime in here. Especially with their analyses of the long-term literature.

In the meantime, Gene, I hope this is useful.
best wishes,

Marty Seligman

To: Dr. Seligman
Sun Jan 29 20:10 1995
From: kaechele@sip.medizin.uni-ulm.de

Dr. Seligman informs the journalist that Psychodynamic Psychotherapy consistently last 10 times as much as the cognitive-behavioral variety. This is an assumption based on the self-proclaimed propaganda of PP therapists:

<The joker in the pack of the Dodo bird hypothesis is duration of treatment. Psychodynamic psychotherapy (PP henceforward) so claim its adherents requires many sessions to be effective. It has therefore not been tested in the wealth of studies that compare various other modalities--all short term therapies--for a variety of disorders.

It would not make sense to compare 100 sessions of PP to 10-20 sessions of cognitive therapy or behavior therapy in its effectiveness.>

Studying the duration of treatment length on a large out-patient sample of a true diversity of treatment modalities (N > 1800) I could not find a systematic difference in treatment length among psychodynamic oriented or behavior oriented or system-family oriented interventions.

Graphically it is easily to demonstrate that treatment duration (measured by session numbers or weeks in treatment) is a continuous, negatively accelerated distribution, in fact replicating Howard et (1986) findings for a much longer period of treatment time; there are no systematic breaks for such things as a short term or long term treatments - these are fashionable shorthand expression of therapists.

The study has been published:

Kächele H (1990) Wie lange dauert Psychotherapie?
Psychother Psychosom Med Psychol 40:148-151

The sample has also been used in

Vessey JT, Howard KI, Lueger RJ, Kaechele H, Mergenthaler E (1994)
The Clinician's Illusion and the Psychotherapy Practice: An Application
of Stochastic Modeling. J Couns Clin Psychol 62: 679-685

I am ready to send a reprint to Dr. Seligman or to provide to special introduction to the journalist why German psychotherapists working under conditions of real practice are more similar than different with regard to the time they need to assist their patients to change.

Horst Kaechele, MD
University of Ulm /FRG

From seligman@cattell.psych.upenn.edu Sun Jan 29 21:14:34 1995 Subject: Re: The Dodo Bird - duration of treatment To: kaechele@sip.medizin.uni-ulm.de (Kaechele, Horst Prof. Dr.) Cc: psycho-pharm@netcom.com, psycho-therapy@netcom.com, sscpnet@bailey.psych.nwu.edu, Luborsky@landru.cpr.upenn.edu
--

The fact that therapists of different orientation spend similar amounts of time with patients is interesting--certainly economically interesting--but way off my main point.

My main point is that no PP that I know would claim to get effective treatment of depression, panic, sexual dysfunction, OCD, marital problems, or phobia in 4-20 sessions. Rather effective PP would take many times that number of sessions--if it worked at all.

Yet there are repeated outcome studies showing effective CT and BT for all these disorders in under twenty sessions.

marty seligman

Therapy for Therapists ?

letters from sscpnet 2/95

Am 28. Februar 1995 um 2.42 pm geriet eine Frage in die SSCPNET-Diskussion, die bei uns entweder schon lange keine Frage mehr ist, oder so noch gar nicht gestellt wurde. Zunächst gab es ein lebendiges Hin- und Her, in dem sich persönliche und berufliche Aspekte sehr lebendig vermischten. Um 20.30 erschien die fortläufig letzte Äußerung zu diesem Thema; in dem folgenden Tagen tauchte sie dann und wann wieder auf. Vielleicht ist es anregend, anhand dieser Diskussion, einen Ein-Blick in die Möglichkeiten der Internet-Kommunikation zu nehmen?

From huppert@acs.bu.edu
Tue Feb 28 02:42:37 1995
To: sscpnet@BAILEY.PSYCH.NWU.EDU²

I have recently been interviewing for practica for next year and have been asked about whether I am in therapy myself by a few places. I was just wondering what thoughts there were out there on the subject of whether it is:
preferred - suggested unimportant - irrelevant
according to the people on the net.

Jonathan Huppert

From sechrest@aruba.ccit.arizona.edu
Tue Feb 28 03:53:13 1995

Outrageous!
Lee Sechrest, Univ. of Arizona

From follette@scs.unr.edu
Tue Feb 28 07:17:00 1995

I've always thought this was one of those unsupported (and unsupportable) heuristic errors internships make-thinking that having been in therapy makes one a good or better therapist. I know of no data to support such speculation. I think that if internships actually respected the notion of confidentiality (which it looks like they don't), they should respect the answer "It's none of your business."

I'm sure someone out there will write in with an anecdotal story about how they found that interns who had been in therapy were so much more empathetic than those who hadn't been in therapy. Ask them if they would be willing to bet their diplomas on whether they could view blind therapy tapes of interns doing therapy and then guess (better than chance) who had been in therapy and who had not.

Bill Follette

From njacob@u.washington.edu

²immer wenn die Empfänger direkt das SSCP-Net ist, wird die Adresse weggelassen

Tue Feb 28 08:32:21 1995
To: follette@scs.unr.edu

Dear Bill,

You are far too cynical about the worthiness of therapy for potential therapists. I have found my support group for ex-APA members extremely valuable, and believe that all prospective psychologists should have to attend at least one meeting (My name is Bill, "Hi Bill" and I'm an APA member). Moreover, I never would have had the courage to share such an intimate experience with thousands of readers of this internet were it not for the support group. Should we require it of our graduate students? Why not? Should people be "on drugs" themselves before being granted prescription privileges? Absolutely.

Neil Jacobson. University of Washington

From kihlstrm@minerva.cis.yale.edu
Tue Feb 28 13:27:11 1995

Any internship site that asks candidates whether they have been in therapy themselves should immediately lose its accreditation.

John F. Kihlstrom,
Professor Department of Psychology, Yale University

From jhart+@pitt.edu
Tue Feb 28 14:05:25 1995

Many graduate programs in clinical psychology also ask similar questions of their PhD applicants. I would be curious to know how many programs represented on this network plan on asking about previous personal therapy in upcoming interviews of such applicants.

Jon Hart (University of Pittsburgh)

From BLANCHARD@medcolpa.edu
Tue Feb 28 14:29:55 1995

I too consider this information irrelevant. The inquiry is at best a violation of privacy.

I would also like to point out that Jonathan is interviewing for *practica* -- not an internship (of course, such a question is equally inappropriate at an internship). Should a graduate program that allows its students to participate in such practica lose its accreditation (practica are not accredited)? My point is that Jonathan's plight raises questions about the quality/ethics of clinical training sites as well as the graduate programs that support such sites with their students.

As graduate programs send their students to community placements so as to 1) provide a greater diversity of clinical training, 2) decrease the supervisory chores of faculty, or 3) provide financial support to students from sources other than the department, I think that those programs must acknowledge their own responsibility in perpetuating training standards that are less than ideal.

Jack Blanchard Medical College of Pennsylvania/EPPI

From kihlstrm@minerva.cis.yale.edu
Tue Feb 28 14:52:25 1995

Practicum sites should not ask this question. Those that do should lose their status. That is, they should no longer be used as approved practicum sites. Internship sites should not ask this question. Those that do should lose their accreditation.

It's that simple.

John F. Kihlstrom, Professor Department of Psychology, Yale University

From BLANCHARD@medcolpa.edu
Tue Feb 28 16:17:59 1995

John Kihlstrom suggested that if a practicum site asks about therapy received by a student this site should no longer be used as an approved site. He also suggested that internships and graduate programs that ask should lose accreditation -- "It's that simple". Agreed.

Aside from awaiting flagrant behavior that may come to the attention of a graduate program how do programs *responsibly* engage the services of community training sites? I wonder what methods programs have for selecting sites other than their availability or minimum experience with that site in the past ("we had a student there a few years back").

My experience with graduate programs (at least in the NE) suggests that there is little or no monitoring of practica by graduate programs. Rather than moan about the violations occurring elsewhere graduate programs might look at their role in such problems. Something that has been missed in this discussion is the question of where did the clinicians at practica and internships receive their training regarding the relevance of such things as projective testing, whether or not a clinician is in therapy, etc.? This is not to suggest that graduate programs are solely responsible -- incompetent or unethical training sites are also culpable.

However, it seems too easy to say that problems in clinical training somehow do not arise until students leave the pristine environment of graduate school.

Jack Blanchard Medical College of Pennsylvania/EPPI

From BOBAPPIC@aol.com
Tue Feb 28 16:30:51 1995

I would be interested to hear from anyone supportive of or tolerant of this practice (inquiring about current/past personal therapy for applicants to internships, grad programs, or practica). What is the rationale for such inquiry? I can guess, but I may be wrong. This is, I realize, a tough house in which to take such a stand - but it might help.

Bob Klepac

From: bsmith+@pitt.edu
Tue Feb 28 16:53:53 1995

To buck the trend on the SSCPnet, I propose that in some cases, it may be valid and appropriate to ask about prior therapy before allowing someone to enter a psychotherapy training program.

For instance, if the person is in therapy to deal with the impulse to have sex with prepubescent children, I think it may be relevant to know about it if I were training people to work with children.

I would also want to know if someone had been in treatment for sexual compulsion or some other behavioral disorder that might compromise ethical behavior with clients.

At some point the obligation to protect clients supercedes the applicant's right to privacy. Thus, certain therapy experiences might screen people out of training opportunities.

Brad Smith

From: k-howard@nwu.edu
Tue Feb 28 17:16:39 1995

Bill Follette suggests an interesting study, as usual. In our clinic, however, every intern is or has been in therapy. Does someone actually know a real (nonbehavioral) therapist who has never been in therapy, or in Europe? More importantly, does anyone know a therapist who does not need therapy as we speak?

Ken Howard Department of Psychology Northwestern University

From njacob@u.washington.edu
Tue Feb 28 17:16:23 1995
To k-howard@nwu.edu>

Ken,
I don't know about therapy, but I do think most therapists need to be on drugs.
Neil

From kwilson@scs.unr.edu
Tue Feb 28 17:12:41 1995

Ken,
Gosh, I hope all of your interns appreciated you self-disclosing for them.
Kelly Wilson University of Nevada

From stickle@aruba.ccit.arizona.edu
Tue Feb 28 17:18:39 1995
To bsmith+@pitt.edu

Brad,

This seems off the mark to me. The data on treating sex offenders suggests that efficacy is extremely poor. This seems to me another example of false confidence because someone has been in the hands of a therapist, regardless of what the data are.

Tim Stickle Department of Psychology University of Arizona

From bsmith+@pitt.edu
Tue, 28 Feb 1995
To stickle@aruba.ccit.arizona.edu

Tim

I believe that you understand my point completely. If someone has been in treatment for these problems they should not be allowed anywhere near a child or patient. I have no illusions that sex offenders are cured by the types of therapy currently employed.

Nevertheless, the Arizona Board of Psychological Examiners recommends therapy for licensed psychologists caught having sex with clients.

Then, more often or not, the "rehabilitated therapists" are allowed to resume practice.

Brad

From kaechele@sip.medizin.uni-ulm.de
Tue Feb 28 18:02:48 1995

In Germany psychodynamic as well as behavioral training regulations have agreed on mandatory "self experience" called training analysis among analysts, training therapy" among psychodynamic psychotherapy people and behavioral training therapy- mainly group sessions - among behaviorist.

Lasting from 150 sessions to eternity

One is not in therapy- mind you - if you are in therapy you are very likely to be turned down for training (they tell you: come back when it is over) you asking for being in therapy is not part of this game We also have no data no this but everyone - including the very tough behaviour therapists agree with the sheer necessity
how come

horst kaechele ulm university, FRG

From kihlstrm@minerva.cis.yale.edu
Tue Feb 28 18:17:13 1995

It's one thing to ask about criminal convictions, it is another thing entirely to ask about therapy.

John F. Kihlstrom, Professor

From gdavis@alnitak.usc.edu
Tue Feb 28 18:37:55 1995

Colleagues,
Hope we can keep two issues separate:

(1) Do graduate, practicum, or internship programs have the right to ask applicants if they are or have been in therapy? (My answer is a firm No, for the reasons already stated on the network. I have always objected also to the requirement that at least used to be made by some programs that applicants submit MMPI profiles.)

(2) Might it be a good idea for people training to be therapists to have been or to be in therapy themselves? (My answer to this is Yes, and my opinion is *not* based on data, at least any data that I am aware of. I associate my leap of faith with the other working assumptions we operate with in our developing science of therapeutic change.)

I hope that my answer to (2) will not result in my email address being dropped from this network.

Jerry Davison

From CAULFIELD.MARIE_B@boston.va.gov
Tue Feb 28 18:44:50 1995

Jerry Davison wrote:

"Might it be a good idea for people training to be therapists to have been or to be in therapy themselves?"
(to which he answered yes)

Dr. Davison: To clarify, do I read you correctly as saying that perhaps all therapists-in-training should have their own therapy experience?

Regarding the other question about asking applicants about their own therapy experience (which I didn't do when interviewing graduate student applicants and I don't do now when interviewing intern applicants): The perplexing thing about this is that some programs ask because they want to rule out pathology, and some programs ask because they want to "rule in" openness to self-exploration.

From stickle@aruba.ccit.arizona.edu
Tue Feb 28 19:11:05 1995
To: Brad Smith <bsmith+@pitt.edu>

Boy that license sure can make a person feel secure about entrusting a treatment provider.

Tim

From plaud@badlands.nodak.edu
Tue Feb 28 19:32:36 1995
To: CAULFIELD.MARIE_B@boston.va.gov, From:

Reply:

I have come up with a comprehensive answer to this question about therapy/personal therapy and training. Here is the comprehensive criteria for appropriate graduate training in this regard, after subjecting my list to exhaustive principal components analysis:

- (1) In order to be a good graduate student you should have a history of operating on your environment (death therapy is therefore ruled out);
- (2) You must have the verbal/covert repertoire to understand what (1) above means.

Essential ingredients of successful graduate experiences solved.

Joseph J. Plaud, Ph.D. University of North Dakota

From bsmith+@pitt.edu Tue Feb 28 19:46:33 1995 To: stickle@aruba.ccit.arizona.edu

Tim

I believe that you understand my point completely. If someone has been in treatment for these problems they should not be allowed anywhere near a child or patient. I have no illusions that sex offenders are cured by the types of therapy currently employed. Nevertheless, the Arizona Board of Psychological Examiners recommends therapy for licensed psychologists caught having sex with clients. Then, more often or not, the "rehabilitated therapists" are allowed to resume practice.

Brad

From follette@scs.unr.edu Tue Feb 28 19:50:40 1995 To: jhart+@pitt.edu>

Jonathan,
University of Nevada, Reno never has and never will.
Bill Follette

From follette@scs.unr.edu Tue Feb 28 19:53:41 1995 To: njacob@u.washington.edu
--

Neil-

Perhaps I should rethink my position. Now that you bring it up, I can think of several notable therapy successes that may justify insisting upon therapy. For example, your experience is certainly heart-warming. I know that Joe Plaud and I have benefitted from social skills training; Bob Carson is making great progress on his Tourette's problem; and I guess I can disclose that the group sharing you, I, Don Klein, John Kihlstrom, and Lee Sechrest do to work on our low self-esteem problems seems to showing some positive effects though we have a long way to go. So, yes, you are right.

Bill Follette

From follette@scs.unr.edu
Tue Feb 28 20:22:51 1995
To: blanchard@medcolpa.edu

Jack

Good points. Not wanting to put words into anyone's mouth, I would presume the source of the justification for a question about one's personal experience with therapy stems from the psychoanalytic training models where the therapist has to work thru their own "transference and countertransference issues" before they can do "competent" therapy. Even the analytic training models makes the supervision process therapy like thus raising additional ethical issues.

Bill Follette

From follette@scs.unr.edu
Tue Feb 28 20:30:41 1995
To: k-howard@nwu.edu

Ken

Don't think I didn't notice that you don't count behavior therapists as "real therapists." Go stick your finger in a wall-socket for 3 seconds for that one. Behaviorally yours,

Bill

From k-howard@nwu.edu
Fri Mar 3 00:36:09 1995

Neil Jacobson has appointed me to be his PERSONAL TRAINER. Please forward any requests regarding goals and outcome assessment to me. I plan to

start with paper training although he may have more talent to heel. Does anyone recommend aversiveness for such an adverse character?

Ken Howard Department of Psychology Northwestern University

From beutler@edstar.gse.ucsb.EDU Fri Mar 3 22:49:49 1995 To: gdaviso@alnitak.usc.edu
--

I agree with Jerry.

The issue of therapy for therapists is a potentially important one and one that has been subjected to empirical investigation. Some of that work is reviewed in the HANDBOOK OF PSYCHOTHERAPY AND BEHAVIOR CHANGE (Bergin and Garfield, 3rd and 4th eds) under Therapist Variables.

Our conclusions from the empirical literature was that there is some evidence that therapists may get poorer effects during the time they are in personal therapy themselves. There is no good evidence of either positive or negative effects in the long run, but some impressions of those who have done some of the best work in this area (e.g. Bergin) that long term effects are positive.

Might the fact that these findings exist warrant an internship center asking about whether one is in or has been in psychotherapy? Might it be a bit like asking if they have had a course on ethics--a course for which there is no evidence that it makes one a better or a worse therapist, but there is some logical rationale for its value in the internship environment?

Larry Beutler

From MIKLOWITZ@clipr.Colorado.EDU Sat Mar 4 03:06:07 1995
--

I've been following the therapist-in-therapy debate with interest. I, too, am opposed to internships that request this information of applicants (in fact, I was asked this question when I was on the job market in 1988, at an academic institution). I am, however, intrigued by the issue of whether being in therapy helps one to become a better therapist. This issue has been discussed primarily in terms of "dealing with one's issues" so that these do not interfere with one's treatment of patients. But what about another possible benefit that has been alluded to in this discussion --learning to do psychotherapy by modeling? The trainee's therapist, if skilled, might continually model for the trainee appropriate responses to clinical dilemmas, much as a clinical

supervisor might. Needless to say, it could work the other way as well -- a therapist trainee could learn some bad habits from his or her own therapist. I'm not suggesting that being in therapy should be a requirement for becoming a therapist, only that it could provide an additional learning experience.
David Miklowitz (U of Colorado)

From MaxM2@aol.com Sat Mar 4 03:56:41 1995

Larry Beutler writes about the effect of therapists' therapy on their effectiveness as therapists (see above):

The problem that I see is that the fact that one has "been in psychotherapy" has virtually no meaning. It certainly cannot mean anything about the amount of useful help an applicant has received, because therapists and therapy are so variable in quality. "I have been in psychotherapy" could mean anything from "I had a disastrous three month experience with a therapist who ended up molesting me after which I was an emotional basket case" to "I had three years of therapy which changed my life in every way for the better" to everything in between. Interviewers could go beyond a simple "therapy: yes/no" question and ask questions like: "What did you get out of therapy?", but it seems to me that: 1) It would be difficult to get more out of the answer than one gets from any other question that requires introspection and thoughtfulness;

2) It is unlikely that anyone is going to say: "Yes, I was in therapy and it was a complete waste of time." The tendency would be to find something meaningful to say about it, possibly distorting the actual meaning of the therapy for the applicant;

3) Most of us probably know someone who has been to a therapist whom we know to be borderline incompetent or worse, and the client reports a very good experience. In some of these cases, an alliance may form which provides actual benefits (although having nothing to do with the therapist's skillfulness), but in some cases I think people are just uneducated consumers of therapy, and think that because someone devotes an hour/week listening to them and nodding, they've gotten a good deal. I would guess that doctoral candidates are not immune to this.

Max Molinaro

From: ART HOUTS University of Memphis Tue, 2 May 1995 To: njacob@u.washington.edu

Neil

Speaking of reading the literature, Stein and Lambert have a review on the training is/is not related to outcome in the latest number of JCCP. The power of the analysis is low, and they conclude that evidence favors training with small effects ($d's < .3$). They get bigger effects in favor of training when the measure is drop out from treatment. I have a number of questions about this review, but it does put forward claims that training matters, which is different from previous reviews.

Art

From: njacob@u.washington.edu Tue, 2 May 1995 10:19:05 To: ART HOUTS University of Memphis
--

This is only the latest in a number of attempts to undo by selective review and sleight of hand the overwhelming body of evidence suggesting that neither training nor experience make any difference in outcome. This literature, summarized in my Psych Science article with Andy Christensen, has been meta-analyzed to death, gone over with a fine tooth comb. In fact, there is no area in psychology where there is stronger support for the null hypothesis: on the average, neither amount of therapist experience nor degree of professional training effects outcome.

Neil

Child-Adult Links

HK's letters from sscpnet 3/95

Date: Sat, 19 Aug 1995 11:59:53 -0400
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)

For a century, we have labored under the dogma that the pain comes from bad events in childhood. The data is simply not bearing this out, and there are a variety of alternate theories: personality structure, genetics, recent events, conscious cognitions, ongoing behaviors, core-conflictual relationship themes, etc. I don't know which theory is correct, but it is time to become open-minded about the possibility that childhood events are relatively impotent in adult life.

I claimed that the major events of childhood--physical and sexual abuse included--have not been demonstrated to have clear causal links to ADULT psychopathology. This is a follow-up to the discussion that ensued on the net.

There is now a large, and well-executed literature, involving large numbers of subjects, careful measurement, and most importantly, direct as well as statistical controls for the genetics of adult psychopathology. It finds huge effects of genetics, large effects of recent life events, large effects of personality, and small or no effects of childhood events. All of us who think that traumatic experience is important in human growth--and I have devoted most of my research life to this issue--MUST read this literature and discuss it.

If you think you can dismiss this literature out of hand because you somehow KNOW that childhood events cause adult psychopathology, you are simply wrong. I couldn't dismiss it and I would like to be able to. The references to it are found in chapter 14 and its notes of What You Can Change and What You Can't (Ballantine, 1994).

There have been at least five responses to the failure to find any clear causal links from childhood events to adult psychopathology that you have raised, and I want to comment on each.

1) Failure of measurement. This is simply a null finding that results from insensitive measurement of childhood events. It's possible, but genetics, recent life events, and personality measures--all imperfectly measured--produce substantial causal paths to, for example, adult depression. Since it is the retrospective report of childhood trauma by adults that convinces so many clinicians of its causal effects, why does this same report have little or no causal pathway, once factors like genetics, recent life events, and personality are partialled out?

2) Ignore the pain. Hardly. Our clients come to us in pain. There are a variety of therapies that work well on adult depression, anxiety, and relationship difficulties. We must continue to do our best administering these. This is not a dispute about neglecting patient's pain--it is a dispute about WHERE that pain

comes from. For a century, we have labored under the dogma that the pain comes from bad events in childhood. The data is simply not bearing this out, and there are a variety of alternate theories: personality structure, genetics, recent events, conscious cognitions, ongoing behaviors, core-conflictual relationship themes, etc. I don't know which theory is correct, but it is time to become open-minded about the possibility that childhood events are relatively impotent in adult life.

3) Bring on the Meds. Hardly. That there are strong genetic and biochemical causes of various forms of adult psychopathology does not mandate medication. Even Ivan Goldberg would probably agree with this. Medication sometimes works, but there is equally strong evidence that a variety of psychotherapies work quite well on most of the problems which have genetic and neurochemical substrata. Indeed, that is a major point of WYCC&WYC.

4) Explanatory style in childhood. I have argued in *Learned Optimism* and just now in *The Optimistic Child* (Houghton Mifflin, 1995) that children learn their explanatory style and that this carries over into adulthood. Does this contradict the findings that childhood EVENTS do not seem to have major causal effects on ADULT psychopathology? I don't think so, but this is a very acute objection. I can give you my speculation, however:

What I believe is learned in childhood that abides is generally SKILLS and their absence. These get repeatedly practiced and become strong habits. Pessimism, good conversation, much of our sexuality, assertiveness, shyness, etc., etc, reflect skills and their absence. Events, such as abuse, can and do influence and retard a whole variety of skills, and it is these that I believe carry over and are the causal mediators. No one has tried to measure this. As therapists, I speculate, that much of our job has to do with skill learning. There are any number of good therapeutic techniques that build trust and love and allow skill learning to proceed with our help. Finding a common enemy, like the perp, sympathetically reviewing childhood, listening well to ongoing problems are all examples of techniques that build a good therapeutic relations. But we should not mistake the main technique--coming up with a credible causal account for the pain that is present that both therapist and client agree on--that builds a therapeutic alliance, for what the actual causes are.

5) I don't care what the data show. I KNOW that childhood events cause adult psychopathology. No empirical data could convince me otherwise.

To those who believe this, I have little to say. My prejudice is that the mark of a good psychologist and the crucial ingredient of a good therapist is openness--openness to new data and new theory and new experience that might change what we "know". I suggest that there is now a serious challenge to the myth of childhood, and it is incumbent on all of us to READ about it, argue about it as we are doing, and only after all that work, make up our minds.

Martin Seligman

From: mbailey@bailey.psych.nwu.edu (Mike Bailey) Date: Sat, 19 Aug 1995 16:10:57 -0500 To: sscpnet@bailey.psych.nwu.edu

Re Marty Seligman's interesting comments about the (un)importance of childhood for adult psychopathology:

I mostly agree, and it's a very important topic for discussion. Faith in the importance of childhood remains both a culture-wide and discipline-wide phenomenon.

The best established surprising fact comes from the study of shared environment, that is the environment that is commonly experienced by siblings reared together. This could include, for example, parental childrearing philosophies, SES, religion, parental psychopathology, etc. The literature, ably reviewed in David Rowe's recent book "The Myth of Experience" (I believe), shows that contrary to conventional wisdom, the shared environment has little influence on most traits. The most direct indicator of this is that the correlation between genetically unrelated children reared together is typically very near zero. (IQ is a possible exception. Unrelated siblings are correlated for IQ as children, though as adults this correlation appears to diminish to zero.)

In contrast, nonshared environmental influences, those aspects of the environment not shared even by identical twins reared together, appear to be substantial. These could be idiosyncratic childhood events, but they could just as easily be random biological developmental influences. Behavior genetics has not been very helpful in illuminating specific mechanisms of development.

One more thing. There has been at least one twin study of attributional style, and my memory is that its results were similar to those for other traits: Heritability near 50%, and the bulk of the remaining variance due to nonshared environmental influences.

Michael Bailey
Department of Psychology
Northwestern University
Evanston, IL 60208
Phone: 708-491-7429
jm-bailey@nwu.edu

SO LASSEN ?????????

From: fresco@cattell.psych.upenn.edu (David M. Fresco)
Date: Sat, 19 Aug 1995 17:47:44 -0400
To: mbailey@bailey.psych.nwu.edu (Mike Bailey)

At 16:10 19/08/95, Mike Bailey wrote:

[big snip]

>One more thing. There has been at least one twin study of attributional style, and my memory is that its results were similar to those for other traits: Heritability near 50%, and the bulk of the remaining variance due to nonshared environmental influences.>

What is the cite for this twin study?

=David M. Fresco

```

Internet: fresco@cattell.psych.upenn.edu =
= Department of Psychology fresco@unc.edu =
= CB#3270, Davie Hall _o =
= Chapel Hill, NC 27599 \, =
= `,'(*) =
= Voice: (215) 204-6851 (*) . ./"""" =
= Fax: (215) 707-1557 """""" =
= Beeper: (215) 552-6012 =
= =
= Co-ListOwner of Helplessness helplessness-1@netcom.com =
-----

```

From: "Peter M. Barach" <pbarach@netcom.com>
Date: Sat, 19 Aug 1995 18:33:47 -0700 (PDT)
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>

Marty,
thanks for replying to those who commented on your position concerning links between childhood events and adult pathology. Perhaps the events themselves are not directly linked to adult pathology. However, where do "personality structure, recent events [such as revictimizations and chronically poor relationship choices], ongoing behaviors [ditto], core-conflictual relationship themes" come from? For some people who have experienced abuse (but apparently not for others), these alternate theories and schemas reflect their abuse experiences along with the contributions of genetics, etc. So it seems that some people have more profoundly negative effects from abuse than other people do. As an additional reason for saying that the methodology may have missed the boat on linking abuse to adult pathology, I'd like to suggest that some of the

studies may have some sample bias. I haven't gone back to the original studies to check this out and I'm sure you would not this since you reviewed the studies for your book--but aren't most researchers unlikely to have studied people who had extremely bad childhoods? That is, people with the most devastating childhoods may not have ended up in the research samples because they are too dysfunctional to cooperate with researchers. Also, some people might deny they were abused if the question was not phrased carefully or if there were no rapport with the examiner or researcher. If this is the case, the available data would seriously underestimate the correlation of abuse with adult pathology.

>What I believe is learned in childhood that abides is generally SKILLS and their absence. These get repeatedly practiced and become strong habits. Pessimism, good conversation, much of our sexuality, assertiveness, shyness, etc.,etc, reflect skills and their absence. Events, such as abuse, can and do influence and retard a whole variety of skills, and it is these that I believe carry over and are the causal mediators. No one has tried to measure this.>

OK, this makes sense. Nobody who's sensible is saying there is a 1.0 correlation between being abused and having pathology. And since abuse (especially the chronic variety) often takes place in a context of psychological neglect by dysfunctional, unskilled, and depressed adults, naturally there will be mega-deficits of basic skills.

Peter M. Barach, Ph.D.

5851 Pearl Road, Suite 305
Parma Heights, Ohio 44130 USA
Phone: Voice-216-8459011 (press 6 if you get voice mail) Fax-216-845-9013

From: plh@ix.netcom.com (Paul R. Lees-Haley, Ph.D.) Date: Sun, 20 Aug 1995 12:15:59 -0700 To: sscpnet@bailey.psych.nwu.edu
--

Mike,

I called Barnes and Noble in Los Angeles, and they say Books in Print list neither a book by the title Myth of Experience nor an author of any book by the name David Rowe. Your suggestion that the literature was ably reviewed by the book makes it sound quite interesting. Could you check on the citation, please? Thank you!

Paul

Thought for the week. . .
Replications of a biased design produce
even more misleading literature.

* Paul R. Lees-Haley, Ph.D., ABPP Telephone 818-981-0138 *
* 5445 Balboa Blvd., Suite 117 Fax 818-981-1715 *
* Encino (Los Angeles), California 91316 Email plh@ix.netcom.com *

This message is not sent in my professional capacity; it is individual.

Date: Sun, 20 Aug 1995 16:44:11 -0500 From: edonchin@s.psych.uiuc.edu To: williams@CATTELL.PSYCH.UPENN.EDU (David R. Williams), seligman@cattell20.psych.upenn.edu (Martin E. P. Seligman)

As I read this very interesting discussion, without the benefit of clinical training of any sort, I am left a bit puzzled. It is not clear from the debate whether a rather important distinction is observed.

It would seem those who argue for the decisive value of clinical observations as a source of support for the importance of childhood experience may be making one of two assertions;

1. Childhood experiences have a causal relationship to adult psychopathology,
2. It is useful to structure the therapeutic interaction as if childhood experiences have a causal relationship to psychopathology.

I assume, without a comprehensive knowledge of the literature, that much of the corpus that supports the importance of childhood experiences is based on the effectiveness of therapeutic interactions ("interventions," some of my friends would call them) that *assume* the truth of assertion (1) above and, therefore, structure the intervention around an examination, or a search for, or an attempt to confront, you name it, these childhood experiences.

It would seem that even if it would turn out that all, or most, or some, of these interventions are successful, such a result has no logical relationship to the validity of assertion (1). It simply does not follow that A indeed caused B, just because at some later date you could eliminate or ameliorate B by talking about A, especially if all you have is someone's memories of A.

This, of course, does not mean that you should not use this assumption if it provides a useful framework for structuring your interventions, or for structuring your internal models of what you were doing.

>From a scientific perspective the validity of of assertion (1) can only be tested within a program that obtains independent, reliable, data on actual childhood experiences, relating these to observed psychopathology, and

having solid data about the base rates of the various experiences and the conditions under which they appear to to be causal of the psychopathology.

Yet, practitioners may validly argue that they find useful the belief in the validity of assertion (1), but that all they are really claiming is that assertion (2) is true.

Of course, this leads to the challenge of finding empirical means of validating assertion (2)...but this belongs to another thread.

My point is that one may reach useful, if erroneous, conclusions about the world and as long as utility is the only issue, then the outcomes are the only thing that matters. Drawing general, theoretical, conclusions is another matter altogether.

To give a homely example. I *know* for a fact, based on many observations of my own experience, that it very likely to rain in London if I am there and if I leave my umbrella in the hotel.
So, the safest thing to do is to carry my umbrella when I go out, which I always do.

This, actually, turns out to be useful because whenever it does rain I have the umbrella and I am not soaked.

So, based on lots of direct clinical observations of my interactions with London I have developed a useful practice. I would, however, hesitate to consider this a proper basis for speculating about the meteorology of the British Isles...

Manny Donchin
Department of Psychology University of Illinois at Urbana-Champaign

Date: Sun, 20 Aug 1995 16:49:54 -0400
From: williams@cattell.psych.upenn.edu (David R. Williams)
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>

Marty,

Shibboleths must fall, but I'm not sure the literature can lead us to ignore childhood abuse and neglect as contributing factors of significant, if not mythic, proportions. It seems to me there is a common kind of caution expressed in different terms in each of the disciplines we share:

- a) learning theory: the "learning-performance" distinction
- b) psychopathology: the "diathesis-stress" distinction

In other words, sometimes you get more--in behavioral possibility-- than what you see in behavioral observation. As you know well, the learning-performance difference was often hard to take into account experimentally. Extended to the case of psychopathology, you have to worry about "as the twig is bent" effects on personality--factors that put people at risk of "being less than they could be". Such factors are hard to assess (there is, by definition, no track record of prior performance to compare with), and convenient to neglect, but necessary to take into account if we want to make psychotherapy fully contributive.

The correlative caution with the "skills" approach is solidly rooted in "learning/performance": whatever skills there are, there is still an organism, a person, who chooses which skills to use, when and how. That's a meta-skill, of course, necessarily acquired on a basis different from the events of the skill-learning itself. Yet it is the exercise of choice at the meta-level that provides the main focus for many people interested in providing, or receiving, psychotherapy. Family and religion, culture and tribe, impact the way a person construes the passing scene, and inspire dispositional decisions that are prior to the deployment of skills. The acquisition of such dispositions most probably depends upon more significant aspects of human relationships than the behaviorally-based "training" literature acknowledges. Unlike regular, actual skills, these are cognitive acts that provide the appearance of "freedom of choice" and "dignity" that the curmudgeonly Skinner could not abide: attributes of character and personality that for Skinnerian reasons do not appeal to the payers of third-party payments, either.

It is easier to affirm something from a literature than to deny it; perhaps, Campbell-like, we should acknowledge myth as such, but affirm interest in the aspects of human experience to which it points.

--david

Date: Sun, 20 Aug 1995 19:59:59 -0400 (EDT)

From: John F Kihlstrom <kihlstrm@minerva.cis.yale.edu>
To: Mike Bailey <mibailey@bailey.psych.nwu.edu>

On Sun, 20 Aug 1995, Mike Bailey wrote:

> Speaking of measurement problems, does it trouble anyone else there that the vast majority of studies linking childhood sexual abuse and adult psychopathology are done retrospectively, on adults? <

No, Mike, apparently it doesn't trouble anyone, and that should trouble those of us who have tried to teach budding clinicians how to compute correlations and make causal inferences. Something went terribly wrong somewhere.

John F. Kihlstrom, Professor
Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205
Telephone (203) 432-2596 Facsimile (203) 432-7172

Date: Mon, 21 Aug 1995 06:18:12 -0400
From: fresco@cattell.psych.upenn.edu (David M. Fresco)
To: mibailey@bailey.psych.nwu.edu (Mike Bailey)

At 17:36 20/08/95, Mike Bailey wrote:

Andrew Leeds writes:

> Since: 1) it is clear ADULT trauma CAUSES ADULT psychopathology; and 2) a vast body of literature demonstrates that CHILDHOOD environmental factors influence ADULT functioning;

Neither 1 nor 2 is all that clear.

True, PTSD and depression can result from severe adult trauma, but there is little evidence that such trauma creates a psychological liability to further depression. A recent analysis by Kendler showed that recurrence in depression, in his sample, appeared only to be due to genes. In other words, vulnerability to depression in his sample was genetic. I stress "in his sample" because it was just one study, albeit a very good one. Based on these observations, I'd make the following prediction: Childhood trauma may cause childhood psychopathology. But if it doesn't, it probably won't cause delayed psychopathology in adulthood.

>Speaking of measurement problems, does it trouble anyone else there that the vast majority of studies linking childhood sexual abuse and adult psychopathology are done retrospectively, on adults? <

Additionally, THAT childhood trauma has been implicated in SO many types of adult psychopathology seems to support Seligman's speculation #4 in his original post:

>4) Explanatory style in childhood. I have argued in >Learned Optimism and just now in The Optimistic Child (Houghton Mifflin, 1995) that children learn their explanatory style and that this carries over into adulthood. Does this contradict the findings that childhood EVENTS do not seem to have major causal effects on ADULT psychopathology? <

I don't think so, but this is a very acute objection. I can >give you my speculation, however:

What I believe is learned in childhood that abides is generally SKILLS and their absence. These get repeatedly practiced and become strong habits. Pessimism, good conversation, much of our sexuality, assertiveness, shyness, etc.,etc, reflect skills and their absence. Events, such as abuse, can and do influence and retard a whole variety of skills, and it is these that I believe carry over and are the causal mediators.

No one has tried to measure this.

The almost total lack of specificity between abuse and certain types of adult psychopathology seems to relegate abuse to, AT BEST, a distal contributory cause of adult psychopathology--perhaps having a more direct effect on the emergence of maladaptive styles of coping to the exclusion of adaptive life skills.

More of Seligman's #4

> As therapists, I speculate, that much of our job has to >do with skill learning. There are any number of good >therapeutic techniques that build trust and love and allow skill >learning to proceed with our help. Finding a common enemy, >like the perp, sympathetically reviewing childhood, listening >well to ongoing problems are all examples of techniques that >build a good therapeutic relations. But we should not mistake >the main technique--coming up with a credible causal account >for the pain that is present that both therapist and client agree >on--that builds a therapeutic alliance, for what the actual >causes are. <

I also think that Seligman provided us with a better approach for handling the abuse history of adult clients in the first paragraph I cited. Yes, allying with the client against the perpetrator will do wonders for attendance in therapy. Still, the acknowledgment of abuse does not prescribe a course of action. Rather, the skilled clinician must still conduct a careful idiographic assessment of behavioral strengths and weaknesses. No uncovering work will spontaneously impart the 'here-and-now' skills needed to function as an adult. Rather, these skills now need to be acquired--mindful of the fact that there acquisition may have been hindered by the emergence of coping styles that eased the pain of abuse (at the time), but are relatively maladaptive in one's adulthood.

David M. Fresco

```

-----
= David M. Fresco      Internet: fresco@cattell.psych.upenn.edu =
= Department of Psychology      fresco@unc.edu      =
= CB#3270, Davie Hall      _o      =
= Chapel Hill, NC 27599      \<,      =
=                               ` ,/ (*)      =
= Voice: (215) 204-6851      (*) . ./ "" ""      =
= Fax: (215) 707-1557      "" ""      =
= Beeper: (215) 552-6012      =
=                               =
= Co-ListOwner of Helplessness      helplessness-l@netcom.com      =
-----

```

Date: Mon, 21 Aug 1995 07:20:07 -0400 (EDT) From: Gregory Paveza AMH <paveza@hal.fmhi.usf.edu> To: mbailey@bailey.psych.nwu.edu (Mike Bailey)

Michael,

If you go back to my post of Sunday, this is exactly my point. In those few types of psychopathologies where some links have apparently been established, we are usually dealing with small N's, and the studies are retrospective. Moreover they are often clinically based and controls are often poor. Moreover, the likelihood of recall bias is magnified as well. As an epidemiologist, I have tremendous respect for well designed retrospective studies in all areas, including in psychiatry and psychology, social work etc. However, I must say that many of the retrospective studies I've looked at and reviewed are often not well designed, and this only compounds the problems.

So yes, I have some problems with the fact that the studies are retrospective in nature, but more problems with the fact that they are often not well designed.

Greg Paveza

Date: Mon, 21 Aug 1995 09:44:58 -0500 (CDT)
From: Richard Gist <rgist@jccnet.johnco.cc.ks.us>
To: "David R. Williams" <williams@CATTELL.PSYCH.UPENN.EDU>

Perhaps the greater unspoken issue implied well in the close of this thoughtful passage is simple--science is a reasoning process built upon the deductive disconfirmation of critical hypotheses derived from inductively constructed theories. Put more simply, affirmation is not POSSIBLE through the scientific process. Those who seek to affirm are fatally misguided.

RMG

Date: Mon, 21 Aug 1995 10:08:55 -0500
From: k-howard@nwu.edu (Ken Howard)
To: sscpnet@bailey.psych.nwu.edu

Over the years I have heard clinicians comment about a case "Oh! he [she] is still talking about his [her] parents. But we should get past this childhood business soon, and get down to doing therapy." Perhaps there is a relationship here with Beutler's conversion ideas. A major principle in the application of any profession is reframing the consumer's problems (hopefully in a diagnostic way or at least a way that would guide intervention). One method is to indocrinate the consumer into a new life story or reconstruction of past experience -- one that brings relief from anxiety and hopelessness. Another way would be to offer some ritualistic cleansing.

I recommend Ari Kiev's (Keiv?) old book "Magic, Faith, and Healing" in these regards.

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Mon, 21 Aug 1995 10:33:34 -0400 (EDT) From: Christopher Peterson <chrispet@umich.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
--

I think the most important point in Marty Seligman's recent posting is that childhood events, abusive or otherwise, affect adult functioning insofar as they dictate the skills and abilities of the individual, which then cascade throughout life. This point comes down to the need to specify mechanisms linking risk factors and consequences, always a good strategy. A similar idea is that early events can channel an individual into one setting rather than another. Consider the divorce of one's parents. If this results in a drastic drop in income for the custodial parent, and I am sure it often does, then the child will have fewer educational and occupational choices, and we can say that the early event had an impact -- not because it created some lurking vulnerability -- but because it changed, in ongoing fashion, the individual's setting. (Sounds like I've become a social learning theorist in my old age.)

Another point to make about the impact of early events, particularly traumatic ones, is that the research literature tends to be trauma-specific. Someone investigates sexual abuse, for example, or parental alcoholism, or whatever, as if this were the only operative factor. However, when research looks simultaneously at several such stressors, the finding is often that they do not occur independently but rather in a complex, which makes it difficult -- if not meaningless -- to expect one-to-one links between early traumas and later difficulties. It seems also to be the case that these correlated horrors are additive if not synergistic.

Thank you, Marty, for rescuing the list, for the time being, from chain letters and mail-order spouses by posting a provocative message.

Chris Peterson

Date: Mon, 21 Aug 1995 11:18:20 -0700 (MST) From: Lee B Sechrest <sechrest@aruba.ccit.arizona.edu> To: Christopher Peterson <chrispet@umich.edu>
--

On Mon, 21 Aug 1995, Christopher Peterson wrote:

> Another point to make about the impact of early events, particularly traumatic ones, is that the research literature tends to be trauma-specific. Someone investigates sexual abuse, for example, or parental alcoholism, or whatever, as if this were the only operative factor. However, when research looks simultaneously at several such stressors, the finding is often that they do not occur independently but rather in a complex, which makes it difficult -- if not meaningless -- to expect one-to-one links between early traumas and later difficulties. It seems also to be the case that these correlated horrors are additive if not synergistic. <

If the stressors do not occur independently, i.e., they tend to be correlated and to occur together, then a trauma-specific approach should result in interpretable findings. In such a case, one would be using a particular trauma as an indicator. The result might be a somewhat less than optimally reliable indicator and, hence, an attenuated effect, but the estimate should be unbiased.

If the results of trauma are additive, the problem is more complicated, and one would be in the position of detecting, for any one level of total trauma, the added effect of the trauma being studied. Again, the result would be attenuated but not biased.

The most difficult problem would arise if the effects of trauma were multiplicative, for obvious reasons, particularly if there were specific interactions. For example, if the effect of one type of trauma did not kick in at all except in the presence of another type of trauma, detecting the end result would be quite difficult. For that situation we need far better theory than we now have, i.e., something better than that "various traumas in childhood may result in various problems in adulthood."

Lee Sechrest, Univ. of Arizona

Date: Mon, 21 Aug 1995 11:56:32 -0700 (MST) From: MIKLOWITZ@clipr.Colorado.EDU To: sscpnet@bailey.psych.nwu.edu

I find intriguing the recent discussion of the relative predictive value of traumatic experiences versus genetics, personality, and recent life events (as pertains to adult psychopathology). I think Marty Seligman's essay would be a useful contribution to an adult psychopathology class, where many graduate students come in already indoctrinated in the "trauma model."

I would like to follow-up, however, on an issue alluded to in some of the recent postings following Marty's essay. Why should we think of genetics, personality, and childhood trauma experiences as "competing predictors" (in the regression sense) when in fact there may be significant intercorrelations between them?

For example, a genetic predisposition to adult depression may be reflected as alcoholism in the biological parent, which may itself predict a greater likelihood of abuse of the at-risk offspring, as well as other childhood psychosocial circumstances (i.e, poor nutrition, neglect, severe family conflict, etc). Likewise, might abuse or other childhood stress experiences affect the development of certain personality characteristics, which then "override" the

predictive value of the abuse experiences in statistically accounting for adult depression (again in the regression sense?).

Can a history of childhood abuse predict a greater likelihood of certain adult life events, which then go on to predict adult psychopathology?

I guess my real question is, Would childhood life experiences have greater prognostic importance if we considered them as mediating variables rather than as univariate predictors of adult psychopathology (as seems to be done in the majority of the trauma literature)?

David Miklowitz Ph.D.
U of Colorado at Boulder

Date: Mon, 21 Aug 95 12:27:06 EST From: "William M. Grove" <wmg@grove.psych.umn.edu> To: DrewMack@aol.com, sscpnet@bailey.psych.nwu.edu

You wrote, in part:

> Case in point: Dissociative Identity Disorder. It has been clearly established that DID is caused by early, pervasive, severe childhood physical and sexual abuse which generally commences before ages 4 or 5.<

Actually this statement is flatly false. See Putnam's book on MPD where he clearly states that the abuse etiology is an hypothesis, and that it has not been clearly established. To be fair, he goes on to state that no clinician who has seen "more than a few multiples [sic]" believes other than that severe early abuse causes MPD. However, it is of course the case that people's opinions aren't data, though they may be sources of fruitful hypotheses.

I believe that a careful review of the published literature will show the following: There exists no methodologically adequate study which shows that *verified* incidents of child abuse, occurring early in childhood or otherwise, are a necessary condition for developing MPD. There is no methodologically adequate study which shows that such incidents are a sufficient condition for developing MPD. There is no methodologically adequate study even showing that verified instances of early childhood sexual and/or physical abuse are more frequently seen in MPD cases than in non-MPD psychiatric cases.

All the MPD reports on child abuse I have ever seen are retrospective, relying on in-therapy recollections. Furthermore, all or almost all are further compromised by the use of suggestive "memory work" techniques such as hypnotism, amytal injections, and guided imagery. The hypnosis literature is replete with evidence that such techniques lead to recovery of more detail, but the detail is frequently inaccurate and/or confabulated. The recovered false recollections after hypnosis cannot be distinguished from the recovered true

recollections, as both seem equally believable to the subject. Hypnosis also reliably increases the confidence that a recollection is correct. So, the memory work techniques, such as hypnosis and related techniques, used by many if not most MPD therapists are highly likely to (a) lead to false recollections of abuse, (b) patient and therapist can't reliably tell the true memories of abuse from the false ones, and (c) the subject will likely believe the false memories of abuse to be true with relatively high confidence.

This is not a recipe for getting good data on the relationship between abuse and MPD.

Regards,

William M. Grove |
galaxy!
Psychology Dept. |
U. of Minnesota

JHVH-1 is a space alien, and still threatens this

--- J.R. "Bob" Dobbs

Date: Mon, 21 Aug 1995 12:27:09 -0400 From: "Robert E. Emery" <ree@fermi.clas.virginia.edu> To: sscpnet@bailey.psych.nwu.edu
--

The weak or zero relation between child trauma and adult psychopathology is not surprising, since there is often only a modest relation between major childhood stressors (e.g., divorce) and *any* form of child psychopathology (as best we can measure it). I don't know of any major childhood stressor or trauma where there is a one-to-one relation between an event and some child outcome. Moreover, there are a huge number of third variable problems, including the correlations between stressors and correlation between genes and environment. The literature on divorce and children's adjustment is a good example of problems with drawing causal inferences, because it is methodologically sophisticated. For example, prospective research indicates that many of the supposed consequences of divorce for children begin *before* a parental separation.

An opposing methodological warning is that our measures are extremely limited, as is our timing of measurement. We therefore need to be especially cautious not to accept the null hypothesis even in the face of null results.

What do you do clinically in the face of uncertain knowledge? Empathize with the client's perspective, candidly admit to our science's limited knowledge of cause, encourage clients to find new, more adaptive explanations/myths (Jerome Frank) for their problems. (I hypothesize that an important process component of behavior therapy is the very optimistic explanation, "You learned this problem, and you can learn new solutions.")

The problem comes when fictions that are adaptive for the individual become public proclamations of fact.

Bob Emery, University of Virginia

Date: Mon, 21 Aug 1995 13:51:00 -0400 (EDT) From: williams@cattell.psych.upenn.edu (David R. Williams) To: edonchin@s.psych.uiuc.edu
--

Manny Donchin wrote:

> those who argue for the decisive value of clinical observations as a source of support for the importance of childhood experience may be making one of two assertions;

1. Childhood experiences have a causal relationship to adult psychopathology,
2. It is useful to structure the therapeutic interaction as if childhood experiences have a causal relationship to psychopathology.>

An excellent point that certainly bears repeating, if it requires mention at all!

> Yet, practitioners may validly argue that they find useful the belief in the validity of assertion (1), but that all they are really claiming is that assertion (2) is true. Of course, this leads to the challenge of finding empirical means of validating assertion (2)..but this belongs to another thread.<

Indeed, and again well said. The point passes over one of the concerns I expressed, however: the problem of the "be-less-than- you-can-be" consequence. A standard dodge of the psychopathologist-- check current functioning against previous levels--is of no use here, and that presents a serious measurement challenge. Ironically, something like an analysis of whether someone who experiences a childhood misfortune does less well than genetic expectations might be one way of looking for damage at the level of major life decisions. Even the best studies leave plenty of variance in the "error" category: some "error variance" may reflect truly random process, but I suspect a good deal of it is merely "not yet explained".

>My point is that one may reach useful, if erroneous, conclusions about the world as long as utility is the only issue...<

Utility is a treacherous ally. The great utility of nomothetic realism confers no right to scientific sainthood: its measures are often too coarse, or too narrowly focused. Measures of "well-being" or "fullness of functioning" may be notoriously speculative and scientifically unattractive for a variety of reasons, but their dis-utility provides no valid basis for neglecting possible sources of contribution to personality and psychopathology from the troublesome Jamesian domain.

Indeed, what motivated my comment was worry that the utility of behaviorally-based statistical purees would inhibit a therapist's legitimate concern with one's quality of personal experience, as a reflection of the story one chooses to tell about one's self. Scientifically, what bothered me was the possibility that the obvious utility of nomothetics could discourage efforts to carry the concept of scientific measurement further into the personal, subjective domain where individual reactions to trauma and insufficiency take place.

>Drawing general, theoretical, conclusions is another matter altogether.

And much too important to be based on considerations of utility, clinical, scientific, or otherwise!

david williams

Date: Mon, 21 Aug 1995 15:55:54 -0400 (EDT) From: John F Kihlstrom <kihlstrm@minerva.cis.yale.edu> To: Lee B Sechrest <sechrest@aruba.ccit.arizona.edu>

I thank Bill Grove for reinforcing a point I have been trying to make on the dissociative-disorders list for more than a year, without any apparent success: retrospective studies of dissociative disorder patients can't demonstrate a causal link to child abuse.

Moreover, even if there were such a link, one would want to tease apart the relative importance of the following types of factors: the fact of having been abused; having been raised in the kind of "dysfunctional" family in which abuse tends to occur; one's social identification (i.e., by parents, teachers, health professionals, etc.) as an abuse victim; and one's self-identification as an abuse victim.

Abuse is wrong, regardless of its effects (the abuse community is somewhat McCarthyite, and requires critics to say obvious things like this), but from a scientific point of view it is important to evaluate abuse and its context separately.

John F. Kihlstrom, Professor

Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205
Telephone (203) 432-2596 Facsimile (203) 432-7172

Date: Tue, 22 Aug 1995 10:38:24 -0700 (PDT) From: Jason Satterfield <satter@itsa.ucsf.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>

I've greatly enjoyed the discussion of childhood trauma/adult psychopathology links but have been puzzled by the (thus far) omission of the neurobiological literature on stress and trauma. The focus on skill deficits and other psychosocial variables (attributional style, etc) could certainly prove to be fruitful but I'd like to offer the hypothesis that childhood trauma creates chronic neurochemical and even morphological CNS changes which predispose an individual to develop psychopathologies later in life.

It is true that our understanding of the biochemistry of psychopathologies is murky at best (not to mention the biochemistry behind psychosocial variables such as attributional style) but I believe an exploration of the current research could provide some important clues in linking childhood trauma with adult psychopathology (biological types, please help me out).

A few ideas:

The phenomenon of kindling (Gold et al., 1988; Post, 1992; Post et al., 1982, 1984). Researchers have hypothesized that painful events such as losses, trauma, neglect which occur early in life may sensitize neural pathways/receptors leading to vulnerability to depression or other affective disorders later in life. Using animal models, Post and his co-workers have demonstrated that repeated application of chemical or electrical stimulation (distant relative to chronic stress?) lead to increasing levels of neurobiological response until the the response eventually occurs even in the absence of external stimulation. Why is early life (or childhood trauma) particularly important? The plasticity of the CNS is much greater in childhood than in adult life allowing more room for permanent neurochemical or even structural changes. It is possible that the skills deficits, pessimism, affective lability or whatever psych variable we want to discuss is influenced by enduring biological changes brought about by childhood trauma.

Another useful heuristic in understanding the interplay between childhood trauma and brain function is primate research.

1. Repeated separations of infant and mother squirrel monkeys causes permanent elevations in plasma cortisol levels in the infants (Coe et al., 1978; Coe et al., 1983). Cortisol has been linked with severity of response to stressors in adult life (in humans) and has been implicated (the jury is still out) in affective disorders.

2. Social isolation and separation has a direct impact on the sensitivity and number of opiate receptors in the infant brain (van der Kolk, 1987). Substance

abuse research has been linking opiate sensitivity to multiple varieties of adult substance dependence disorders.

3. Other long-lasting brain function effects of "childhood" trauma include changes in adrenal gland catecholamine-synthesizing enzymes, hypothalamic serotonin secretion, and noradrenergic receptor sensitivity. All of the above have been implicated in adult and childhood psychopathology.

(For a very quick, beginning review of the interplay between biology and therapy see Gabbard, G.O. (1992). Psychodynamic psychiatry in the "Decade of the Brain", American Journal of Psychiatry, 149(8), 991-998.)

I realize I've had to use the words "associated, implicated, etc" and nothing causal has been demonstrated (and perhaps never will be). However, I think it's important to remember that psychosocial stressors have an indisputable effect (whatever it might exactly be) on brain function which in turn affects mood, cognition, memory, pathology.

Perhaps when a child is faced with serious psychosocial stressors (plus genetic vulnerability, etc) his/her CNS is permanently and adversely altered—more so than if that same trauma occurred in adulthood.

Of course, finding causal or associated biological changes in childhood doesn't mean treatment has to be an exploration of childhood trauma, recovering "repressed memories", or even treatment with medications. But that's a discussion for another day....

Jason M. Satterfield, Ph.D.

Department of Psychiatry
University of California at San Francisco
401 Parnassus Ave, Box CPT
San Francisco, CA 94143-0984
(415) 476-7569

Date: Tue, 22 Aug 1995 17:09:17 -0700 (PDT) From: Jason Satterfield <satter@itsa.ucsf.EDU> To: Lee B Sechrest <sechrest@aruba.ccit.arizona.edu>

On Tue, 22 Aug 1995, Lee B Sechrest wrote:

> Dr. Satterfield provides a very interesting and useful summary of ways in > which early trauma might result in lasting changes in the brain and dispose the individual toward later pathology.

Then why is it not relatively simple to demonstrate a relationship between early trauma and later pathology? That is the missing link that is required to make the argument truly useful. >

Good question and one that is not easily amenable to empirical investigation. Because of the nature of the beast (complexity and ethical issues around abuse) we cannot run the studies that need to be done to find "the" links (if any) between early trauma and later pathology. We do have weak correlations between abuse and degree of sx albeit in studies with some less than ideal methodologies (Roesler & McKenzie (1994). Effect of childhood trauma on psychological functioning in adults sexually abused as children. Jnl of Nervous and Mental Disease, 182(3), 145-150.)

My pet hypotheses for the missing link would be:

1) Constitutional and other attenuating factors that minimize the impact of trauma. This includes minimizing the impact on brain function. These factors could be genetic, chemical, or psychosocial. The subjective interpretations of aversive events and ways of coping would fall into this camp.

2) The traumatic experience is but one powerful event that shapes the life (neurobiology and psychology) of the individual. It seems logical that powerful positive events could also impact brain function and ability to "recover" from trauma. And yes, there will certainly be critical subjective elements (e.g. assigned meanings) here too. Trying to predict present behavior (or anything for that matter) by looking at an arbitrary and isolated chunk of childhood denies the development of the individual (and perhaps his/her psychopathology) over time. Predictions from time 1 to time 2 are just too simplistic. But to deduce from this that childhood trauma has NO effect is a bit of stretch.

-Jason
Jason M. Satterfield, Ph.D.
Department of Psychiatry
University of California at San Francisco
401 Parnassus Avenue, Box CPT
San Francisco, CA 94143-0984
(415) 476-7569

Date: Tue, 22 Aug 1995 17:38:11 EDT From: "David Freides" <FREIDES@fs1.psy.emory.edu> To: sscpnet@bailey.psych.nwu.edu
--

Satterfield has amply introduced the available data biological consequences of early stress and Sechrest has again raised the question of why no clearer relationship --

I digress slightly to make a point. A child of 6 loses a parent. Everything I know theoretically and otherwise tells me that the consequences of that loss will depend on how that child mourns and what substitute relationships are developed, or pre-existed.

The point. Stress does not produce psychopathology. Levine and many others demonstrated that way back. The consequences of stress depend on how the stress is coped with. If a research design does not attempt to encompass this complexity, the researcher is oversimplifying what we know about living systems evolved to adapt to changing environments. He or she is also getting null results.

David Freides, Emory University

Date: Tue, 22 Aug 1995 23:14:41 -0400 From: fresco@cattell.psych.upenn.edu (David M. Fresco) To: David Rosenhan <rosenhan@psych.stanford.edu>

At 15:57 22/08/95, David Rosenhan wrote (see above):

You raise some important ideas when you talk of capturing the "subjective perception of that experience." Yet, the methodologist in me is really scratching his head as to how to test the subjective experience a la Mischel and Shoda. I also have to confess that I have not yet seen the M & S paper--so I must rely on my subjective read of your subjective appraisal of M & S :-)

Perhaps I am having trouble emerging from a strict helplessness perspective, but I don't see the 5 mediating units that you describe above as much different than the diathesis component of the now myriad diathesis stress models to account for various pathologies. I also believe that we have developed a relatively sensitive methodology to psychometrically test the interplay of mediating units and the stresses present in one's life.

It seems to me that within the Helplessness/Hopelessness framework (or from Beck's and Lewinsohn's perspective for that matter) there is some

acknowledgement and agreement that people acquire mediating units. These mediating units do just that mediate, filter, or exacerbate the impact of current life stress. That's why we go out of our way to assess our hypothesized diatheses in our psychopathology studies. We are (ostensibly) capturing the degree of mediation by these units here.

Consequently, we in turn want to assess life stress in as objective a manner as possible. We have Brown & Harris and Monroe to thank for propelling our awareness of bias introduced (methodologically) by subjective distress ratings--and in bailing us out by developing means of objectifying life stress with less influence from mediating units.

We also have Metalsky & Joiner to thank for teaching us (or at least me) how to effectively use set-wise multiple regression, Cohen's & Cohen's Analysis of Partial Variance (APV) and Baron & Kenny moderation/mediation analyses.

So, to describe the here-and-now of our clients/subjects, I think we have the bases pretty much covered when it comes to capturing the subjective perception. The area that needs the greatest articulation now is on the acquisition of mediating units. All of the diathesis stress theories remain relatively silent on this point. Taking them all together we have a smattering of genetic inheritance, attachment theory, conditioning history--all within a broad social-learning perspective.

I think Marty's original post hints at a compelling way to describe the acquisition of mediating units following childhood trauma/abuse, but may actually speak to the acquisition of mediating events from our entire learning history too.

=-=-

= David M. Fresco Internet: fresco@cattell.psych.upenn.edu =
= Department of Psychology fresco@unc.edu
= CB#3270, Davie Hall
= Chapel Hill, NC 27599

Date: Wed Aug 23 15:45:11 1995
From: Lee B Sechrest <sechrest@aruba.ccit.arizona.edu>
To: Jason Satterfield <satter@itsa.ucsf.edu>

Dr. Satterfield provides a very interesting and useful summary of ways in which early trauma might result in lasting changes in the brain and dispose the individual toward later pathology.

Then why is it not relatively simple to demonstrate a relationship between early trauma and later pathology? That is the missing link that is required to make the argument truly useful.

Lee Sechrest, Univ. of Arizona

Date: Wed Aug 23 15:45:21 1995
From: David Rosenhan <rosenhan@psych.Stanford.EDU>
To: Jason Satterfield <satter@itsa.ucsf.edu>

Following the line of issues raised by Satterfield, Sechrest and Friedes, it may well be that the focus of the C-A investigations has been misplaced to the *experience* of say, child abuse, instead of the *subjective perception* of that experience. In that regard, it's useful to reread the recent Psych Review paper by Mischel and Shoda (1995).

They identify five mediating units that individuals bring to any experience, among which are: 1) encodings about the self, others, events and situations; 2) expectancies and beliefs about the social world; 3) affects; 4) goals and values; and 5) competencies and skills. The evidence strongly suggest that within the normal ranges of personality, as well as in the ranges of development (as Marty points out), experience per se, has no direct and unassailable meaning. Without evidence regarding the construal of that experience, one cannot hope meaningful relationships between past and present.

Date: Wed, 23 Aug 1995 19:24:35 -0500
From: talmadge@onramp.net (John M. Talmadge, M.D.)
To: David Rosenhan <rosenhan@psych.stanford.edu>,
Jason Satterfield <satter@itsa.ucsf.edu>

At 3:57 PM 8/22/95, David Rosenhan wrote (see above):

I'm not familiar with Psych Review. Can you advise me how to find it? I'm sure others will be interested as well. The hierarchy you summarize makes a

good deal of sense to me, but I'm something of a curmudgeon constructionist by nature anyway. I wish I'd never discovered epistemology.

JT

John M. Talmadge, M.D.
Psychiatry and Addiction Medicine
9535 Forest Lane, Suite 110
Dallas, Texas 75243
(214) 680-0400 ext 285

The Trout Spies a Fly

If you wish your misery to end,
seek also to lose your wisdom --
the wisdom born of human illusion,
that which lacks the light
of God's overflowing grace.
The wisdom of this world increases doubt;
the wisdom of faith releases you into the sky.

Rumi, Mathnawi, II, 3200-203
(Persian Poet, 13th Century)

Efficacy issues once again

including
APA's MANUALS FOR EMPIRICALLY VALIDATED TREATMENTS

letters fromsscpnet 4/95

Letters from the editor:

Much more than anyone would think the SSCPnet discussions revolve around the scientific and clinical value of efficacy and effectiveness studies. Anyone concerned with ongoing discussions - which of the many treatments are truly and in a clinical sense meaningful - should immerse herself or himself in these fascinating discussions.

The German readership of the fourth edition of the "Letters from SSCPNET" is asked to treat these documents as < shareware > implying some small financial reimbursement for the copy provided.

Account Nr 7 611 454 Sparkasse Ulm (BLZ 630 500 00)

Forschungskonto Prof. Kächele (per issue DM 10.-).

Only those of my readers that interested as demonstrated by this small contribution will receive further editions of the LETTERS.

yours horst kaechele

Date: Sat, 24 Jun 1995 18:19:30 -0700 (PDT)
From: Neil Jacobson <njacob@u.washington.edu>
To: section3 <:sscpnet@bailey.psych.nwu.edu>
Subject: clinical significance, CBT for depression, and possible , rapproachment with
Larry Beutler?

Dear Colleagues,

I just returned from SPR, where I served as discussant for a panel on the clinical significance of CBT as a treatment for depression. After my remarks, a number of interesting questions and comments were made by colleagues in the audience, notably, Tracie Shea, Ken Howard, and Ellen Frank. The format didn't allow sufficient time for discussion. However, the issues raised by the panel and the comments afterward are very important ones, and I was reminded about how much room there is for reasonable minds to disagree on the extant literature, and what it means. Delete without reading if you are tired of these issues. But I thought about them for the rest of the day, and think that some of them are of general interest.

First, it seems clear to me that clinical trials have generally provided optimal tests of the efficacy of CBT and other psychotherapies, and undoubtedly overestimate the value of these treatments as practiced by your typical clinician who is not rigorously trained, monitored, or supervised during the course of a trial. Our research and that of others raises questions about the exportability of CBT as a treatment for depression into naturalistic settings, since competence seems to drift downward even among highly experienced therapists who were well-trained to a certain level of competence, unless supervision/calibration remains quite intensive throughout the trial. Any treatment which requires such intensive supervision, even with highly experienced and well-trained therapists, may not generalize well to typical practitioner settings where there is little training and certainly no supervision. In fact, in all likelihood most practicing cognitive therapists for depression have at most read the Beck et al. manual, and perhaps attended a didactic workshop.

Second, the results in clinical trials are further skewed by subject selection procedures designed to homogenize the sample and detract from its representativeness to clinical practice. As Ellen Frank and Tracie Shea pointed out at the meeting, this subject selection practice cuts both ways. On the one hand, the single most often used basis for exclusion in depression trials is that subjects are not sufficiently depressed: so subjects treated in clinical trials are in all likelihood MORE depressed than those typically seeking therapy from practitioners. On the other hand, the exclusion of bipolar I and bipolar II patients, patients with dual diagnoses, including current episodes of substance abuse, panic disorder, eating disorders, and the like limit generalizability. But these two opposing processes (selecting for real major depressive disorder while at the same time screening out diagnostic complexity) result in a sample of patients that in some ways are easier to treat, and in other ways are harder to treat.

Third, even taking into account the highly select sample of patients selected for these trials, the efficacy of CBT, as well as other available pharmacological and psychosocial treatments, are exceedingly modest from the standpoint of clinical significance. Whether psychosocial treatments, including CBT, exceed rigorously constructed placebos is still very much open to question.

Fourth, and perhaps more importantly, I was challenged to explain an apparent inconsistency from the conclusions of my review with Andy Christensen, showing that neither professional training nor years of therapist experience correlate with outcome, and the argument that competence in CBT appears to drift and may effect outcome without constant intense supervision. Andy and I found that whether one receives a professional degree does not enhance his or her clinical outcomes, on the average. Moreover, given a professional degree, one is just as good as another, on the average. This raises the question of whether doctorally trained psychologists will, on the average, be better therapists than those with masters degrees, bachelors degrees, or even degrees in geography. This same literature also suggests that being in practice longer does not make one a better therapist, on the average.

However, it is entirely possible that, despite these conclusions, clinical outcome WILL be enhanced for specific problems with training in specific skills. We do not know that for sure: for example, it could be that no matter how much a therapist "looks like" a good cognitive therapist, outcomes will not be effected. These questions remain to be tested, and current scales designed to measure competence, such as they are, do not seem to be correlated with outcome. But it is possible that training and supervision in specific skills does effect clinical outcome for particular problems, even though, on the average, neither professional training nor years of experience matter. Until this possibility is eliminated, it is incumbent upon investigators to ensure that therapists "look like" competent cognitive therapists, to the extent that these behaviors have been operationalized.

Finally, Ken Howard thought that my pessimism about CBT for depression was unwarranted, cognitive behavior therapist that he is. He alluded to the overwhelming evidence that psychotherapy is effective, presumably referring to meta-analyses such as the one by Smith, Glass, and Miller. But he was talking about apples, and I was talking about oranges. He was talking about the average size of a statistical effect summed across hundreds of studies, mostly done with mildly disturbed patients, showing that psychotherapy is better than nothing. His apple is irrelevant to my orange. For one thing, the size of a statistical effect has nothing to do with its clinical significance. But more importantly, those meta-analyses tell us nothing about CBT or any other psychosocial treatment for depression. In fact, if one wants to play meta-analytic baseball, you need go no further than the Robinson et al. meta-analytic review of psychotherapy for depression, which shows that psychotherapies have not been shown to outperform placebos. To the extent that they show effects, they are not specific to unique aspects of technique or strategy within a form of therapy but rather nonspecific effects that may be

nothing more than the person in the white coat listening in a caring way, making pronouncements consistent with one's theoretical convictions, with an involved client who wants to believe in psychotherapy, and has every reason to be involved in the process.

Finally, I should note that everything I am saying about psychotherapy could be applied to pharmacotherapy, except the part about not exceeding placebo-which many antidepressants clearly do. Thus, in a recent paper Steve Hollon and I question the conclusion that pharmacotherapy is more effective than CBT because it performed better in the NIMH Collaborative study. I doubt whether in the long run drugs work any better than psychotherapy, but since they are cheaper, the burden of proof is on us, the psychotherapists.

Neil S. Jacobson, Ph.D.

University of Washington
Center for Clinical Research
Department of Psychology
Fax: 206 685-3293
Phone: 206 543-9871

Date: Wed, 28 Jun 1995 08:55:00 -0600 From: Mack Stephenson <MACKS@stlgate.byu.edu> To: SSCPNET@BAILEY.PSYCH.NWU.EDU Subject: Treatment Efficacy & placebo

Neil Jacobson wrote:

>Finally, I should note that everything I am saying about psychotherapy could be applied to pharmacotherapy, except the part about not exceeding placebo-which many antidepressants clearly do.<

As far as I can tell, the psychopharmacologists have not proven that the antidepressants are better than placebo. I know Don Klein hates it when I bring up Fisher & Greenberg's stuff, but it is convincing--the trials are not truly blind, and the blinder the trial, the less difference between active & placebo.

The fatal flaw has been the lack of active placebos on a regular basis. How on earth can someone argue that they are controlling for placebo effects when the placebo experience is markedly different than the active treatment??! Don Klein is correct in criticizing many psychotherapy treatments for their problems in dealing with placebo control, but on the flip side, the pharmacologists have an easy and obvious way to make a meaningful placebo and don't.

Finally, we may be vastly underestimating how powerful it is for a patient to receive a treatment that both they and the doctor believes in. The article in

Clinical Psychology Review two years ago (Vol.13 #5) examined five medical/surgical treatments for "physical" disorders. All of the five had nothing to do with the actual physiology of the malady, yet all five were passionately believed in by their proponents, and all five had scientific studies to back them up. Combining the five treatments, 40% had "excellent" response, 30% had "good," and 30% had "poor" response. Seventy percent got better to some noticeable degree!!

Now, a depressed person goes to a physician, who tells him or her that he or she has a "biochemical imbalance" and that pills can be taken to correct this imbalance. The physician is utterly convinced of this "fact" (and has considerable prestige and authority). The patient also believes this, given the cultural popularity of biochemical explanations (as a sidelight, Ely Lilly is running an ad campaign urging people to "see your physician" if you feel depressed because it's a chemical imbalance--they don't identify themselves as the maker of Prozac). Given the fact that both doctor and patient thoroughly believe the explanation and the medication produces distinct bodily sensations, it seems reasonable to see antidepressant placebo response rates up into the 60 and 70 percent range. I could go on, pointing out how the placebo washout may selectively exclude those who respond to an inactive placebo and retain active placebo responders, and that the placebo rate is probably higher in practice than in research, etc.etc.

Suffice it to say, I don't think medication has convincingly demonstrated its superior efficacy to placebo.

Mack Stephenson

Macks@stlgate.byu.edu

From: ANTONYM@cs.clarke-inst.on.ca
Date: Wed, 28 Jun 1995 12:24:25 -0400 (EDT)
To: sscpnet@cs.clarke-inst.on.ca
Subject: active vs. inactive placebos

I believe that the term "active placebo" may have been used to refer to a placebo that mimics the side effects of a particular medication but without the therapeutic benefits.

Martin Antony
University of Toronto

From: ssoldz@warren.med.harvard.edu
Date: Thu, 29 Jun 95 12:46:27 EST
To: Bill Follette <follette@unr.edu>,
Subject: Re: two sigma!

I'm in agreement that the clinically significant criteria is too strict in requiring return to the normal distribution. This is a nice goal, but may be often unattainable. Schizophrenics may never return to the normal range in many areas of functioning, but this hardly means that treatment is irrelevant. The same may be true with depression. We may not be able to accomplish that goal, but we may markedly improve that state.

It would seem that the first criteria a treatment should make is that any improvement to noticeable (to at least one of: Self, or significant others). Dan Ozer (in J. of Personality, 1994) has extended the Just Noticeable Difference (JND) from psychophysical scaling to personality assessment. Perhaps we should extend it to clinical assessment. The preliminary criteria for treatment should be that its effect is greater than the JND for the dimension (eg., depression). If the treatment accomplishes greater effects, that is, of course, much better, but not a requirement for clinical significance.

A friend's brother recently had a liver transplant. The outcome is that he didn't die, surely a clinically significant outcome! Yet, he may never return to the normal distribution in terms of functioning. At SPR some years ago, in responding to the sovering followup data from the Collaborative Study, Varda Solomon commented on our tendency to think we can be as effective as god. Accomplishing less, but something, shouldn't be diminished.

Stephen

Stephen Soldz
Director of Research

Boston Institute for Psychotherapy
520 Commonwealth Ave.
Boston, MA 02215
&
Dept. of Psychiatry
Harvard Medical School
(617) 469-3576

Date: Wed, 23 Aug 1995 14:08:26 -0400 (EDT) From: Stephen Scott Ilardi <ssi@acpub.duke.edu> To: Jason Satterfield <satter@itsa.ucsf.edu> Subject: CBT and nonspecific factors
--

On Fri, 18 Aug 1995, Jason Satterfield wrote:

> Do CPT, IPT, BMT, DBT specific interventions give us any significant power over and above nonspecific factors or are they just creating a convincing myth to encourage a "conversion" experience in the patient? Before reflexively flaming, take another look at Ilardi & Craighead 1994 or the recent "Clinical Psych:Science and Practice" issue which discussed "common factors". It's a fascinating puzzle. >
Dr. Satterfield,

Although Ed Craighead and I did argue that Beck's CBT appears to promote a very rapid reduction in depressive symptomatology largely due to the protocol's ability to capitalize on nonspecifics (such as providing a credible tx³. rationale, or "myth"), we were also careful to note that the efficacy of CBT is not "merely" a function of nonspecific mediation (Clin Psych:Sci & Pract 1:138-156, 1994). Clinical improvement in the latter stages of therapy has been tied specifically to the successful application of cognitive restructuring techniques (DeRubeis et al., 1990; Persons & Burns, 1985; Teasdale & Fennell, 1982). Further, several investigations have indicated that CBT confers some prophylactic benefit with respect to depressive relapse -- a phenomenon which we consider more consistent with a "compensatory skills" model of change in CBT (Barber & DeRubeis, 1989).

Along these lines, K. Howard et al. (1993) have proposed a "three phase" model of psychotherapy, in which nonspecific factors are tied to (1) remoralization and (2) rapid early reduction of depressive symptoms, while the final phase of treatment (3), beginning about the 5th session, is characterized by improvement across several domains of functioning (e.g., the acquisition of a set of cognitive and behavioral skills to employ in the face of distressing life events). It is in this latter phase of therapy that one might expect to find a central role for the specific effects of manualized therapy protocols, including CBT.

³tx = therapy

Thus, I'm concerned that anyone would conclude that CBT (or DBT, IPT, BMT) is efficacious solely on the basis of nonspecific effects (or, as you put it, "therapy [is] just a high-tech version of shamanism"). Rather than see the issue in all-or-nothing terms, I'm inclined to see a central role for both specific and nonspecific treatment effects in therapy, even though the relative contributions of each may vary as a function of several variables, including time.

Sincerely,

Steve Ilardi, Ph.D.
University of Colorado at Boulder

Date: Wed, 23 Aug 1995 11:51:51 -0700 (PDT) From: Jason Satterfield <satter@itsa.ucsf.EDU> To: Stephen Scott Ilardi <ssi@acpub.duke.edu> Subject: Re: CBT and nonspecfic factors

On Wed, 23 Aug 1995, Stephen Scott Ilardi wrote:

[big snip of cogent and appreciated arguments]

> Thus, I'm concerned that anyone would conclude that CBT (or DBT, IPT, BMT) is efficacious solely on the basis of nonspecific effects (or, as you put it, "therapy [is] just a high-tech version of shamanism"). Rather than see the issue in all-or-nothing terms, I'm inclined to see a central role for both specific and nonspecific treatment effects in therapy, even though the relative contributions of each may vary as a function of several variables, including time. <

I appreciate your post. It wasn't my intention to imply that the effects of CBT, IPT or whatever was due solely to nonspecific treatment effects. I hope others did not get that message so strongly. In fact, I said my concern was that therapeutic change was "partly or even greatly" due to non-specific effects and "I am beginning to wonder if therapy is just a high-tech version of shamanism". Not quite so all-or-nothing.

I did get and agree with your point that later in therapy ("stage 3" or whatever) specific effects can take place. I am also familiar and even excited about the work of Rob DeRubeis et al (he was my clinical supervisor when I was in training at Beck's Center for Cognitive Therapy). However, I would like to know how much extra "mileage" treatment specific effects actually give us. If the added benefit of specific effects is small compared to non-specific factors then it seems like a lot of our research energies are misdirected. Of course, the "relative contributions" question is an empirical one - a challenge I hope someone will accept soon.

My best,

Jason Satterfield

Department of Psychiatry
University of California at San Francisco
401 Parnassus Avenue, Box CPT
San Francisco, CA 94143-0984
(415) 476-7569

Date: Thu, 24 Aug 1995 06:06:03 -0400 From: fresco@cattell.psych.upenn.edu (David M. Fresco) To: Jason Satterfield <satter@itsa.ucsf.edu> Subject: Re: CBT and nonspecific factors

At 11:51 23/08/95, Jason Satterfield wrote:

(see above)

I wonder if Needles and Abramson might be a place to start? I have always found it to be an intriguing paper.

Needles, D. J., & Abramson, L. Y. (1990). Positive life events, attributional style, and hopefulness: Testing a model of recovery from depression. *Journal of Abnormal Psychology*, 99(2), 156-165.

A model of a recovery process from depression that is compatible with the hopelessness theory of depressive onset is proposed. This model predicts that depressives who have an enhancing attributional style for positive events (i.e., make global, stable attributions for such events) will be more likely to regain hopefulness and, thereby, recover from depression, when positive events occur. This prediction was tested by following a group of depressed college students longitudinally for 6 weeks. Although neither positive events alone nor attributional style alone predicted reduction in hopelessness, depressives who both showed the enhancing attributional style for positive events and experienced more positive events showed dramatic reductions in hopelessness which were accompanied by remission of depressive symptoms. Thus, attributional style for positive events may be a factor that enables some depressives to recover when positive events occur in their lives.

David M. Fresco Internet: fresco@cattell.psych.upenn.edu
= Department of Psychology fresco@unc.edu
= CB#3270, Davie Hall
= Chapel Hill, NC 27599

Date: Thu, 24 Aug 1995 13:52:16 -0400 (EDT) From: Stephen Scott Ilardi <ssi@acpub.duke.edu> To: Jason Satterfield <satter@itsa.ucsf.edu>
--

Subject: Re: CBT and nonspecific factors

On Wed, 23 Aug 1995, Jason Satterfield wrote:

[snip]

> I would like to know how much extra "mileage" treatment specific effects actually give us. If the added benefit of specific effects is small compared to non-specific factors then it seems like a lot of our research energies are misdirected. Of course, the "relative contributions" question is an empirical one - a challenge I hope someone will accept soon.

This is a great point.

Do you think the issue is related to the Dodo verdict, which has been rendered vis-a-vis just about every comparative therapy outcome trial, "Everyone has won, and all must have prizes"? That is, to the extent that a protocol's "specific" components make a sizeable contribution, couldn't we expect that therapy to generate superior outcomes compared with a protocol which relied solely on nonspecific effects (e.g., the "placebo plus clinical management" condition in the NIMH TDCRP)? <

Along those lines, last night I dug up a meta-analysis by Svartberg & Stiles (JCCP, 59:704-14, 1991) in which CBT was found to be significantly superior to "short-term psychodynamic psychotherapy". Then again, maybe CBT is simply better at capitalizing on nonspecific effects. As you say, the question is an empirical one.

My regards,

Steve Ilardi, Ph.D.
University of Colorado at Boulder

Date: Fri, 25 Aug 1995 15:23:42 -0700 (PDT)
From: Jason Satterfield <satter@itsa.ucsf.EDU>
To: Stephen Scott Ilardi <ssi@acpub.duke.edu>
Subject: Re: CBT and nonspecific factors

On Thu, 24 Aug 1995, Stephen Scott Ilardi wrote:

> This is a great point. Do you think the issue is related to the Dodo verdict, which has been rendered vis-a-vis just about every comparative therapy outcome trial, "Everyone has won, and all must have prizes"? That is, to the extent that a protocol's specific components make a sizeable contribution, couldn't we expect that therapy to generate superior outcomes compared with a protocol which relied solely on nonspecific effects (e.g., the "placebo plus clinical management" condition in the NIMH TDCRP)?<

A protocol which relies solely and explicitly on non-specific effects is hard thing to imagine. It seems that providing a theoretical framework is essential to instill confidence, hope, therapist's enthusiasm, etc. Nonetheless, I've chosen group therapy to begin to address this question of non-specific vs specific effects. There's a lot more room for non-specifics and lot more research on group dynamics to build operationalizations upon. I'd be happy to send details to interested parties.

Jason Satterfield, Ph.D.
Department of Psychiatry
University of California at San Francisco
401 Parnassus Avenue, Box CPT
San Francisco, CA 94143-0984
(415) 476-7569

Date: Thu, 24 Aug 1995 13:47:06 -0500
To: Lee B Sechrest <sechrest@u.Arizona.EDU>
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: Two of my favorite things (baseball and therapy)

Lee:

See the Crits-Christoph paper I referenced in my last message. Compared to within therapist (outcome) variance, the variance between therapists is trivial (2-4%). The issue I raised is whether this is a fair way to evaluate differences.

With regard to your first response about hitters. What if you had to choose between two cardiac surgeons for some very difficult last-ditch procedure. One has a .225 success rate and the other has .350 success rate (this difference is

statistically significant at the .05 level). Each has done about 100 such procedures. The amount of variance explained is quite small.
Do you have a preference?

Ken Howard

Date: Thu, 24 Aug 1995 12:53:24 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: Ken Howard <k-howard@nwu.edu> Subject: Re: Two of my favorite things (baseball and therapy)
--

On Thu, 24 Aug 1995, Ken Howard wrote:
(see above)

Of course I have a preference, just as any baseball manager would almost always prefer to have a better batter at the plate. *Ceteris paribus*, go with the best record, even if it is weensy.

But a manager will sometimes go with a weaker hitter in a given spot because that hitter is a better fielder, etc. And if one surgeon offers to do the surgery for \$5000 and the other wants \$500,000, then I may have to take a chance on the lesser of the two.

If we could increase the success rates of psychotherapists by doubling the amount of time we keep them in training, should we do that? Depends.

Lee

Date: Thu, 24 Aug 1995 12:12:48 -0500 To: "Robert W. Montgomery, Ph.D." <psyrwm@panther.Gsu.EDU> From: k-howard@nwu.edu (Ken Howard) Subject: Re: Baseball as life

The median number of sessions of therapy is quite dependent on the provider type and the definition of psychotherapy. In our clinic, for example, the median is about 14 sessions, in general settings it seems to be about 6 sessions.

The mode is another story altogether. The modal number of visits is 1 (the modal number of prescriptions of Prozac is also 1).

Ken Howard

Date: Thu, 24 Aug 1995 12:39:31 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: Ken Howard <k-howard@nwu.edu>
Subject: Re: Baseball as life

On Thu, 24 Aug 1995, Ken Howard wrote:
see above

Boy, psychotherapy and Prozac are both really powerful stuff! What is the modal number of martinis?

Lee

Date: Thu, 24 Aug 1995 12:47:22 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: "Robert C. Carson" <rcarson@acpub.duke.edu>
Subject: Re: Iatrogenic Effects (fwd)

On Thu, 24 Aug 1995, Robert C. Carson wrote:

Much deleted

< I also believe we have been lax in this respect in "legitimate" training programs in terms of some of the people we certify as ready to be inflicted on the public; for example, I would not refer my worst enemy to some graduates of my own program. <

Well, Bob, you had better think carefully about who referred those students to your program.

Lee

Date: Fri, 25 Aug 1995 08:45:33 -0500 (CDT)
From: "Stephen M. Saunders" <SAUNDERSS@vmsa.csd.mu.edu>
To: sscpnet@bailey.psych.nwu.edu
Subject: baseball and therapy

Both baseball and therapy involve both ability and achievement. With baseball, there is ability, which is hard to measure, and achievement, such as batting average, which is probably imperfectly measured. With therapy, measuring either is probably even more difficult.

My prior posting to this net was meant to point out that, obviously, baseball ability and achievement are related. However, apparently we cannot statistically account for that relationship very well, that is, our statistical methods are incapable of detecting the relationship. Might this also be true with therapy?

Also, to conclude that a relationship does not exist because studies cannot find it is, it seems to me, to accept the null hypothesis.

Stephen Saunders
Marquette University

Date: Fri, 25 Aug 1995 09:36:56 -0500 From: k-howard@nwu.edu (Ken Howard) To: John F Kihlstrom <kihlstrm@minerva.cis.yale.edu>Subject: Re: Iatrogenic Effects (fwd) Cc: sscpnet@bailey.psych.nwu.edu
--

There is certainly a problem in interpreting the phrase "based on science." In epistemology knowledge claims are dependent on the so-called "evidence requirement" -- x has the right to know that y. This right to know can be based on empirical facts (e.g., research results) and/or on credentials.

Ken Howard

Date: Thu, 31 Aug 1995 10:11:08 -0500 To: sscpnet@bailey.psych.nwu.edu From: k-howard@nwu.edu (Ken Howard) Subject: RE: Prozac

I would like to echo Bob Carson's admonition about attacking an author because he represents a far out point of view. How can people (scientists, no less) in good faith read Dawes' arguments about treatment efficacy or Jacobson's arguments about experience and training as being objective (not representing their points of view) and then trash (without reading) the next person's (also presumably a scientist) arguments?

Ken Howard

Date: Fri, 8 Sep 95 12:53:40 PDT
From: beutler@edstar.gse.ucsb.EDU (Larry Beutler)
To: ssi@psych.colorado.edu
Subject: Re: CBT and nonspecific factors (fwd)
Cc: sscpnet@Bailey.psych.nwu.edu

The placebo analogy for discussing nonspecific or common factors bothers me. Moreover, I think it will be very difficult to distinguish between what you call "common factors" and what have been called "nonspecific factors". What if common factors are only common in that these are aspects of patient hope, faith, work, collaboration, etc that are activated by therapist techniques. Moreover assume that some techniques almost universally activate some level of these qualities while other techniques only do so in some people. Further, assume that when these latter techniques are used with compatible individuals they add to the magnitude of the qualities developed by the techniques (perhaps, empathy, caring, support, optimism) of the therapist that almost always evoke these patient responses. But, other techniques attenuate or even counteract the effect of these latter therapist behaviors.

In the foregoing, there are common responses of patients (hope, faith, motivation, etc.), therapist qualities that are common to various therapies but that may or may not be discussed in those theories (empathy, caring, support, etc), and therapist actions (techniques) some of which are common to two or more therapies, some of which are unique, and some of which may be common to all that exert an additive or subtractive effect on the "common" responses of patients. Now, which are the common factors--the therapist behaviors that cut across some or all therapies or the responses of the patients? Where is therapeutic alliance--a therapist behavior or a patient response--and is it common? And, what do we do about something like empathic reflection if it turns out to be common across all therapists (in different quantities) but to exert additive effects with some patients and subtractive ones with others?

It seems to me that we must differentiate between things that are therapist controlled and those that are not and between the reactions of patients (hope, remoralization) and the things that evoke these reactions. Currently, common factors sometimes refer to patients' responses as much as to therapist actions. Until we can say how re-moralization occurs and how it is affected by empathetic reflection etc., we ought not combine the two, but both may be nonspecific and both may be common--to different participants and to different degrees. This is a case where moderating and mediating variables need to be clarified.

Larry Beutler (UCSB)

Date: Fri, 8 Sep 1995 17:38:19 From: Stephen Ilardi <ssi@psych.colorado.edu> To: Larry Beutler <beutler@edstar.gse.ucsb.edu> Subject: More on common factors

On Fri, 8 Sep 1995, Larry Beutler wrote:

> The placebo analogy for discussing nonspecific or common factors bothers me<.

Jerome Frank has suggested that pill placebo likely owes its considerable efficacy to the same mechanism -- remoralization (or positive expectancy, hope, faith, etc.) -- hypothesized to underlie the efficacy of common factors in psychotherapy. Inasmuch as drug trials seek, in a (granted) clumsy manner, to parse out the nonspecific effects of each psychoactive drug by including a placebo control (presumably, a specific effect is demonstrated if the drug in question outperforms the placebo), I was merely suggesting that psychotherapy investigators may do well to distinguish, if possible, between those facets of the protocol which owe their efficacy to the mechanism of remoralization, and those facets which are efficacious due to some more "specific" mechanism (e.g., cognitive restructuring). I do not believe that psychotherapy, even a protocol which owes much of its efficacy to remoralization effects, is tantamount to placebo.

[Larry Beutler continues...]

> Moreover, I think it will be very difficult to distinguish between what you call "common factors" and what have been called "nonspecific factors"<.

I agree: the two terms tend to be used synonymously, and I would not seek to differentiate between them. The majority of investigators who have written about "common factors" or "nonspecific factors" usually refer to the same four constructs outlined by Frank over 30 years ago (rationale, ritual, healing setting, and emotionally charged relationship with designated "healer"). What I suggested was the need to distinguish between common (or nonspecific) factors and the "causal mechanism(s)" (remoralization) through which such factors mediate clinical improvement. The former refers to therapist behaviors, while the latter refers to a reaction on the part of patients. As you noted...

> It seems to me that we must differentiate between things that are therapist controlled and those that are not and between the reactions of patients (hope, remoralization) and the things that evoke these reactions<

So, I think we both agree that much unnecessary confusion has ensued from the practice of conflating therapist behaviors and patient responses, subsuming both under the rubric of "common factors".

[Larry Beutler continues...]

> Until we can say how re-moralization occurs and how it is affected by empathetic reflection etc., we ought not combine the two, but both may be nonspecific and both may be common--to different participants and to different degrees.<

Someone (perhaps Twain?) once remarked that "common sense is not that common", and I sometimes wonder if the same may not be true of common factors -- especially, as you allude to above, the phenomenon of empathetic reflection. Would you be in favor of dropping the "common factor" label from all therapist behaviors which cannot be empirically demonstrated to occur in virtually all therapies? I think it makes common sense.

My best,

Steve Ilardi, Ph.D.
University of Colorado at Boulder

Date: Fri, 8 Sept 95, 13.05 From: Lee B. Sechrest <sechrest@u.Arizona, EDU> To: Larry Beutler <beutler@edstar.gse.ucsb.edu> Subject: Re: CBT and nonspecfic factors (fwd)
--

On Fri, 8 Sep 1995, Larry Beutler wrote:
> see above

Someone else, sometime back, suggested that ordinary face-to-face therapy should be regarded as the placebo condition in all psychotherapy (generically meant) research. I agree. Any treatment that cannot do better than routine psychotherapy does not deserve further attention.

But Larry's view suggests that another useful approach in psychotherapy research would be to figure out how to enhance the nonspecific effects that appear to produce whatever change we find. We need a new and improved placebo intervention. Years ago (1966) Goldstein, Heller, and Sechrest reviewed research that appeared to bear on such questions as how interpersonal attraction and influence could be enhanced in psychotherapy, and Arnold G. did a fair amount of research on such problems. Maybe we ought to push the development of nonspecific therapy effects as far as they will take us. They may be the best thing we have going for us, at least for those who are going to insist on continuing to pursue the chimera of cure by conversation. ter therapist behaviors.

Lee Sechrest

Date: Wed, 13 Sep 1995 07:12:28 -0400 From: DonaldK737@aol.com To: psydoc@netcom.com Cc: hollon@u.washington.edu, DonaldK737@aol.com, fu03c2dj@fub46.zedat.fu-berlin.de, seligman@cattell.psych.upenn.edu, njacob@u.washington.edu Subject: consensus hollon/jacobson/klein

dear colleagues

At the end of 1994 an interpsych email discussion began concerning the research grounds for asserting the specific efficacy of psychotherapy. In particular the multi-site NIMH collaborative treatment study of depression (Elkin,Irene-PI) that incorporated imipramine/case management,placebo/case management, interpersonal psychotherapy and cognitive psychotherapy became a focus of discussion. The debate dealt with the particulars of that study but also general methodological concerns about the role of pill placebo controls, validation of treatment methods,sample definition,etc.

Several wellknown experts, e.g.,jacobson,hollon,barlow,seligman, mcnally chimed in. The editor of the J clinical consulting Psychology felt this was of sufficient interest that it should be published. As far as I know that is the first academic use of informal email debate. Hopefully, not the last.

The published format will be a statement by jacobson/hollon ,a critique by klein, commentaries by elkin and mcnally, and a wrap up by jacobson.

However Klein felt that rather than wait for publication(1/96?) it would be useful to the field to email out a mutually agreeable critical statement of areas of agreement and disagreement after discussion jacobson and hollon agreed and manuscripts were exchanged asymptotically. The reader may find the level of agreement surprisingly high. Some of our practical suggestions may improve the process of therapy evaluation. Our joint statement is below
donald f. klein

joint statement

1. We agree that imipramine was superior to psychotherapy (particularly cognitive behavior therapy) in the acute treatment of more severely depressed and impaired outpatients in the NIMH sponsored Treatment of Depression Collaborative Research Program (TDCRP).

Although we recognize that there were no differences between the treatment conditions with respect to the less severe patients, we do not necessarily agree about how this lack of differences should be interpreted.

Klein believes that it is possible that many of the less severe patients were atypical depressives for whom imipramine is relatively ineffective, although

both MAOIs and SSRIs have been shown to be specifically useful in this condition.

The analyses for atypical depression effects are feasible from the data but have not been pursued as yet.

Jacobson/Hollon are less inclined to believe that the less severe sample was comprised largely of atypical depressives or that they would have shown a superior response to a different type of medication, but agree that the possibility should be explored both with respect to the existing data and in subsequent studies.

However, we disagree about the relevance of differences in between-site treatment effects in tempering this conclusion.

Klein believes that he addressed the question of site differences with negative results in an earlier reanalysis of the TDCRP data (see Klein & Ross, 1993, Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report, Neuropsychopharmacology, 1993, 8, 241-252).

Jacobson/Hollon, on the other hand, believe that the TDCRP was "underpowered" to detect site by treatment interactions and suggest that the magnitude of the differences between the sites in the efficacy of CBT were large enough to warrant caution in interpretation despite the fact that these effects did not reach conventional levels of significance.

Although we disagree with respect to how these indications should be interpreted, we agree that the issue is an important one and regret that it has not been examined more fully since it was first raised by Elkins and colleagues in 1989.

2. We agree that none of the treatments in the TDCRP performed particularly well with respect to their ability to produce lasting recovery from major depression, but disagree about the significance of these findings.

Klein believes that the theoretical surprise is that the psychotherapies showed no prophylactic effect superior to that of placebo, whereas Jacobson/Hollon believe that the failure to replicate differences found in other studies favoring CBT over drugs in rates of relapse following treatment termination is consistent with the notion that psychotherapy may not have been adequately implemented in the TDCRP.

(Nobody expects drugs to have a prophylactic effect after termination, which is why maintenance medication is necessary for recurrent patients treated with pharmacotherapy only.)

Given that questions have been raised about the adequacy of implementation of psychotherapy between the sites, we concur that it would be instructive to

see actual rates of relapse for each treatment as a function of level of severity within sites.

3. We agree that the lack of difference between CBT and placebo during both acute treatment and follow-up would call into question the existence of any specific active ingredient in CBT, but disagree as to whether or not CBT was implemented in a representative fashion in the TDCRP.

Similarly, the fact that less severely depressed patients (who represented over half the sample) failed to show any indication of a "true" drug effect raises doubts as to whether imipramine has any specific effect with such patients or the adequacy with which it was implemented in the TDCRP.

Klein believes that CBT may have been adequately implemented in the TDCRP, whereas Jacobson/Hollon have their doubts. In particular, they note that only one of the sites had any particular history of experience with CBT (the same can be said for IPT) and that all three sites were under pressure to find therapists who would participate in the trial, whether those therapists had any particular experience with or allegiance to the modality to which they were assigned. Moreover, although the CBT supervisors were quite competent and initial training during the pilot phase appeared adequate, the bulk of the training was provided off-site and the frequency of supervision decreased markedly once the therapists graduated into the study proper. In this respect, the TDCRP was as much a test of the feasibility of training as the efficacy of CBT (or IPT). (At the very least, the TDCRP showed that it may be harder to learn to do CBT in a competent fashion than had previously been suspected, something that may have implications for the ease with which it can be exported to applied clinical settings.)

In the same vein, it would have been helpful had greater detail been provided in the original report regarding medication dosage and schedules, particularly with respect to precisely how problems with side effects, nonresponse, and noncompliance were handled.

This would have facilitated determining whether the less severely depressed patients treated in that study were truly not pharmacologically nonresponsive, as opposed to the possibility raised by Klein that pharmacotherapy was less than optimally implemented.

4. We agree that the inclusion of pill-placebo controls in evaluative studies comparing drugs and psychotherapy is extremely desirable. However, we disagree about the interpretability of drug versus psychotherapy comparisons in the absence of a pill/placebo comparison group.

Jacobson/Hollon believe that several studies exist in which both pharmacotherapy and CBT have been adequately implemented; in those trials, drugs and CBT typically have been comparably effective.

Klein believes that since these are all samples of convenience, it remains moot just what equivalence in these studies demonstrate. Perhaps neither

treatment was working. We agree that those studies that have suggested that drugs were inferior to CBT in reduction of acute symptoms typically have implemented pharmacotherapy in a less than adequate fashion.

Although we disagree somewhat on the interpretability of existing drug-psychotherapy studies that lack pill-placebo controls, we are largely in agreement on the desirability of including such controls in future comparisons. As Klein has pointed out repeatedly over the last two decades, in the absence of such controls, it is far more difficult (Klein would say impossible) to determine whether pharmacotherapy was adequately implemented or the representativeness of the sample (assuming that it has already been determined that a particular population is indeed pharmacologically-responsive).

In most cases, drugs are chosen for comparison to psychotherapy that have already been shown to have a specific pharmacological effect in the population of interest. However, it is rarely the case that even a well-established effect is so robust that it can survive inadequate implementation or so universal that it holds for all individuals in the population. That being the case, it simply makes sense to include a pill-placebo controls in drug-psychotherapy comparisons, both because it provides a manipulation check for the adequacy with which pharmacotherapy was implemented in that particular study and because it is potentially informative about the nature of the specific sample selected.

Although we concur that such a strategy is at least desirable, we also agree that it is not sufficient, and further agree that the quality of pharmacotherapy (and psychotherapy) must be evaluated independently of whether drug/placebo differences are evident in trials that do include pill-placebo controls,

5. We disagree as to whether CBT can be considered a viable alternative to pharmacotherapy in the treatment of more severe outpatient depression.

Klein believes that medication is not only likely to be more effective with such patients, but also faster and cheaper.

Jacobson/Hollon, on the other hand, point out that the TDCRP is the only study that has found an advantage for drugs over CBT in the treatment of such patients (including speed of response) and that indications of an enduring effect for CBT found in other studies, if true, may call into question the notion that drugs are necessarily less expensive than psychotherapy in the long run.

We concur that replication is necessary for definitive conclusions and that no single study, not even one so large as the TDCRP, is sufficient unto itself, particularly when its findings are not consistent with other studies in the literature. On the other hand, we also concur that it would be grossly premature to conclude that CBT does indeed have a prophylactic effect, since

none of the relevant studies have yet ruled out differential retention during the acute treatment phase as a rival plausible alternative (that is, that "sicker" patients, at greater risk of symptom return, also may be more likely to complete and respond to drugs than to CBT).

6. We agree that behavior therapy is a promising intervention worthy of proper evaluation in the treatment of panic disorder. However, we disagree about the efficacy of CBT and the solidity and experimental validity of existing studies evaluating this approach for such patients.

In this regard, Klein emphasizes that very high placebo response rates have been reported for this diagnostic group, particularly with regard to diminished panic frequency or cessation.

Jacobson/Hollon note that CBT has been found to be superior to a variety of comparison conditions in those earlier studies, the majority of which should have been at least as effective as any as type of pill-placebo or nonspecific control.

Klein points out, however, that these earlier studies demonstrating CBT superiority were conducted by CBT originators, which raises the problem of allegiance effects. Studies by non-therapy originators (Shear, Ost) have found CBT equivalent to non-cognitive therapies.

7. We agree with regard to the value of large scale, multi-site clinical trials, but disagree as to their potential risks.

Klein considers such multi-site comparative trials as absolutely essential, since he believes that they are the only way to deal with allegiance effects.

Jacobson/Hollon agree that multi-site trials can be of value, but believe they are open to misinterpretation unless potential allegiance effects are carefully balanced within each site.

8. We agree that much that passes for scientific discussion is often little more than a partisan debate between various interest groups (something we have been guilty of ourselves) and that the utmost rigor and completeness in data presentation and analysis is therefore required.

We thank the authors of the TDCRP for making their data available to the field for independent analysis. We recognize that the study authors state that they feel an ethical obligation not to link site to pattern of response, but do so with deep regret, since such identification is necessary to determine whether allegiance effects may have operated in that study.

In this respect, it is a curious and vexing paradox that we are less able to evaluate the possible impact of allegiance effects in the TDCRP than we are

with respect to other single site studies, despite the laudable openness of the study authors with respect to making the bulk of their data available.

One possible solution for this conundrum would be for the study therapists to release the authors from their pledge not to identify the sites and we call on them to do so, for the good of the field. Anything less will only contribute to the promulgation of science by rumor and undercut a willingness to make data from the study available to the larger field.

Date: Wed Sep 13 18:30:43 1995 From rcarson@acpub.duke.edu To: Neil Jacobson <njacob@u.washington.edu>
--

Gentle persons:

Thank you for this excellent synopsis of the outstanding issues in the TDCRP study, which happen to be those central to the entire drug vs. psychosocial intervention controversy.

Bob Carson

Date: Wed Sep 13 20:14:16 1995
From njacob@u.washington.edu

Dear Colleagues:

Earlier today, I forwarded you a consensus statement, drafted initially by Don Klein, indicating areas of agreement and disagreement in our JCCP exchange. As Don indicated, this debate began on the internet, and, largely due to Larry Beutler's tenacity and creativity, was converted into a debate that will be published in JCCP. Also, as indicated by Klein, Steve Hollon and I wrote the lead article, which Klein responded to. Rejoinders to both Jacobson & Hollon and Klein were written by Elkin and colleagues and by McNally. Steve Hollon and I concluded with a wrap-up piece. Both of the Jacobson and Hollon pieces have now been accepted in final form, and as I understand it, with any luck the exchange will appear in the February issue of JCCP. Although I have no authority or access to the final versions of the Klein, Elkin et al., or McNally articles, I would be happy to send along preprints of the two Jacobson and Hollon pieces for anyone who is interested. Instead of requesting those pieces via e-mail, however, please send any requests to my Program Coordinator: Joan Giacomini, University of Washington, Department of Psychology, Center for Clinical Research, 1107 N.E. 45th St., #310, Seattle, WA 98105.

Neil J.

Date: Wed Sep 13 21:07:41 1995
From rjm@wjh.harvard.edu

Dear Colleagues:

Like Neil, I am happy to provide preprints of the final version of my article, entitled "Methodological Controversies in the Treatment of Panic Disorder." As he mentioned it is part of the JCCP debate/series (nb: It is solely a critique of Klein's critique of CBT panic treatment studies; it does not concern depression at all).

My address is:

Richard J. McNally, Ph.D.
Department of Psychology
Harvard University
William James Hall
33 Kirkland Street
Cambridge, MA 02138

Date: Thu Sep 14 17:51:29 1995
From k-howard@nwu.edu
To: Neil Jacobson <njacob@u.washington.edu>

Neil,

Could you just fill us in on some little details about the Collaborative study?

It is my understanding that less than 50% of randomly assigned patients completed any of the four comparison conditions (about 65% completed at least 12 of the scheduled 16 weeks). So we are left with the comparison of four self-selected groups.

There were a number of outcome criteria. Were these combined in a way that would allow for the determination of whether a patient improved or not (or reached some clinical-significant change criterion)? If so how many patients in each condition improved? How many improved in the analyses in which severity was used to disaggregate?

Ken Howard

Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Thu Sep 14 20:10:55 1995
From njacob@u.washington.edu
To: Ken Howard <k-howard@nwu.edu>

Dear Ken,

Your questions are good ones and some are even crucial. In our lead article for the JCCP exchange, Steve Hollon and I looked carefully at the published data, the published reanalyses, and the other research-not only in the TDCRP, but also in CBT for depression and panic. I cannot possibly do justice to this summary now, although it appeared on the internet earlier in abbreviated form. Similarly, when we received the actual data from the TDCRP investigators, we performed some analyses of our own, which we summarized in the second Jacobson & Hollon article, which serves as the concluding article in the exchange. We are still analyzing data from the TDCRP tape. As I said yesterday, those pre-prints are available via my Assistant, Joan Giacomini, at the address posted yesterday, and they deal with some of the issues that you raise.

As for the questions regarding drop-out and self-selection, you are correct. Those are serious problems, as they are in many if not most randomized clinical trials. I refer you to the Shea et al. (1992) article published in "Archives", in which the TDCRP investigators report on the 18-month follow-up.

Table 1 in that article makes it clear that none of the treatments performed impressively from the standpoint of clinical significance. In fact, the probability of completing treatment, getting well, and staying well for 18 months following the 16 week treatment ranged from 19% with Imipramine to 32% with CBT. As the mother-in-law of a close friend and esteemed colleague once said when presented with these data, "And they pay you for this?"

Neil

Date: Fri Sep 15 17:05:00 1995 From k-howard@nwu.edu To: njacob@u.washington.edu
--

Neil,

Part of my previous post was aimed at determining whether it was going to be worth spending time reading my esteemed colleagues' thoughts regarding reanalyses of data or comments on findings from the collaborative study. It is my understanding that, at most 120 patients completed the research protocol (and that these patients were not representative of any important patient population, certainly not Depressives in clinical sites). That would be 40 at each of the three sites and 10 per treatment per site (four treatment conditions). Then, there was disaggregation on the basis of initial severity -- let's say 50% again. So, there were five severe patients at each site treated in each condition, or fifteen in all (across sites). According to Klein, the 15 patients who received Imipramine did better (had a statistically significantly higher mean on some outcome variable[s]) than did the 15 patients who received CBT. Is this really worth all the effort? What are we to make of such a finding? Did any patient actually improve in either condition? How could there be across site effects with samples this small?

Ken Howard

Date: Mon, 18 Sep 95 9:48:56 CDT
From: Lili Pintea-Reed - Adjunct Psychology <lilip@cuok.cameron.edu>
Subject: Re: consensus hollon/jacobson/klein
To: DonaldK737@aol.com

All of this points out the difference of care of acute illness and chronic illness. An analogy would be, if you have pneumonia you need some anti-biotics, but once you are over the acute illness one has to look at long-term health changes needed to keep the pneumonia away. I think mental illness fits in this model. Acute episodes certainly will (in my clinical experience) show quick response to medication, but long term work is needed to stop reoccurrence..

In a nutshell acute episode vs. chronic problem....

My 2 cents for the day...Lili

Lili Pintea-Reed, MA
Adj. Prof. Psy.
Cameron U

Date: Mon, 18 Sep 1995 20:35:31 -0400
From: DonaldK737@aol.com
To: lilip@cuok.cameron.edu

Lili Pintea-Reed, MA Adj. Prof. Psy. Cameron U makes the argument that chronic illness requires "long term work is needed to stop re-occurrence" perhaps so but data is lacking so it shouldn't be assumed that lengthy, expensive treatments are indicated. the TDRCP study showed no more prophylactic effect than placebo. Its unfortunately a unique body of data.
don klein

Date: Mon, 18 Sep 1995 11:18:21 -0400
From: stricker@sable.adelphi.edu
To: sscpnet@bailey.psych.nwu.edu
Subject: outcome measures

A man went to his physician and was found to have high blood pressure. The physician administered one of several empirically validated medications and the blood pressure fell. Treatment was a success. However, as often happens with these medications, his sexual functioning was impaired so he stopped taking the medication, his blood pressure rose, and treatment now was a failure (an example of efficacy vs. effectiveness). If he continued to take the

medication, the blood pressure would have stayed low, but would treatment have been a success?

Different hypertensive man goes to a psychologist, is given an empirically validated treatment regimen, follows it, his blood pressure falls, and treatment was a success. However, as time goes by, he relaxed his vigilance, his blood pressure rose, and treatment now was a failure (another example of efficacy vs. effectiveness).

To make it more complicated, as with all aggregated data, some people will not show the general effect. Some successful cases will live a life filled with diminished sexual capacity or excessive jogging (neither sound great to me) and then die early anyway (This doesn't even get into anomalies like getting mugged while jogging). Others will show the effect and live longer in order to get Alzheimer's. Is the proper measure blood pressure, longevity, or quality of life?

If we have this trouble with a simple function like blood pressure, what if the goal (at last, the point) is to change a human life? Patient satisfaction does not correlate well with other measures, such as status, but which is the key measure. Perhaps status doesn't capture the reason why the patient is in treatment, or perhaps satisfaction is subjective and unrelated to "real" change.

Another example - A patient enters treatment with dysphoria and discovers that he is experiencing strain from a marriage and family previously thought to be satisfying, and so he gets a divorce. Is the treatment successful? The patient, immediately after treatment thinks so, but in the long term may find himself missing his former satisfaction, or may end up in another relationship that cause strain. The symptom may or may not change, depending on the balance between strain and satisfaction. The wife and family surely think treatment failed, as their lives have been disrupted. The therapist may or may not think treatment succeeded, depending on how the decision was made (and on her own experiences with intimate relationships, but that is another story). Friends and family will vary depending on their feelings about the wife and their own relationships. The insurance company will vary, depending on how long it took to make the decision. "Society" will vary, depending on whether taxes are increased because of the loss of a joint tax return or there is an increase in welfare payments. So, which is the proper measure of outcome? They will not converge because they reflect different issues, each of which may be valid in its own right. Perhaps we will have to get into values, truth, reality, and world views, but those are metaphysical concepts that we like to avoid. Or perhaps we will have to respect each of those measures rather than arguing for the "real" superiority of any of them, and settle for a descriptive account of the multiple effects of psychotherapy.

George Stricker

George Stricker
Internet address (STRICKER@SABLE.ADELPHI.EDU)

Telephone (516) 877-4803
Fax (516) 877-4805

Date: Tue, 19 Sep 95 09:51:00 EDT
From: "Ogles, Benjamin" <OGLESB@email1.msmail.ohiou.edu>
To: SSCPnet <sscpnet@bailey.psych.nwu.edu>
Subject: NIMH- Clinical Significance

I have been lurking on SSCPNet for awhile and this is my first attempt to join the discussion. My apologies if my message doesn't make it.

We (Ogles, Lambert, and Sawyer (JCCP, 63, 321-326)) recently published a re-analysis of the NIMH data using Jacobson's clinical significance methods. When looking at the completer sample (all those completing at least 12 sessions and 15 weeks) using the Beck Depression Inventory, HRSD (Hamilton rating scale for depression), and Symptom Checklist (HSCL), there was remarkable concordance among measures in terms of identifying subjects who made clinically significant change.

34 clients were identified by all three measures as not making clinically significant change and 84 were identified by all three measures as making clinically significant improvement. That is all three agreed for 73% of the 162 subjects. Measures did not agree for the remaining subjects.

Observed Frequencies for the Loglinear Model.

Significance Status Treatment Group						
BDI*	HRSD	HSCL	CBT	IPT	Imipramine-CM	PLA-CM
NS	NS	NS	11	5	6	12
NS	NS	Signif	1	6	4	1
Signif	NS	NS	0	1	0	0
NS	Signif	NS	3	1	0	1
Signif	NS	Signif	1	2	1	1
NS	Signif	Signif	5	5	3	6
Signif	Signif	NS	0	0	1	1
Signif	Signif	Signif	19	27	23	15

* Row one reports the number of individuals in each treatment group identified by all three measures (BDI, HSCL, & HRSD) as having non-clinically significant change. Row two reports the number of individuals identified by the HSCL as meeting the criteria for clinically significant change, but as non-clinically significant change by the BDI & HRSD. . . . Row eight reports the frequencies when the three measures agree that the subjects clinically improved.

** total Ns for the completer treatment groups - CBT = 40, IPT = 47,

Imipramine- CM = 38, PLA-CM = 37.

(I hope this table makes it into cyberspace with real columns).

Another interesting point made in the article is that some patients deteriorated (up to 3 who completed the full course of treatment and about 11 of the original 239 who started treatment). This corresponds to earlier literature indicating from 5 to 10% deterioration in outcome studies.

Date: Wed, 20 Sep 1995 22:45:14 -0400 From: PSYCHLING1@aol.com To: PsyNetUSA@aol.com Cc: stricker@sable.adelphi.edu, sscpnet@bailey.psych.nwu.edu Subject: Re: N] Explorations of issues in Outcome measurement

George -

What's your point? You describe much. You state more.

You state that there are several metaphysical concepts that we do not like to get into. I disagree. Those are things that we DO like to get into. Psychotherapists do this. So does the rabbi, the teacher and the priest/ess. We get into that with delightful and enriching friendships. We struggle through these as we share life, time and experiences with one another. We feel the deepening of these things when we know loneliness, ecstasy, tragedy, wonder.

I guess I have a hard time with your comments because I am focusing on something dreadfully narrow: what will the (third-party) person who is paying for this therapy consider worthy? It is unlikely that the third party will pay for metaphysical musings, though these may be the more fundamental meanings.

Yet, this is a very important point: helping persons through hard times does not mean that they will get to perfection/fulfillment on the other side? It just means that they got through this hard time. And they could have navigated an entirely different route and reached an entirely different destination. Better they should get through the hard time than they should not. };>

- Dan Fallon / Chicago
Psychling1@aol.com

Date: Sun, 24 Sep 1995 09:59:38 -0700 (MST)

From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Facilitated Psychotherapy (?)

Lee Sechrest, Univ. of Arizona

The following consists of a series of quotes or paraphrased statements from what I think is a truly important article that all of you should read--and recommend widely to others. The topic is "facilitated communication" (FC) and how it came to be widely popular and disseminated in the treatment community.

In this part of what I am going to say, I will refrain from comment. Probably no comment is needed, but that will not keep me from a second posting in which I will make some additional observations.

Jacobson, J.W., Mulick, J.A., and Schwartz, A.A. (1995). A history of facilitated communication: science, pseudoscience, and antiscience. American Psychologist, 50, 750-765.

How do controversial treatments come to be adopted and their premises so widely accepted?. The critical elements, we believe, are an unexpected but apparently dramatic treatment response, a superficially plausible "theory" for this effect, and a disavowal by the proponent of conventional standards of scientific procedure and proof." (p. 752)

These early, and foundational, studies of FC used naturalistic observation, ethnographic methodologies, no standardized measures at baseline or post-treatment

"Some of this literature explicitly criticizes 'psychologists' for holding people with disabilities to unreasonable (i.e., objective) standards for demonstrating their communications skills." (p. 754)

"...noteworthy is that some communications disorders and special education professionals, their students, and some parents of people with severe disabilities began to promote the use of FC through publication and training opportunities in the absence of a substantive scientific foundation...Much of the publication of articles promoting the use of FC has also occurred outside of the peer-reviewed literature; popular magazines, newspapers, television, and newsletters...." (p.754)

"The phenomenon of FC satisfies many of the criteria of pseudoscience: Demonstrations of benefit are based on anecdotes or testimonials; baseline abilities and the possibility of spontaneous improvement are ignored, and related scientific procedures are disavowed; and therapists who use FC unsuccessfully are blamed for not doing it correctly or not believing it will work." (p. 757)

"Controlled research that disconfirms the phenomenon is criticized on grounds of rigor and procedure not even remotely addressed by studies that purport to demonstrate the effectiveness of the technique...Selected findings, rather than the composite findings, of controlled studies are critiqued and interpreted." (757)

>From a book on FC (Biklen, 1993) "Some critics of facilitated communication appear to approach it from a positivist perspective. The tradition of positivism attempts to address social science questions as if human experience could be understood in cause and effect terms, much as one understands the natural science of physics. This perspective contradicts the perspective to which I hold, which is that objects, events, observation, and understanding are by definition socially constructed."

"In the brave new world of postmodern analysis, terms from the physical sciences, like the uncertainty principle, chaos, and paradigm shift are loosely and often inaccurately used in an Alice-in-Wonderland revision of scientific method." (p. 670)

Date: Mon, 25 Sep 1995 09:43:39 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Facilitated Psychotherapy. Part I
--

Lee Sechrest, Univ. of Arizona

This posting is, obviously, a continuation of my last one that quoted extensively from an American Psychologist article on facilitated communication. Equally obviously, I meant to suggest that that article should serve as a warning, a "wake up call" to use the cliché du jour, to the field of psychotherapy research and advocacy.

A while back, my esteemed colleague George Stricker, the gentleman from Adelphi, stated his "belief that efficacy research is inadequate as a sole source of data to inform clinical practice. I want to make it clear that I am not attacking George specifically here, and I reject any inference that I associate him with such nonsense as facilitated communication--or even facilitated psychotherapy. His comments do, however, provide a convenient stimulus for some that I want to make.

First, I do not know anyone who has ever said that efficacy research *should be* the sole source of data to inform clinical practice. Efficacy research is certainly never alleged to be the sole source of data to inform medical

practice, and it is in medicine that the concept of efficacy research evolved. Efficacy research has a narrow, but specific function.

George goes on to say, "Efficacy research is consistent with Plato in its belief in ultimate truth, but not as to the methods by which we arrive at it." That is undoubtedly true. Efficacy research evolved in the context of scientific inquiry, about which Plato was uninformed. The method of introspection "rules out" experimentation. That is simply true by definition and says nothing about the usefulness of either method. Scientific method can be and should be limited in its definition; not every approach to inquiry and truth is science. Not every scientific method is appropriate to efficacy research.

Then George says, "I wish we would do more effectiveness research as well as (not instead of) efficacy research." Here is where the rub comes in. Until the *efficacy* of a treatment (or a diagnostic maneuver) is established, there is no basis for effectiveness research. Facilitated communication jumped right over the efficacy obstacle. It was precisely efficacy that was never established until the appropriate and telling experimental trials were done. Once those were done, the whole "house of cards" should have collapsed. That it did not is an indictment.

Is there sufficient evidence for the efficacy of psychotherapy to move right into effectiveness trials? I think not. If there were, we would not be having all these discussions. The very fact that the major attempt to establish efficacy of treatment for depression several years later requires the kind of debate that has been going on is testimony to the lack of solid evidence for efficacy.

George's complaint, however, about the need for more effectiveness research is strange. There is a tremendous amount of "effectiveness" research on psychotherapy. It is efficacy research that is in woefully short supply. Ken Howard has alleged (I am being journalistically cautious) that there are more than 500 studies on the "outcome" of psychotherapy. Only a handful of those, however, are successfully realized experimental trials with random assignment and good clinical outcome measures.

OK, enough for this time. Part II is coming. Watch this space.

Lee Sechrest, Univ. of Arizona:

Date: Mon, 25 Sep 1995 11:31:02 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Facilitated Psychotherapy: Part II

Lee Sechrest, Univ. of Arizona:

This is a continuation of my comments on "facilitated psychotherapy." This series is longer than I had intended, but your delete key should still be working. Part III will be the last, I promise you.

The TDCRP study has been criticized by any number of observers, recently by Ken Howard on this network on the grounds, partially, that attrition was so extensive that the "experiment" did not end up being an experiment at all, but a quasi-experiment. If he is right, and I suspect that he is, then we have N-1 efficacy trials of psychotherapy. I am more than a bit bothered by continued attempts to resuscitate the moribund TDCRP on the grounds that one or the other therapies, whichever one is being defended, was not done properly. (See the quotes from the article on facilitated communication.)

Now here is the rock bottom, fundamental truth: the fact that a strong research design cannot be implemented does not make the results from a weak design more plausible.

Years ago Kiesler, somewhat prematurely I think, announced the victory of psychotherapy, claiming that it had clearly been demonstrated to be successful and that all that remained was to determine for whom, by whom, for what problems. Maybe that is something of what George Stricker had in mind in wishing (on a star, George?) for more effectiveness research.

But what research methods and designs are going to be used for that research? To find out for whom psychotherapy is most effective, won't it be necessary to be able to assign persons randomly to conditions of treatment? Will the problem of who is effective in doing therapy be resolved without strong experimental designs? Can, as George Stricker implies, such issues be resolved by observation and introspection? I don't think so.

Moreover, these are not problems that can be resolved among ourselves. "We," meaning us academicians and researchers might get together and develop some sort of a West Bank agreement that we are going to eschew the rigors of randomized experiments and are going to start believing in a lot of other methods. That does not mean, however, that anyone else is going to honor our agreement. Review committees, funding agencies, insurance companies, and our colleagues in other fields, let alone our colleagues in our own field, are not going to fold their tents and steal away from the arena when we show that patients who volunteer for treatment do better than those who do not or that psychotherapists who are asked to introspect on their activities conclude that they are highly effective even if in ways that cannot be detected by current methods of assessment.

Take a look at the history of facilitated communication.

Date: Mon, 25 Sep 1995 16:05:08 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Facilitated Psychotherapy: Part III
--

Lee Sechrest, Univ. of Arizona

This is Part III of my comments on facilitated psychotherapy and, as I promised in Part II, the last part.

I am as sympathetic as anyone to the problems involved in doing research in natural settings. I teach a course on methods for research in the field that I subtitle "Soft methods for hard problems." I urge, however, that we employ the best, strongest methods available for the research we do. Too often, by my observations, researchers settle for far less than the best designs they could use or for far weaker measures than are possible.

We cannot *make* anyone believe our research findings. They will be acceptable and believable in proportion (although $r < 1.00$) to their methodological integrity. I believe that our field is characterized by a good bit of whining about the reluctance of our colleagues, the health establishment, the public to "trust us" because we are doing the best we can (we are not) and because we have such good intentions (which I will grant). We ought to get on with the business of improving our research methods, with toughening ourselves to produce the best possible research rather than the most expedient, and with development of integrative theory that will enable us to tie our very diverse ideas and findings together in a plausible package.

In the meantime, I suggest the application of Sechrest's Methodology Assessment Crucible (SMAC) to all our proposals to bring light and truth to our field. SMAC asks a simple question: would the methods we propose for examining our own problems be acceptable to us for examining the problems of other people with whom we disagree? For example, I would suggest that the SMAC for psychotherapy research should include facilitated communication, astrology, religious observances, marijuana, and so on. So if we are going to take seriously the results of weak designs in support of the effectiveness (let alone the efficacy) of psychotherapy, then we ought to be prepared to accept the results of those same designs for virtually any other interventions for the same kinds of problems.

One more comment about facilitated communication and facilitated psychotherapy. Near the beginning of this network we had a discussion about the "necessity to do something" on the part of psychotherapists dealing with people with difficult problems. I suggested in a posting "Don't just do something; stand there!" that the honest thing to do might be to do exactly nothing in some instances. Or at least nothing specific.

Imagine the plight of workers dealing with autistic or otherwise seriously impaired children whose parents desperately want something done for them. Well, why not facilitated communication? After all, one must do *something*!

Date: Mon, 25 Sep 1995 18:14:27 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: maddis@vms.clarku.edu Subject: Re: Response to Lee

I think Michael Addis makes an excellent point. In the NIMH TDCRP, where supervision was not continued once therapists met competence criteria, performance (based on outcome) was generally worse than in other trials where supervision was continued throughout the study. There is good reason to believe that clinical trials provide an optimal estimate of treatment efficacy, because of this close supervision and adherence/competence monitoring. It is safe to assume that "efficacy" overestimates "effectiveness" in naturalistic settings. Efficacy research may be necessary, but it certainly is not sufficient.

Date: Mon, 25 Sep 1995 20:24:47 -0500 (EST)
From: maddis@vms.clarku.edu
Subject: Re: Response to Lee
To: Bill Follette <follette@unr.edu>

Bill,

It seems to me there is an important distinction missing from the debate over whether clinical practice should be guided only by empirical research. While Bill Follette is entirely correct that the consumer should be informed whether the therapist's approach is empirically valid, what about the all too common grey area? i.e. where a therapist is, for example, cognitive-behavioral in orientation but does not practice Beck's treatment for depression according to the letter of the law. Should this therapist inform the client that his or her practices are "generally" governed by empirical evidence? That seems a bit ambiguous. In reality, we have no data demonstrating the extent to which practicing clinicians are adhering to empirically validated treatments as practiced in controlled (hopefully) clinical trials. Self-report would be a first step but would by itself be suspect. Thus, it is not clear to me at what point any practicing clinician can claim (at this point) to be practicing an empirically validated treatment approach.

Food for thought,

Michael Addis

maddis@vax.clarku.edu
Department of Psychology
Clark University
950 Main St.

Worcester, MA 01610-1477

Date: Tue Sep 26 19:51:45 1995
From: "Stephen M. Saunders" <SAUNDERSS@vmsa.csd.mu.edu>
To: sscpnet@bailey.psych.nwu.edu
Subject: efficacy and effectiveness

Perhaps readers more familiar with clinical trials can answer the following. Is it possible to offer the same treatment to patients who succeed and to patients who fail? It seems to me that the progress of the patient would interact (feedback, confound, influence, determine) with the progress of therapy. If this is true, then it must be concluded that patients who succeed received a different therapy than patients who failed. Alternatively, if the same therapy is offered to both, then progress or lack thereof must be ignored, which must be extremely difficult for caring persons (i.e., therapists).

This is, I think, related to the issues of efficacy and effectiveness.

(While we're on the topic: Can anyone point me to a clinical trial wherein abatch of therapy data was dropped because therapy integrity was compromised?

I'm not familiar with any such literature.)

Stephen Saunders
Marquette University

Date: Tue, 26 Sep 1995 13:54:15 -0400 From: stricker@sable.adelphi.edu To: sscpnet@bailey.psych.nwu.edu Subject: One more time

This note is motivated by two separate issues. First, I will be leaving for Toronto Thursday morning and will be out of touch with my computer for 4 days, so I don't want to let this thread move at a leisurely pace. Second, Lee seemed to imply that I am devoted to introspection and observation as approaches to science. I am happy to plead guilty to my regard for introspection and observation, but not as substitutes for experimentation. I would like to propose two exercises.

Exercise I -

Take a piece of paper, divide it into three columns, and then write at the top of the columns the words thinking, observing, and experimenting. Then think of the great scientists throughout history, what their major contribution was, and put the name in the proper column. Which one fills up first? Last? Now do the same for psychologists. Start with the most universally applauded person (and deservedly so) on this Net, Paul Meehl. Which column does he belong in ? Try the group gadfly, my old and good friend (and this is not sarcastic), Lee Sechrest. Where would his contributions place him? Perhaps the division of science into thinking, observing, and experimenting isn't so easy to do, or so fruitful.

Exercise II -

Take another piece of paper, divide it into two columns. and, in the left hand column, write the names of the great scientists. In the right hand column, write the experiment associated with the name. I'll go first. Isaac Newton, arguably the greatest genius of all. In that 18 month period that he was away from Cambridge because of the plague, when he formulated the law of gravity, the laws of thermodynamics, the calculus, and the foundation of optics, did he do any experiments? Albert Einstein, whose name is synonymous with intelligence in the 20th century. Isn't he the one who, when brought the results of research that confirmed relativity, said rather disinterestedly, that the research couldn't have been otherwise because the

theory was accurate? And wasn't that research observational? How about the three men, all of whom considered themselves scientists, who revolutionized the way people saw themselves and their place in the universe - Copernicus, Darwin, and Freud. Plenty of thinking, a good deal of observation, but not much in the way of experiments.

Lest I be misunderstood, this is not a polemic against science. It has brought knowledge further in the past 350 post-Baconian years than was accomplished in the previous thousands of years of recorded history. It also is not a polemic against the experimental method. Experiments have played an important role in scientific developments. It is a plea not to discard other sources of scientific advance, and to respect the interrelationships among thinking, observing, and experimenting. There are many approaches to knowledge, and they should be viewed as complementary rather than competitive.

George Stricker

George Stricker Internet address (STRICKER@SABLE.ADELPHI.EDU)
Telephone (516) 877-4803
Fax (516) 877-4805

Date: Tue, 26 Sep 1995 22:42:46 -0400 From: DonaldK737@aol.com To: SAUNDERSS@vmsa.csd.mu.edu Subject: Re: efficacy and effectiveness/treatment integrity

dear colleagues,

I know of no trial rejected for lack of treatment integrity. in the usual drug vs placebo trial a significant finding is taken as a warrant that some patients resembling the entry criteria wil have specific pharmacological benefit. if the treatment was poorly done then the significant effect size is viewed as an underestimate so nobody cares. the headache comes from comparative treatment trials where if $a > b$. perhaps b was poorly done as b aficianados will hasten to point out. i am reanalysing the NIMH Collaborative Treatment of Depression Program regarding medication.. it is evident that the manual guidelines were varied from greatly with regard to medication . it will really be hard to estimate if this matters.after all perhaps it was the difficult cases that caused protocol variation. we will try to see.

over one third of the placebo patients did not achieve the prescription guidelines!!

note there has been no statement in some 13 years about the integrity of this study's psychotherapies and this data is still not publically available. clearly in comparative trials protocol variations must be explicitly identified and

explained and built in to the analytic process. we need to progress to this design necessity.

cordially
don klein

Date: Fri, 29 Sep 1995 08:25: From: Neil Jacobson <njacob@u.washington.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: Jacobson's comments on Lee's Comments on..

I once published such a study in Behavior Research and Therapy on a comparison of those two conditions in BMT. The flexible group did better.

Date: Fri, 29 Sep 1995 09:39:39 -0500 To: sscpnet@bailey.psych.nwu.edu From: k-howard@nwu.edu (Ken Howard)
--

We seem to have muddled the waters again:

Efficacy research shows that under some highly specified conditions, on average, an intervention CAN be shown to produce better results than some comparison condition.

Effectiveness research inquires whether, on the average, the intervention DOES do better than some comparison condition. (Service research asks "How many people are helped, at what costs?")

Ken Howard

Date: Fri, 29 Sep 1995 09:30:
From: k-howard@nwu.edu (Ken Howard)
To: sscpnet@bailey.psych.nwu.edu
Subject: Jacobson's comments on Lee's Comments on..

In a recent message, Neil stated that "This accountability issue in efficacy studies appears to override the factors that you mention: clients getting to choose their treatments, ending treatment when they want to, etc."

As usual, Neil has made a brilliant insight into how we could advance our discussion. Let's do some empirical studies of these methodological issues.

The paper by Greenberg et al. in JCCP (in which a comparison is made between results of studies in which the blind had been compromised or not) is the kind of work that will lead to some leveling of the playing field in the sense that no one methodology is likely to be sufficient.

I have tried to voice my concern that reliance on the ideal randomized clinical trials, for a host of reasons, may be unfounded. That we should be prepared ahead of time for secondary analyses that will be inevitable given the compromised execution of such trials. That some variables (e.g., sexual abuse vs. no sexual abuse) could not be randomized in any case. Etc. All of this was meant to say that we need lots of ways of establishing facts.

Ken Howard

Date: Fri, 29 Sep 1995 10:44:
From: Seligman@CATTELL.PSYCH.UPENN.EDU
To: k-howard@nwu.edu (Ken Howard)
Subject: Re: Jacobson's comments on Lee's Comments on..

Even better, how about a design in which half of the whole group of subjects are placed in random assignment, fixed duration, manualized psychotherapy of some kind, and the other half choose therapy and therapist, the duration is tailored to their progress, and the therapist can correct the trajectory of techniques which fail--all for some well-defined disorder.

Neil, Lee, Ken...wanna place your bets on how this will come out empirically?

marty seligman

Date: Fri, 29 Sep 1995 10:59:51 -0500
To: seligman@CATTELL.PSYCH.UPENN.EDU
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: Jacobson's comments on Lee's Comments on..

<see above>

THIS IS EXACTLY THE KIND OF EMPIRICAL APPROACH THAT
NEIL'S COMMENTS SUGGESTED WE PURSUE.

Ken Howard

Date: Fri, 29 Sep 1995 12:35:36 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: k-howard@nwu.edu (Ken Howard)
Subject: Re: Jacobson's comments on Lee's Comments on..

Ken,

I have some concerns re a host of comments you and others (e.g. in the "efficacy versus efficiency debate") have made about randomized trials and secondary analysis.

I have always believed--and was reinforced in this belief by the articles in the June 1994 issue of Science (1538-1541) that assignment in randomized trials must be defined in terms of *intention to treat.*

If, for example, practitioners using type X therapy have some variance in how they do it, well that is part of the assignment to type X therapy. If, for example, an exercise program is so grueling that about 90% of those assigned to it drop out, well that is an effect of assignment to the program, etc.

I'm not just being "purest." The problem is that defining "treatment" in any other way presents insoluble logical problems. Suppose, for example, we define group therapy treatment as actually showing up for group meetings. That sounds reasonable. But not all do. OK But :(i) we have no idea who in the control group would versus would not have shown up had they been assigned to the experimental group therapy group, and (ii) if in fact there is some benefit of group therapy that occurs only for those who show up, then since they form a subset of those assigned to group therapy, the entire set as a whole of those assigned to group therapy will be better off than the entire control group. There might be a power problem, but that means mainly that we should design our studies well enough--i.e. have an adequate enough n -- that we can deal with the "subset benefit" phenomenon and still find a difference. If some therapists in other contexts are terrible, then once again we must have an adequate design with enough subjects to find a difference

anyway. If so many are terrible that no difference is there (in the hypothetical population from which we sample), then the therapy doesn't work.

There is also the problem of assessing those who drop out, but a well-designed study is going to assure at least some contact with them to assess at least some important outcome measures anyway. A more poorly designed study may use covariance or imputation as a substitute, but there's a big problem with **all** such matching techniques--the more successfully you match on observed variables the greater the difference on unobserved ones when people behave differently. Well, you can match for similarity--i.e. those in the control group with subsets in the experimental group who behave differently there, but then when the hypothetical group assignment is not perfect, you have exactly the type of regression effect that will make a difference within the experimental group that has nothing to do with the treatment assignment (e.g. those who did or did not take clofibrate more than 80% of the time) appear to be a difference between experimental versus control groups. (Is everyone aware of the clobfibrate example? only 15% new heart attacks in the next 5 years for those who took it more than 80% of the time, 25% for the others-- **exactly** the same figures or those who did or did not take the placebo The worst thing to do, of course, is to assume that dropping out is a negative outcome. People drop out for a number of reasons, including that they weren't that sick to begin with and/or that they got better.

Robyn

Date: Fri, 29 Sep 1995 14:03:17 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman) Subject: Re: Jacobson's comments on Lee's Comments on..
--

Marty,

See the work of Bill Miller (New Mexico) and others of "matching" for alcohol treatment. Randomly half or clients are randomly assigned, the other half assigned to a treatment based on an interview indicating whether their own beliefs and attitudes are more compatible with behavioral or with AA treatment. (Earlier studies indicated a whopping **crossed** interaction or outcome for those randomly assigned to behavioral or AA as a function of these beliefs and attitudes)

Robyn.

Date: Fri, 29 Sep 1995 12:26 From DonaldK737@aol.com To: k-howard@nwu.edu Subject: efficacy research

ken howard says:

"Efficacy research shows that under some highly specified conditions, on average, an intervention CAN be shown to produce better results than some comparison condition."

The problem is that without proper design that allows the inference the mere fact that one group of people do better on treatment a than another group on treatment b may have nothing whatever to do with the treatments. That's the rub. Randomization is like democracy, its the worst possible procedure except for everything else. (It takes care of group differences on variables you don't even know about.)

However there have been some interesting developments. My colleague Don Ross, elaborating on a paper by Zhang and Robbins, has demonstrated that inferences are possible under conditions of treatment group assignment above and below a severity level (with quite loose assumptions). Paper in press J Psychiat Res. Does that un-muddy things?

Marty Seligman suggests

"even better, how about a design in which half of the whole group of subjects are placed in random assignment, fixed duration, manualized psychotherapy of some kind, and the other half choose therapy and therapist, the duration is tailored to their progress, and the therapist can correct the trajectory of techniques which fail--all for some well-defined disorder"

I take it Marty means the two halves are due to random assignment. Something similar was discussed at the scientific advisory committee (SAC) of the TDCRP. The idea was to have a "treatment as usual (TAU)" track that one randomized group could work with in the usual fashion of a clinic applicant. Most clinics don't allow you to choose your therapist or the mode of treatment nor does the usual practitioner. If you wanted to include such an unusual practice it would require a separate track. The TDCRP SAC scientific objection was that TAU was so ill defined that it would be hard to say anything concrete. For instance if the treatment was no better than placebo what have we learned except some clinic may not be too sharp?

I argued that if the TAU did as well as the fancy manualized therapies (or better as Marty thinks) that would be meaningful. However budgetary considerations were overriding as that would have been a fifth track. The NIMH staff were also unenthusiastic.

Methodologically if you do a manualized vs a TAU smorgasbord you couldn't contrast meaningfully within the TAU group because patients with different

prognoses might choose different treatments. However I don't see any reason not to average over the whole group.

It seems impossible to have "the duration ... tailored to their progress" since treatment effects would be irretrievably time confounded. With waxing and waning illnesses if you wait long enough you are likely to hit a good patch. That's the time to declare victory!

Another problem would be determining if the act of therapist selection (rather than being assigned) is in itself anti-demoralizing (empowering?) So that any advantage to the self selected therapies may have more to do with access rather than therapy. That might be detected in the first sessions or require yet another contrast group. Things are getting rococco. Would Marty settle for usual clinic practice ?

Another possibility is to tell all patients that they will be assigned to various treatment tracks but they will not be told the method of assignment other than it is not calculated to put anyone at a disadvantage. This might reduce a differential feeling of special capacity.

I still think that any psychotherapy evaluation requires a pill placebo and therefore an active medication arm also. This makes scientific and public health sense. To complicate the issues: shouldn't we have two TAU tracks, one a pure psychotherapy track with optional psychiatric consultation (as is usual in most treatment clinics) and another where psychiatric consultation and medication evaluation are routine (as is usual in departments of psychiatry)?

Cordially
Don Klein

Date: Sat, 30 Sep 1995 03:34:36 -0400 From: jmraraback@aol.com To: PsyNetUSA@aol.com, sscpnet@bailey.psych.nwu.edu Subject: Fwd: Outcome Studies. (fwd) - Priorities defined

Thought this might be of interest to the list members.

John

John M. Roraback, Ph.D. <jmroraback@aol.com>
Southpark Psychology 309-797-2900 voice
2100 Fifty-Second Avenue 309-797-2147 fax
Moline, IL 61265-6366 [GMT -6]

Forwarded message:

From: "TERENCE C. DAVIES, M.D." <TCD@WORF.EVMS.EDU>
Organization: Eastern Virginia Medical School

The NY Times noted yesterday that the Foundation for Accountability met last week and identified ten conditions for intensive outcome study. The ten were depression, coronary artery disease, asthma, breast cancer, diabetes, hypertension, low back pain, arthritis, pregnancy and maternity care (ranked in order of their incidence/health-economic significance). The Foundation is a coalition of major industry reps who are out to monitor HMO's in terms of cost and quality. Looks like something that we should be watching.

Terence C. Davies, M.D.

Professor and Chairperson, Family and Community Medicine
Eastern Virginia Medical School
(804)446-7461 <voice>; (804)446-5196 <facsimile>; tcd@worf.evms.edu <email>

Date: Sun, 1 Oct 1995 11:48:

From: Lee B Sechrest <sechrest@u.Arizona.EDU>

To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>

Subject: Re: Jacobson's comments on Lee's Comments on..

On Fri, 29 Sep 1995, Martin E. P. Seligman wrote:

< see above >

OK, here goes.

First, since I do not have a particularly optimistic view about the outcomes of therapy anyway, my bet would be that neither approach would have impressive outcomes, and there would not be a great deal of difference between them. Don Klein is exactly right, however, about the lack of any justified conclusions that would ensue from letting therapists who win it to decide when to declare victory and quit. (This morning I was in a hardware store and the clerk had trouble finding the price sticker on the item I bought. Finally she did and remarked, in some exasperation, "How come the sticker is always in the last place I look!"

I am glad to see, Marty, that you recognize the necessity--absolute, I would say--of a randomized design in order to warrant conclusions from even your thought experiment.

My bet in any actual experiment would depend on the specifics of the arrangements for the study. To begin with, are we talking about graduate student therapists, MA therapists, "expert" doctoral level therapists, or what? Ceteris paribus, if we take a randomly selected group of "therapists" out in the community, I believe I would bet on the manual.

Your proposed experiment actually has more than one dimension, subjects, along which questions of potential bias (external validity) of the experiment would have to be dealt with. If therapists are not to be assigned randomly to therapies, then the outcome may reflect differences between therapists who choose one or another mode of therapy rather than simply differences between therapies per se. For example, the therapy mode elected by therapists may reflect their personality styles, their training, their age, their intelligence, their theoretical persuasions, etc.

A second issue has to do with what kinds of patients with what kinds of problems are to be involved. If you decide to assign 100 phobic or panic patients to the two treatments, I would be inclined to bet on the manual. I might do the same for CBT, depending on the nature of the problems presented by patients. On the other hand, if patients are presenting with vague complaints of ennui, dissatisfaction with their lives, and so on, and the manual chosen were IPT, I might well decide to bet on the instincts of the therapists--if I really had to bet.

My bet would also depend on the outcomes to be measured. If you choose patient satisfaction, I would almost certainly bet on the flexible therapy. Placebo, after all, means "I please." If functional outcomes are to be assessed, I would not, as indicated, be overly optimistic, but, again, *ceteris paribus*, I think I would bet on the manual.

Finally, you would have to specify more completely how the therapy is to be carried out. Your description could be seen as pitting a rigid, mechanistic approach (manual) against a much more flexible, sensitive, and compassionate approach (winging it). That need not be the nature of the distinction between therapies, however. Therapy manuals, as I have noted before, do not require therapists to do stupid things. Therapy manuals can be used for guidance but used flexibly by sensitive, compassionate therapists. On the other hand, a study such as you propose would have to be carried out with concern for the integrity of the therapy, i.e., we would have to be sure that manualized therapy were done "correctly," and that the winging-it therapy were also done in a way that more or less objective judges would recognize as "good" therapeutic practice.

No, one more thing. Cost-effectiveness would also have to be considered quite seriously. If you are going to let winging-it therapists decide when therapy is completed, my bet is that therapy will be a very long enterprise. It will then have to be a lot better than manualized therapy in order to be judged cost-effective. Especially if, as I assert, there is not likely to be a lot of difference between the two therapies in outcomes.

My thoughts on your thought experiment. Which is not going to get done. And that may be a pity. We need experiments of the sort you proposed--if they are randomized so that they can teach us something.

Date: Mon, 2 Oct 1995 07:47:02 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: clark lee anna <laclark@blue.weeg.uiowa.edu>

Subject: Re: Jacobson's comments on Lee's Comments on..

I will not include the original posting by M. Seligman that led to my comments, but I am now glad to have Lee Anna Clark's response

On Sun, 1 Oct 1995, clark lee anna wrote:

>Isn't this what Marsha Linehan did in her study of Dialectical Behavioral Therapy for borderline personality disorder? The "control" group received "treatment as usual in the community" -- they were free to pick and choose, etc. DBT did better than-or the same as--treatment as usual, including at 1-year followup, on a variety of outcome measures.<

Thanks much for the example. As Samuel Johnson said to Bishop Berkeley, I say now to Marty, "Thus do I refute you!"

Lee Sechrest, Univ. of Arizona

Efficacy & Effectiveness

including
APA's MANUALS FOR EMPIRICALLY VALIDATED TREATMENTS

letters from sscpnet 5/95

Date: Tue, 3 Oct 1995 08:44:41 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: The Seligman Design

I have changed the subject title of this string in order to make clear what the focus now is

On Mon, 2 Oct 1995, Martin E. P. Seligman wrote:

--Previous postings deleted, but Marty was responding to L.A. Clark's reference to M. Linehan's study of treatment of borderline personality cases--

> Treatment as usual in the community does not wash for the design I am proposing. The kind of treatment, e.g., CBT for panic, and the disorder must be held constant, e.g. dsm4 panic. But then patients are entered randomly into either a fixed duration, manualized, random therapist set of cells, or the other half: choice of therapist, length of therapy adjusted to progress, order of techniques adjusted to progress of patient. Changing the treatment modality completely does not test what I am suggesting needs testing.<

Marty, your current version of your Design lacks verisimilitude as a study of psychotherapy effectiveness. A major problem as the study is now structured is that patients are *assigned* not only to the manualized therapy condition but are assigned, apparently, to therapists. Meanwhile, the alternative treatment condition, let us call it "idealized psychotherapy" permits patients to choose their own therapist.

From the standpoint of generalizability of the study, you have two choices:

- 1) To do a study that would compare a, let us say, Biodyne model with an idealized model. In the Biodyne model, people come into a clinic and are assigned to a therapist who will practice according to a fixed protocol, etc. In the idealized model, virtually nothing is fixed.
- 2) To do a study that compares fixed protocol therapy done under more or less free choice conditions with therapy done without a fixed protocol.

The first choice, the test of the Biodyne model confounds the fixed protocol with constraints on patient choice. That is not a necessary confounding. It is imaginable that in the real world patients would go to see therapists who might differ in whether they work by a fixed protocol or simply wing it. Patients might not even be aware of the difference.

A problem exists in defining what is meant by "choice of therapist." Most patients probably have relatively little "choice" in the sense that patients do not have a description of all the therapists available, they are constrained by geography, they get referred, and so on. If you have in mind that patients should have the opportunity to "shop around" for a therapist, you have to remember that such shopping around can add greatly to the eventual total cost of treatment. If patients do not find it necessary to shop around much, then the case for choice is to that extent vitiated.

I do not have a ready solution to the design problem that preserves randomization, but I am glad that we both continue to see that without randomization, the results of the study would be substantially uninterpretable. No, let us say that the results might well be such as to leave substantial residual uncertainty.

Date: Tue, 03 Oct 95 12:02 From: John Hunsley <HUNCH@acadvm1.uottawa.ca> Subject: Seligman's research proposal To: SSCPNET <sscpnet@bailey.psych.nwu.edu>
--

In the context of the promotion of empirically validated treatments, it seems to me that the type of research design suggested by Marty Seligman (random assignment to "by the book" manualized treatment vs. treatment from a therapist who practices in general accordance with the "book") is a definite necessity. Most clinicians who will claim to be practicing empirically validated treatments will be self-educated on the basis of limited workshop sessions and the use of some of the manuals listed as resources for those who want to practice empirically validated treatment (e.g., many of the chapters in David Barlow's 2nd edition of Clinical Handbook of Psychological Disorders). In the spirit of the enterprise of clinical research suggested over 15 years ago by Agras and Berkowitz (Clinical research in behavior therapy: Halfway there? 1980, Behavior Therapy, 11, 472-487), the evaluation of training methods in established treatments and the evaluation of field efficacy are the among the final steps in developing effective interventions.

John Hunsley, University of Ottawa

Date: Tue, 3 Oct 1995 09:41:02 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: John Hunsley <HUNCH@acadvm1.uottawa.ca> Subject: Re: Seligman's research proposal

On Tue, 3 Oct 1995, John Hunsley wrote:

< see above >

John is probably right about the training of clinicians "who will claim to be practicing empirically validated treatments."

But, to be rectilinear about all this, the training of clinicians who will claim to be practicing standard "book" therapy needs to be understood as well. Here is my guess: they will have had practicum plus an internship and will have

"seen" maybe 8 or 10 patients in what might be called psychotherapy training with anything from virtually zero to maybe pretty good supervision, probably averaging OK. From that point on they will have been self-educated by means of books written by someone regarding him or herself as affiliated with Milton Erickson or will have gone to a few workshops given by people regarding themselves as affiliated with Milton Erickson. Or maybe they will both have read books and gone to workshops.

I, myself, do not believe we are ready for tests of the field effectiveness of psychotherapies of any kind. We need to know whether when the therapies are done right by people with training we judge as having been excellent the therapies result in clinically (and socially) important changes beyond those attributable to well-conceived and implemented placebo treatments.

Lee Sechrest, Univ. of Arizona

Date: Tue, 3 Oct 1995 09:10:13 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: Lee B Sechrest <sechrest@U.ARIZONA.EDU> Subject: Re: The Seligman Design
--

Lee,

You have picked up on the one major design flaw in Marty's proposal, the "choosing a therapist". But if that part of the design is eliminated, and subjects are simply assigned randomly to "research structured" versus "clinically flexible" conditions, with therapists held constant, and choice of client held constant, this becomes an excellent design for testing the performance of psychotherapy under typically research versus naturalistic (or relatively naturalistic) conditions.

Best,
Neil

Date: Mon, 9 Oct 1995 13:42:17 -0500 From: k-howard@nwu.edu (Ken Howard) To: sscpnet@bailey.psych.nwu.edu Subject: Robyn Dawes' Comments

Robyn:

I am reluctant to start a debate with you in this medium because I am not sure that you can keep to a superficial level. But I don't want you to think I have been ignoring your insightful comments. So, here goes a little try on one of your points.

Premise: Efficacy research is intended to show that under some specified conditions, for carefully selected patients and well-defined treatments, intervention T can be shown to be superior (yield a statistically reliable higher mean outcome score) to some comparison condition, C. (These requirements of specificity, selection, and definition will tend to limit the external validity [generalizability] of any findings.)

One requirement for such a conclusion ($T > C$) is that the two groups were equivalent (comparable) to begin with. Depending on how strict the definition of "equivalence" is in any given experiment, it is very unlikely that random assignment will equate two relatively small groups with regard to causal influences (other, unspecified causally efficacious independent variables). (For example, as you well know, if one took 60 coins and randomly assigned them to two groups of 30 each, it would be very unlikely that the two groups would show the same number of heads and tails.) By the time a study is completed, moreover, whatever comparability that had been established through random assignment will have been compromised by attrition from the two groups.

One solution that has been proposed is to trust the comparability putatively established by the original randomization and to preserve the integrity of the groups by defining the treatment (e.g., CBT) as offering the treatment to the patient (i.e., randomly assigned to CBT, whether or not the patient actually attended any sessions). It is my understanding that this is what has been called "intention to treat." This raises some other problems.

1. It is unlikely that "termination" or follow-up outcome scores will be available for all of the dropouts. So, it will probably be the case that even an intention to treat analysis will be compromised by attrition.

2. People who drop out of the study will certainly try to do other things about their problems (e.g., seek other treatments) -- they do not go into hibernation. So the intention to treat analyses will compare groups that will have to be carefully redefined and will probably not be comparable (i.e., after dropping out, but before the outcome assessments, more of one group may choose medication, more of the other group may be severely disturbed).

3. If the secondary analysis of treatment completers (which will certainly be done regardless of the results of the intention to treat analyses) indicates that the treatment has a notable effect on those patient who are exposed to a sufficient dose, it would be hard to convince proponents that they should not put more effort into treatment compliance.

I agree with you completely that treatment compliance is an important element in evaluating a treatment. The problem is what to do with the non-compliers. One solution is to treat them all as failures (impute the lowest outcome score). It would be useful to know that a treatment found little patient acceptance. Clinicians would want to know how effective the treatment was for patients who were willing to comply with it. Because, if the treatment seemed good enough, they may want to try methods for obtaining higher compliance.

You indicate that you would prefer making sure that you get outcome data on everyone, and evaluate mean differences. If these exist (and $T > C$), then presumably the subset that actually received a therapeutic dose of T did enough better to carry the analysis -- the T-completers had higher average outcome scores, while the T-dropouts, the C-completers, and the C-dropouts had lower means. As you point out, this would require a large N and/or a large treatment effect (as well as actually getting all subjects to complete termination and followup assessments). More importantly, we would certainly have to examine (in secondary analyses) the T-completer results to ascertain that this was the source of the significant difference.

Perhaps one solution would be to use treatment compliance as an inclusion criterion. After all, the investigator is free to set these parameters and often chooses such hidden criteria as "agreed to sign an informed consent," "completed pretest forms," etc. Why not "completed treatment?" I guess one "why not" is that most exclusion criteria are applied pre-randomization, while treatment compliance is a post-randomization issue.

(With regard to compliance, we and others have found that the majority of patients who are prescribed Prozac never take a dose. Of course the medical literature is replete with this issue and no one seems to have solved it.)

Otherwise, the only question that RCT methodology can be expected to address is whether being offered Treatment T yields better average outcomes than being offered a comparison intervention, C. Given this scenario, in the service of full disclosure and scientific integrity, the investigator would have to state that the aim of his or her RCT is to determine if offering patients T results in better average outcome scores than does offering patients C. The investigator could not take into account whether any patient actually received any T or C or the amount of T or C any patient received. Further, the investigator could not take into account any other outcome-related (e.g., other

treatments) activities of the patients. (Of course, these could be assessed and used to describe the overall conduct of the study, they just could not be used to inform or modify the planned comparison.) I am not very confident that such a study would ever be funded. Personally, I do not find such a research question very relevant for clinical practice.

Ken Howard

Date: Tue, 10 Oct 1995 09:02:38 -0400 (EDT) From: Saul Shiffman <shiffman+@pitt.edu> Subject: Re: Robyn Dawes' Comments To: Ken Howard <k-howard@nwu.edu>
--

Just some thoughts about intent-to-treat analyses and drop-outs (telling you things you already know):

- 1) the picture is very much complicated by the fact that drop-out is likely related to outcome, not just an independent nuisance variable. the patients who are getting no benefit presumably are the most likely to drop out. So, participation in treatment can't simply be treated as a 'dose' variable.
- 2) There are statistical methods for dealing with incomplete data: i.e., using the last available measure from drop-outs or using all their measures. All the solutions are imperfect, but surely better than considering all drop-outs as failures or not analyzing their data.

Saul

Date: Tue, 10 Oct 1995 10:17:26 +0000 From: "Paul D. Rokke" <rokke@plains.nodak.edu> To: sscpnet@bailey.psych.nwu.edu Subject: Tangential thought to Shiffman: Dropouts
--

Saul Shiffman wrote:

>Just some thoughts about intent-to-treat analyses and drop-outs (telling you things you already know):

1) the picture is very much complicated by the fact that drop-out is likely related to outcome, not just an independent nuisance variable.

The patients who are getting no benefit presumably are the most likely to drop out. <

I realize that this is a little tangential to the issues being discussed, but it raised a question that I've been curious about.

The key word above is 'presumably.' Although it is often assumed that it is the treatment failures that drop out, I wonder how much we really know about dropouts.

My situation may be a little unique, but we've actually had the opposite problem with dropouts. After having difficulty recruiting sufficient numbers of older adults to participate in a treatment outcome study we changed our inclusion criteria. It seemed that we had several older adults seeking treatment for depression, who were in fact experiencing significant levels of distress, but who were not qualifying for a diagnosis of major depression. I argued successfully to the granting agency that it might be prudent to loosen our criteria to include participants who were mildly depressed. Thus, I hoped, we would increase our N and obtain a sample that was more representative of those who sought help.

Unfortunately, loosening the criteria didn't help our N much. Many of the less severely depressed participants have come in, enrolled in the program, and after 2 or 3 sessions (10 were scheduled) drop out. Not because they were not benefitting, but because by that time they were feeling very good and had difficulty justifying the additional time and effort for what was perceived as little additional benefit.

Has anyone done a review, aggregating across studies offering similar treatments and that have described dropouts sufficiently, to describe the modal characteristics of dropouts and their reasons for discontinuing? And then, by extension, further describe and distinguish the effects of treatment? (In my experience, more data from drop outs would be extremely useful. But of course that is what makes them dropouts - they don't provide more data.)

Paul

Paul D. Rokke
Department of Psychology

North Dakota State University
Fargo, ND 58105-5075
Voice: 701/231-8626
Fax: 701/231-8426
Internet: rokke@plains.NoDak.edu

Date: Tue, 10 Oct 1995 09:33:20 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: "Paul D. Rokke" <rokke@plains.nodak.edu>
--

On Tue, 10 Oct 1995, Paul D. Rokke wrote:

--much deleted--

>The key word above is 'presumably.' Although it is often assumed that it is the treatment failures that drop out, I wonder how much we really know about dropouts. Has anyone done a review, aggregating across studies offering similar treatments and that have described dropouts sufficiently, to describe the modal characteristics of dropouts and their reasons for discontinuing? And then, by extens distinguish the effects of treatment? (In my experience, more data from drop outs would be extremely useful. But of course that is what makes them dropouts - they don't provide more data.)<

Dropouts are scarcely likely to be a homogeneous set. Dropouts occur for different reasons. 1. Paul is right in supposing that some people drop out of experiments because the calculus of costs and benefits shows a negative value for the latter. We ought to be more explicit in paying attention to the costs of participating in experimental studies, e.g., time costs, travel costs, inconvenience costs, etc.

2. Some people drop out of experiments because they find the intervention toxic. I am involved in a dietary fiber study in which dropping out is substantially correlated with side effects of the high fiber diet. Some psychological interventions unquestionably produce side effects that are unpleasant.

3. Still other people drop out of studies because their lives are, for whatever reasons, chaotic, and they simply cannot get themselves organized to keep appointments and meet all the other demands put on them by researchers.

4. And still other people drop out for other reasons that some of you may be able to imagine or for reasons that none of us can quite imagine. I am reminded of one large focus group study of high school dropouts that resulted in a conclusion that a great many students just sort of drift away from school. They do not make any particular decision. Things happen, and one day they find that they simply are not going to school anymore.

All of the above may occur in any given study. If we are going to study and learn how to analyze data properly, we will need to have a much more differentiated view of the dropout problem than we now do.

Date: Tue, 10 Oct 1995 15:27:59 -0500 To: sscpnet@bailey.psych.nwu.edu From: k-howard@nwu.edu (Ken Howard) Subject: Invalidated treatments

The major goal of efficacy research is to show that a treatment can work (at least under the specific conditions of the experiment). There has been considerable discussion about determining which treatment should be considered validated (have sufficient empirical evidence to support the determination of efficacy).

It is my understanding that current lists of validated treatments only consider treatments that have a reasonable manual and have been subjected to clinical trials (usually randomized). Treatments that do not make this list have either been shown to be invalid or have not been manualized and evaluated via randomized clinical trial.

At my staff meeting this morning a graduate student asked: "Various and diverse forms of psychotherapy have presumably met the standard of efficacy. A list of scientifically validated treatments has now been disseminated. To lend support to the concept of "scientific validation", it would be helpful to define scientifically invalid treatments. Have any scientifically invalidated treatments been determined? Or, are the rules of science applied in such a way that any treatment can be shown to be efficacious if the procedures, subjects and dependent variables are designed to highlight the strengths of the particular intervention?"

Can someone please provide me with a list of invalidated treatments (preferably using the same criteria as the Division 12 report)?

Ken Howard

Date: Wed, 11 Oct 1995 09:06:46 -0400 (EDT)
From: Saul Shiffman <shiffman+@pitt.edu>
Subject: Re: Robyn Dawes' Comments
To: Ken Howard <k-howard@nwu.edu>

I grant both issues, but note that they are somewhat contradictory. If dropouts don't differ in relevant respects, then the assumption of random attrition may be reasonable.

Saul

Date: Thu, 12 Oct 1995 06:49:46 -0700 (PDT)
From: Jake Jacobs <jakej@alnitak.usc.edu>
To: DR ARNOLD A LAZARUS <RBDL01A@prodigy.com>
Subject: Re: Valid Treatments

On Wed, 11 Oct 1995, DR ARNOLD A LAZARUS wrote:

>I had an interesting experience at a talk I gave in New York this past Saturday. After my talk, over iced tea and carrot cake, there were several clinicians present who strongly argued that there are no valid treatments, only good therapists using effective methods (in a contextual and reciprocal interaction). This is not new to anyone, but I was struck by the vehemence that was displayed and the strong sentiments that were expressed against treatment manuals. If these were peripheral folks or some touchy-feely or psychodynamic personnel, I would walk away unmoved by their prejudices. I am referring to people who are highly respected in several different quarters and who are so adamant about this position that they are ready to take legal action. They feel that experimenters are dictating to clinicians how therapy out to be done and that managed health care concerns will adopt arbitrary standards of validity to the detriment of all. When I tried to discuss this with an eminent authority, he simply laughed them off as misguided fools. I do not think that this is the way to deal with their objections. I think they need to be taken seriously and disabused of their views via logic, reason, facts and ongoing debates.<
Arnold

For the long haul, a good place to start might be in abnormal classes at the undergraduate level. Continuing the lesson through graduate and post-doctoral training might not hurt. Most of these people are, after all, a product of University training.

W. Jake Jacobs

Date: Thu, 12 Oct 1995 11:08:26 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: sscpnet@bailey.psych.nwu.edu Subject: efficacy versus effectiveness

Arnold's exasperation is poignant. While I am a fan of the power of efficacy studies, it must be said there is some, considerable justification--epistemologically--for the discontent, although not for the indifference, of many practioners to efficacy studies. Here's why: The following six properties characterize psychotherapy as it is done in the field. Each of these properties are absent from an efficacy study done under controlled conditions. If these properties are important to patients' getting better, efficacy studies will underestimate or even miss altogether the value of psychotherapy done in the field:

- 1) Psychotherapy (like other health treatments) in the field is not of fixed duration. It usually keeps going until the pateint is markedly improved or he quits. In contrast, the intervention in efficacy studies stops after a limited number of sessions--usually about twelve--regardless of how well or poorly the patient is doing.
- 2) Psychotherapy (again, like other health treatments) in the field is self-correcting. If one technique is not working, another technique or even another modality is usually tried. In contrast, the intervention in efficacy studies is confined to a small number of techniques, all within one modality and manualized to be delivered in a fixed order.
- 3) Patients in psychotherapy in the field often get there by active shopping, entering a brand of treatment they actively sought with a therapist they screened and decided to work with. In contrast, patients enter efficacy studies by the passive process of random-assignment to treatment and acquiescence with whom and what happens to be offered in the study (Howard, Orlinsky & Lueger, 1994.)
- 4) Patients in psychotherapy in the field usually have multiple problems, and psychotherapy is geared to relieving parallel and interacting difficulties. Patients in efficacy studies are selected to have but one diagnosis by a long set of exclusion and inclusion criteria.
- 5) Psychotherapy in the field is almost always concerned with improvement in the quality of life, as well as amelioration of a disorder and relief of specific, presenting symptoms. Efficacy studies usually focus only on specific symptom reduction and whether the disorder ends.
- 6) In the field people tend to choose the psychotherapist they want to work with and the modality they most believe in. Random assignment may turn out to be a major flaw of the efficacy method. Many (but assuredly not all) of the problems that bring consumers into therapy have elements of what was called wanhope in the middle ages and is now called demoralization. Choice and control by a patient, in and of itself, counteracts wanhope (Seligman, 1990). Random assignment of patient to a modality or to a particular therapist not only undercuts the remoralizing effects of treatment, but it also undercuts the nonrandom decisions of therapists in choice of modality for a particular patient. Assume, for the moment, that therapists are canny about who need

drugs plus psychotherapy and who can do well with psychotherapy alone. The therapists assign those patients accordingly and so appropriate patients get appropriate treatment. This is just the same logic as a self-correcting trajectory of treatment in which techniques and modalities are modified with the patient's progress.

It is hard to imagine how one could ever do a scientifically compelling efficacy study of a treatment which had variable duration, self- correcting improvisations, and was aimed at improved quality of life as well as symptom relief, with patients who were not randomly assigned and had multiple problems. But this does not mean that the effectiveness of treatment so delivered cannot be empirically validated. Indeed it can, but it requires a different method, and the effectiveness methods may serve us better here.

Martin E. P. Seligman

Date: Thu, 12 Oct 1995 08:14:59 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: DR ARNOLD A LAZARUS <RBDL01A@prodigy.com> Subject: Re: Valid Treatments
--

On Wed, 11 Oct 1995, DR ARNOLD A LAZARUS wrote:

> see above >

Or else they need to face up squarely to the implications of their position, which are that clinicians will need to be *individually* validated and that reimbursement should be allowed *only* for the services of clinicians demonstrated to be effective.

Lee Sechrest, Univ. of Arizona

Date: Thu, 12 Oct 1995 08:45:14 -0700 (PDT) From: Gerald Davison <gdaviso@alnitak.usc.edu> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Validated Treatments
--

The two recent postings by my friends and colleagues Arnold Lazarus and Jake Jacobs move me to make a downbeat response.

Arnold is correct that many people "in the field" are suspicious of efforts to stipulate empirically validated treatments, and many are unimpressed with controlled research that downplays or overlooks the complex realities of clinical interactions and, dare I say, clinical intuition.

And Jake is correct when he says that we should attend carefully to how we educate our undergraduates and graduate students.

The problem is, I think, that even graduates of our doctoral programs often drift away from the lessons that I suspect we SSCP members try to teach them as well as the undergraduates. In a way, then, the "experiment" that Jake suggests has already been conducted: we have pursued the s-p model of education and training for many years, and yet, if there is an enemy, the enemy may be us (well, "we," actually).

Like most on this network, I have wrestled with this problem for many years, and student members of the network are no doubt experiencing it in their practica and internship placements. Do the research-oriented clinical internships (the ones Section III tries to steer our students to) do any better in helping academic programs produce people who, five years later, are hewing to the s-p model whether they are in academic or applied settings? I agree with Arnie that we need to resist the temptation to write "them" off, but it is less clear to me what kinds of dialogues will improve a divisive and unproductive situation.

-- Jerry Davison

P. S. Just read Lee Sechrest's posting, to which I offer the following response. I am becoming less confident that validated treatments or validated therapists are going to win the day. They should, I agree. And most of the time I think they surely have to. But I would not underestimate the political clout of those who hold other views. The recent APA Council action on prescription privileges provides yet another lesson. I respect and admire Robyn Dawes, Neil Jacobson, and no doubt others for resigning from APA, but I don't think that's going to do it. AAAPP's efforts on the drug business make a lot of sense to me, and I am lending my hand to those initiatives. Too bad that, as far as I can tell, APS has not really entered the fray. I think that most non-clinical and non-counseling psychologists continue to be blind to the serious implications that our disputes with the less data-oriented applied people have for the integrity and vitality of academic departments and programs.

Date: Thu, 12 Oct 1995 11:00:37 -0500
From: k-howard@nwu.edu (Ken Howard)
To: Jonathan Huppert <huppert@acs.bu.edu>
Subject: Re: Invalidated treatments

>Wouldn't treatment studies like those showing that "psychodynamic and client-centered" treatment groups did worse than no-treatment groups which did worse than functional family therapy groups for conduct disordered children, on the rating of future court contact (63%, 59%, 40%, and 20% respectively) be a start for a list of ineffective (and harmful) treatments?

Jonathan Huppert<

If the two treatments were manualized and these were the results of at least two RCTs, then we should conclude that client-centered therapy and psychodynamic therapy are invalid interventions for the treatment of conduct disordered children. I do not know the details of the studies in your posting, but is it fair to say that these treatments should be blacklisted (in the same spirit that we are approving validated treatments)? Any objection?

Ken Howard

Date: Thu, 12 Oct 1995 11:09:07 -0500
From: k-howard@nwu.edu (Ken Howard)
To: Lee B Sechrest <sechrest@U.ARIZONA.EDU>
Subject: Re: Invalidated treatments
Cc: sscpnet@bailey.psych.nwu.edu

>Lee Sechrest, Univ. of Arizona wrote:
my scientific superego prompts me to point out that what your student and you are asking is akin to affirming the null hypothesis. It is quite possible to show that a treatment is *better* than no treatment or than a placebo and, hence, to "validate" the treatment in some degree. It is not so easy to establish that a treatment is *not* at all better than no treatment or a placebo.<

You are becoming a rabble rouser. You know as well as I that the null hypothesis is the least plausible alternative hypothesis known to man and is certainly almost always false. However, we have ways of determining if a given (a priori clinically meaningful -- say 10% difference in efficacy) difference has been established (e.g., confidence intervals). We can test to see if a given difference is statistically significantly small enough to ignore -- a variant of equivalency testing.

The main issue is to find treatments that meet the methodological criteria used by the Division 12 panel and where the decision is "no established efficacy despite trying the best we could."

Ken Howard

Date: Thu, 12 Oct 1995 13:46:46 -0500 To: RBDL01A@prodigy.com From: k-howard@nwu.edu (Ken Howard) Subject: Re: Valid Treatments
--

I have held similar discussions with people who have conducted studies using manualized treatments with session limits. Not one of these clinical researchers follows such procedures with their own patients.

Ken Howard

Date: Thu, 12 Oct 1995 13:51:15 -0500 From: k-howard@nwu.edu (Ken Howard) To: sscpnet@bailey.psych.nwu.edu Subject: Re: Validated Treatments

One way to engage in the validated treatment debate would be to have clinicians have some input on outcome criteria. It is not just design and methodological restrictions that make research alien from practice, it how the treatment and outcome are defined.

Ken Howard

Date: Thu, 12 Oct 1995 10:52:58 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Re: Valid Treatments (fwd)

Forwarded at the request of M. Strauss

----- Forwarded message -----

Date: Thu, 12 Oct 1995 12:56:27 -0400
From: Milton E. Strauss <mes3@po.cwru.edu>
To: Lee B Sechrest <sechrest@u.Arizona.EDU>
Subject: Re: Valid Treatments

>Lee Sechrest, Univ. of Arizona

>
>On Wed, 11 Oct 1995, DR ARNOLD A LAZARUS wrote:
>
>> see above>>

Perhaps I am missing something, but what is the issue about "only good therapists using effective methods." I haven't seen anything in the exchanges that denies the importance of nonspecific factor in treatment (there is still a lot to be learned from J. Frank), nor has anyone said that a manual for an effective treatment is all that is necessary for clinical effectiveness. Isn't a core question whether or not people who might be good therapists (I don't want to open the can of worms about what that means, btw) might be doing their patients better to use treatments of demonstrated efficacy?

Milton E. Strauss, Ph.D.
Department of Psychology
Case Western Reserve University
Cleveland, OH 44106-7123

Date: Thu, 12 Oct 1995 19:17:37 -0700 (PDT) From: Jason Satterfield <satter@itsa.ucsf.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: efficacy versus effectiveness

Marty,

I'm not quite ready to throw in the towel on efficacy research. The "six properties [that] characterize psychotherapy" can be addressed through RCTs or similar methodologies.

On Thu, 12 Oct 1995, Martin E. P. Seligman wrote:

>Arnold's exasperation is poignant. While I am a fan of the power of efficacy studies, it must be said there is some, considerable justification-- epistemologically--for the discontent, although not for the indifference, of many practitioners to efficacy studies. Here's why:

The following six properties characterize psychotherapy as it is done in the field. Each of these properties are absent from an efficacy study done under controlled conditions. If these properties are important to patients' getting better, efficacy studies will underestimate or even miss altogether the value of psychotherapy done in the field:

1) Psychotherapy (like other health treatments) in the field is not of fixed duration. It usually keeps going until the patient is markedly improved or he quits. In contrast, the intervention in efficacy studies stops after a limited number of sessions--usually about twelve--regardless of how well or poorly the patient is doing.<

Perhaps this was true before the days of managed care. The typical clinical scenario now is to receive a case with clear time-limitations and little

flexibility (usually around 6 sessions). Clinicians are now forced to focus on specific, concrete goals just as researchers have done (for the most part) in efficacy trials. "Psychotherapy" and efficacy trials are becoming more and more similar given the health care environment.

>2) Psychotherapy (again, like other health treatments) in the field is self-correcting. If one technique is not working, another technique or even another modality is usually tried. In contrast, the intervention in efficacy studies is confined to a small number of techniques, all within one modality and manualized to be delivered in a fixed order.<

Several comments: 1) I think you're being too generous. Lots of clinicians stick with treatments/methods that don't work. 2) Manualized treatments are not inflexible or mindless. Aaron Beck and David Burns repeatedly drummed into my head that I needed to have 30 different ways of cognitively intervening with a particular patient problem. If one doesn't work, you try another but all of them are in the manual. 3) This sounds like the old eclecticism argument. Any evidence that eclecticism is advantageous? I'd guess it would diminish the confidence/remoralization effect (your number 6 below) if your therapist is always changing hats.

>3) Patients in psychotherapy in the field often get there by active shopping, entering a brand of treatment they actively sought with a therapist they screened and decided to work with. In contrast, patients enter efficacy studies by the passive process of random-assignment to treatment and acquiescence with whom and what happens to be offered in the study (Howard, Orlinsky & Lueger, 1994.)<

The randomization is a "passive" process but the study selection/recruitment allows the patient to make some choice. Patients also choose whether or not to continue treatment. Perhaps the sample which completes the study (or some cutoff point) meets with your criteria of psychotherapy.

>4) Patients in psychotherapy in the field usually have multiple problems, and psychotherapy is geared to relieving parallel and interacting difficulties. Patients in efficacy studies are selected to have but one diagnosis by a long set of exclusion and inclusion criteria.<

Efficacy studies do seek out a "pure" sample but it's really a matter of degree. Most clinicians will pick target symptoms or prioritize goals and very rarely are all "problems" treated. I also doubt if a research treatment targeted at say depression only affects depressive symptoms. If the right outcome measures are used we might see an affect on marital discord, parent effectiveness, anxiety, job success, etc.

>5) Psychotherapy in the field is almost always concerned with improvement in the quality of life, as well as amelioration of a disorder and relief of specific, presenting symptoms. Efficacy studies usually focus only on specific symptom reduction and whether the disorder ends.<

I'd hope efficacy studies not only look for specific symptom reductions but also ask the quality of life questions - what I think of as true clinical significance.

>6) In the field people tend to choose the psychotherapist they want to work with and the modality they most believe in. Random assignment may turn out to be a major flaw of the efficacy method. Many (but assuredly not all) of the problems that bring consumers into therapy have elements of what was called wanhope in the middle ages and is now called demoralization. Choice and control by a patient, in and of itself, counteracts wanhope (Seligman, 1990). Random assignment of patient to a modality or to a particular therapist not only undercuts the remoralizing effects of treatment, but it also undercuts the nonrandom decisions of therapists in choice of modality for a particular patient. Assume, for the moment, that therapists are canny about who need drugs plus psychotherapy and who can do well with psychotherapy alone. The therapists assign those patients accordingly and so appropriate patients get appropriate treatment. This is just the same logic as a self-correcting trajectory of treatment in which techniques and modalities are modified with the patient's progress. It is hard to imagine how one could ever do a scientifically compelling efficacy study of a treatment which had variable duration, self-correcting improvisations, and was aimed at improved quality of life as well as symptom relief, with patients who were not randomly assigned and had multiple problems. But this does not mean that the effectiveness of treatment so delivered cannot be empirically validated. Indeed it can, but it requires a different method, and the effectiveness methods may serve us better here.<

I agree effectiveness or clinical services research are important. But designing creative and effective interventions which we understand seems to first require some basic bench work - efficacy studies.

Just my 2 cents.

Jason

Jason M. Satterfield, Ph.D.
Department of Psychiatry
University of California
401 Parnassus Avenue, Box CPT
San Francisco, CA 94143-0984
(415) 476-7569

Date: Thu, 12 Oct 1995 19:52:3
From: DonaldK737@aol.com
To: rd1b+@andrew.cmu.edu
Subject: Re: Jacobson's comments on Lee's Comments on..

Robyn defends intent to treat designs with much logic. however it should be made plain that such a design switches the question from determining if a treatment agent has an active ingredient over and above placebo, expectancy, natural course etc to whether a treatment program -that has various degrees of success in delivering the agent-is better than placebo, expectancy, natural course etc. logically the failure of an itt design to show specific activity may be as much programmatic as substantive. if, however a specific benefit is found it is likely to be a lower bound. an upper bound is usually yielded by a last observation carried forward analytic approach, so thats a useful bracket. much interest in work of rubin/lavori in attempts to model dropout process (propensity scores) and to include this in analysis which might get us back to original question.

cordially
don klein

Date: Thu, 12 Oct 1995 19:53:01 -0400
From Donald Klein
To: sechrest@U.ARIZONA.EDU
Subject: Re: CBT and nonspecific factors (fwd)

dear all
sechrest says " Actually, I suspect that psychotherapy *is* a placebo, that is it consists of nonspecific effects that result in sensed empathy, caring, support, and induced optimism, etc." i think the term nonspecific is problematic. if you believe that the major target of nonspecific psychotherapy is the quite specific state of demoralization then you might have a handle on how to improve the therapeutic alliance placebo eg guided mastery experiences. i think this particularly useful when avoidant behavior is maintained by anticipatory anxiety and corrective experience sabotaged by demoralization.
don klein

Date: Fri, 13 Oct 1995 06:24:03 -0500 (CDT) From: "Joseph J. Plaud" <plaud@badlands.NoDak.edu> To: DonaldK737@aol.com Subject: Re: CBT and nonspecific factors (fwd)

On Thu, 12 Oct 1995 DonaldK737@aol.com wrote:

< see above >

It is always nice to see converts to behavioral psychology. SSCPnet can bring out the best in people. To wit, if, as Don Klein (or is that don klein?) suggests, the "target" of nonspecific therapy is a behaviorally-specific treatment regimen (such as "mastery experience" training), then it can be concluded that when we can define behavior specifically, as well as its intervention, nonspecific effects can become quite specific. This is a strength of behavior therapy: we can understand and manipulate the socially-mediated reinforcers that are responsible for the so-called nonspecific effects, and in so doing transform them into specific effects. Therefore, while the issue of nonspecific effects pertains to bt as well as to other therapies (including the so-called hot-tub variety of eclectic approaches), the difference (which turns out to be most important) is that in bt we can utilize directly the social reinforcement that produces nonspecific effects in the first place. Good going, Don. And Lee, right on as usual.

Joe Plaud

Date: Fri, 13 Oct 1995 08:25:47 -0400 From: DonaldK737@aol.com To: plaud@badlands.nodak.edu Subject: Re: CBT and nonspecific factors (fwd)

dear all,
joe plaud welcomes my conversion

< see above >

but i hope he agrees with me that there is still only slight evidence that directed social reinforcements have been shown to produce more than anti-demoralization/therapeutic alliance placebo effects. thats the nub of the problem. sounds good in principle but we need the demonstration. the liebowitz/heimberg social phobia study is the only really convincing one i know. in my phobia study systematic hierarchical desensitization prover no better than simple nondirective support.

see Klein, DF, Zitrin, CM, Woerner, MG & Ross, DC: Treatment of Phobias: II. Behavior therapy and supportive psychotherapy: Are there any specific ingredients? Arch Gen Psych 40(2):139-145, 1983.

don klein

Date: Fri, 13 Oct 1995 09:09: From: k-howard@nwu.edu (Ken Howard) To: DonaldK737@aol.com Subject: Re: Jacobson's comments on Lee's Comments on..

Don Klein wrote:

>if ,however a specific benefit is found it is likely to be a lower bound. an upper bound is usually yielded by a last observation carried forward analytic approach, so thats a useful bracket.

>cordially don klein <

I know of no evidence to support this contention. Any inference about specific effects based on an intention to treat design implies an examination of self-selected treatment compliers and defeats the point of the itt.

Ken Howard

Date: Fri, 13 Oct 1995 08:41: From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Unmentionable variables
--

There he goes again! Ken Howard will stop at almost nothing in his vain (in two senses of the word) efforts to shore up the case for psychotherapy.

The idea that in relatively small experiments random assignment does not protect against biases in the form of unmentionable (well concealed) variables associated with therapy outcomes just does not wash. Here is why.

First, there probably aren't any such variables! After decades of futility, psychotherapy mavens still have no predictors of psychotherapy outcomes, except maybe for the YAVIS syndrome. How likely is it that in even a reasonably well conducted and documented therapy study there would be important but unnoticed YAVIS differences between groups? Or differences in initial severity? Everyone does use pretests, you know. But just for the hell of it, let's hear about some examples of randomized therapy studies with as

many as 10 cases per treatment group in which it has ever been discovered that, by golly, the groups were different in some important way that no one ever noticed until after the study was published.

The second reason is that differences of any consequence are improbable. And, in any case, the same considerations that work against statistical power for detecting differences when samples are small operate against detecting effects of biases. "Chance" differences are *no more likely* in small than in large studies, that is true by definition and the immutable laws of probability. "Chance" differences may be larger in smaller studies, but then their effects on outcomes must also be larger in order to be statistically significant. We need some simplifying assumptions here. Let's assume a total of 40 patients are to be randomly assigned to two equal size ($n = 20$) groups. There are smaller therapy studies, but they are underpowered anyway. Assume also that some unmentionable variable with a favorable impact on response to therapy has a frequency in the population of .5. It will then be the case that a split of 13- 7 favoring good outcome in one group and a symmetrical split of 7- 13 against good outcome in the other group will occur less than 5% of the time. That is, splits of that degree of extremity will be quite infrequent. We would, of course, find equally frequent splits in the opposite direction, i.e., against therapy. Out of every 20 such therapy studies we might expect two with that degree of bias. On average cancelling each other out. So Ken's handwringing over deeply buried biases in therapy studies is overdone on the basis of frequency alone. (Note that splits on the unmentionable variable(s) that depart from .5 will result in fewer "biased" cases in either group and, therefore, an attenuated effect. Note also that if there are multiple unmentionable variables that are uncorrelated, their effects will tend to cancel out, just as randomization intends.)

And, since any biases of the sort imagined by Ken would be random, the effect on estimates of therapy outcome would be *unbiased*, i.e., the bias would as often operate in favor of the experimental as the control group. Across a series of studies, e.g., as in a meta-analysis, one would not expect any bias in estimates of effect size unless it were true that unmentionable variables tended to underestimate the effects of treatment and, therefore, produce a file-drawer problem. (Does anyone know of an otherwise well-conducted therapy study that was never published because of failure to find an effect?)

But we are not finished yet. Even if the groups differ in terms of some factor *favorable* to a good therapy outcome, that factor will be a fallible indicator of therapy outcome. Let us say that our indicator, for which selection bias has occurred, correlates .50 with therapy outcome (a larger relationship than we are likely to find). Using Rosenthal's BESD (binomial effect size display), we would expect .75 of persons having the factor to have a favorable therapy outcome, etc. So the actual therapy outcome results, assuming no real therapy effect, would be about 12 positive cases in the experimental group and only 8

in the control group. But with 40 cases, that difference would not be statistically significant. Even the fairly extreme bias would not have resulted in an erroneous conclusion.

Lee Sechrest, Univ. of Arizona

Date: Fri, 13 Oct 1995 09:28:30 -0500 From: k-howard@nwu.edu (Ken Howard) To: "Richard Suinn" <richard_suinn@cnsmail.natsci.colostate.edu> Subject: Re: Invalidated treatments

To Richard Suinn:

Thanks for your thoughtful reply to the issue of invalidated (or failed to validate) treatments. The student was responding to the attempt to determine a list of validated treatments. We had been discussing problems with establishing the efficacy of a treatment. It seemed that outcome criteria for the RCTs used to validate the treatments could be idiosyncratic, integrity checks for manualized therapies seemed to never result in throwing out cases (i.e., all cases in a clinical trial seem to have adequate integrity), attrition was seldom addressed, but, more importantly, all treatments subjected to an RCT seemed to be efficacious.

Having sat on review committees, I know that we tend to require that the research proposal provides pilot data that supports the efficacy of a treatment. Given promising pilot results a RCT is mounted that seems to always result in support for the efficacy of the treatment. If this is the case, why bother to do an RCT? Are RCTs just scientific window dressing to support what we already know? If not, then there must be some examples of interventions that (in replicated RCTs) have been shown to not be statistically superior to a therapeutically inert comparison condition.

I think it would be interesting to compile a list of these "failed to be validated treatments." Someone has suggested light therapy for SADs, but did not provide the necessary replicated RCT references.

Ken Howard

Date: Fri, 13 Oct 1995 08:06:01 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: DonaldK737@aol.com
Subject: Re: Jacobson's comments on Lee's Comments on..

On Thu, 12 Oct 1995 DonaldK737@aol.com wrote:

--some deletion--

> if ,however a specific benefit is found it is likely to be a lower bound. an upper bound is usually yielded by a last observation carried forward analytic approach, so thats a useful bracket. <

We ought to think more in terms of putting bounds on estimates anyway.
There is no one right way to analyze any set of data worth analyzing.

Lee Sechrest, Univ. of Arizona

Date: Fri, 13 Oct 1995 11:18:35 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Invalidated Treatments
To: sscpnet@bailey.psych.nwu.edu

This has developed into an interesting discussion, but to me it seemed a bit narrow. When I started trying to think of some treatments that had been invalidated, my list was based on a wider casting of the net.

For example, the "Doman-Delacato" treatment for cerebral palsy has been cast into outer darkness by the research community for a long time. Ed Zigler has written at length about the lack of evidence for its validity.

A special issue of the Journal of Clinical Child Psychology on child psychopharmacology a few years ago cast considerable doubt on the efficacy of so-called antidepressant medications for treating depression in pre-pubertal children, although they often were effective in treating ADHD.

In the field of applied behavioral analysis, reinforcing children for paying attention in class turned out not to generalize to academic productivity. However, reinforcing academic performance did generalize to measures of attention.

Is this kind of wider discussion helpful to you regulars on the SSCP Net? I am a new participant in this discussion.
Don Routh

Date: Fri, 13 Oct 1995 10:45:06 -0500 From: k-howard@nwu.edu (Ken Howard) To: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Invalidated Treatments

>Is this kind of wider discussion helpful to you regulars on the SSCP Net? I am a new participant in this discussion.<
>Don Routh<

A wider discussion would be interesting, but then we would get back to having lots of validated treatments (almost all brands of psychotherapy). The committee that came up with the list of validated treatments used specific criteria -- a treatment manual, replicated Randomized Clinical Trials, etc. I wanted to use the same criteria to find "failed to be validated treatments" to ascertain whether the determination "validated" just meant that someone went to the trouble to do two RCTs (and the results were a foregone conclusion).

Ken Howard

Date: Fri, 13 Oct 1995 10:58:43 -0500 From: k-howard@nwu.edu (Ken Howard) To: Lee B Sechrest <sechrest@U.ARIZONA.EDU> Subject: Re: Unmentionable variables

Lee Sechrest, Univ. of Arizona, wrote:

< see above >

Be still, my heart.

Lee Sechrest agrees with me that we should allow patients to choose the therapy they want in a study. Randomization doesn't accomplish anything important (as Lee points out) and attrition reduces us to self-selected groups in any case. Finally, some sense is emerging.

Ken Howard

Date: Fri, 13 Oct 1995 09:05:43 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: Ken Howard <k-howard@nwu.edu> Subject: Re: Invalidated treatments
--

On Thu, 12 Oct 1995, Ken Howard wrote:

> You are becoming a rabble rouser. You know as well as I that the null hypothesis is the least plausible alternative hypothesis known to man and is certainly almost always false. However, we have ways of determining if a given (a priori clinically meaningful -- say 10% difference in efficacy) difference has been established (e.g., confidence intervals). We can test to see if a given difference is statistically significantly small enough to ignore -- a variant of equivalency testing. The main issue is to find treatments that meet the methodological criteria used by the Division 12 panel and where the decision is "no established efficacy despite trying the best we could."<

Ken, I *am* rabble! God loved us; he made so many of us.

I agree with you about the issues here and that we need to learn how better to argue for the "no difference despite our best efforts" conclusion. In fact, I have written about that problem several times, and were your scholarship up to the level of your psychotherapy-partisan activities, you would surely be aware of that fact.

A highly desirable strategy would be to move in the direction of specification of exact quantitative hypotheses (even if it would be painful to try to be so exact) and then do likelihood ratio tests for the tenability of the two hypotheses, e.g., is it more likely that the effect of an intervention is 40% (placebo value) or 60%, a clinically useful increment over placebo? Or more likely that the change on a well-being scale is 5 points (trivial) or 20 points (important)?

On reflection, Ken, I will have to admit that what I have written on accepting the null (actually, uninteresting difference) hypothesis is published in places that would have become evident only to an assiduous scholar.

your humble rabble representative.

Lee Sechrest, Univ. of Arizona

Date: Fri, 13 Oct 1995 11:03:48 -0500
From: k-howard@nwu.edu (Ken Howard)
To: Lee B Sechrest <sechrest@U.ARIZONA.EDU>
Subject: Re: Unmentionable variables

>Lee Sechrest, Univ. of Arizona
But just for the hell of it, let's hear about some examples of randomized therapy studies with as many as 10 cases per treatment group in which it has ever been discovered that, by golly, the groups were different in some important way that no one ever noticed until after the study was published.<

The crucial issue is not only group differences but "let's hear about some examples of randomized therapy studies with as many as 10 cases per treatment group in which" secondary analyses were not conducted on subgroups to tease out differential treatment effects (e.g., post hoc analysis of severity in the Collaborative).

Ken Howard

Date: Fri, 13 Oct 1995 15:35
From: Donald Klein@
To: k-howard@nwu.edu
Subject: Re: Jacobson's comments on Lee's Comments on..

dear all,
there seems to be some confusion re intent to treat designs. I can't follow Ken Howards concerns. It is moderately complicated so please read the the best reference i know: Clinical Evaluation of Psychotropic Drugs. ed by Prien and Robinson, Raven Press. NY, 1994
see

Laska, EM, Klein, DF, Lavori, PW, Levine, J & Robinson, DS: Design issues for the clinical evaluation of psychotropic drugs.

and particularly. chapter 7 Statistical issues for the clinical evaluation of psychotropic drugs. by lavori laska and uhlenhuth.

cordially,
don klein

Date: Fri, 13 Oct 1995 15:09:44 -0500
From: k-howard@nwu.edu (Ken Howard)
To: sscpnet@bailey.psych.nwu.edu
Subject: Whoops

>Could you please send me clearer copies of Jacobson and Sechrest in the men's room at APA? Thanks.<

Please excuse this message; it was meant for the sexual dysfunction network. Sorry.

Ken Howard

Date: Fri, 13 Oct 1995 15:30:25 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: Ken Howard <k-howard@nwu.edu>
Subject: Re: your mail

On Fri, 13 Oct 1995, Ken Howard wrote:

> Could you please send me clearer copies of Jacobson and Sechrest in the men's room at APA? Thanks.<

I am glad to know that Ken remembers at least that some years back I did work on graffiti (Homosexuality in the Philippines and the US: the handwriting on the wall; Graffiti in four institutions of higher education). That work was *about* writing on men's room walls. Actually, however, my last publication *in* a men's room was when I was about 15. Perhaps Neal has been more recently involved in that sort of thing. Still, Ken appears to have a strange approach to scholarship in the field. Lots of people carry things into men's rooms to read. Newsweek, GQ, JCCP, Rolling Stone, Psych Bull, other journals.

Lee

Date: Fri, 13 Oct 1995 18:55:48 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: Lee B Sechrest <sechrest@U.ARIZONA.EDU>, k-howard@nwu.edu (Ken Howard) Subject: Re: Unmentionable variables

Ken,

I'm preparing a longer response. But I must disagree with your assertion that "attrition reduces us to self-selected groups in any case." We should, in so far as possible, be looking at **all** those assigned to the treatment--not just completers. The problem is that a comparison within the treatment group of completers with noncompleters is an invalid (ambiguous) one, because these two groups may differ on all sorts of characteristics other than those resulting from completing or not. The appropriate comparison is of completers in the experimental group with those in the control group who would have completed had they been assigned to the experimental group but weren't assigned to it. The problem is that we don't know who those are, and **THUS FAR** anyway, statistical means for constructing "hypothetical membership" in them have--unless someone is very silent about success--not been good enough to overcome the regression problem. Again, consider the colfibrate example, only assume that (as in most psychotherapy studies) there is no placebo (so that we would have no way of knowing that those who took SNY pill conscientiously had fewer subsequent heart attacks).

The point is that the **whole** control group in an randomized trials experiment does indeed have counterfactual "hypothetical" membership in the experimental group because it was randomly constructed. That logic certainly doesn't preclude peoples' choosing treatments--just that they won't necessarily get the ones they choose, at least not right at the time of the experiment. Moreover, people are always trying to do things on their own, but that's not a unique feature of psychotherapy as opposed to medical trials. We don't tell people in control groups in medical trials experiments to lie over and die if that is the outcome of passivity. The problem might be worse in psychotherapy studies, but that is the nature of the phenomenon we're studying--people. They (we) are active, seeking of explanations and better states, etc., etc., etc. Isn't such seeking an intrinsic part of what's going on, rather than a hindrance to discovering some sort of "pure" measure of efficacy, or whatever you want to call it?

Robyn

Date: Mon, 16 Oct 1995 08:10:18 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Howard's End

I am distressed to learn that my old friend Ken Howard is symptomatic again. You see, he, along with many others known to all of you readers, suffers from an apparently interminable disease called Psychotherapy Obsession Disorder (POD). This disorder is characterized by an obstinate attraction to and absorption in psychotherapy. Whereas to most of us, psychotherapy is a possible (but not likely) means by which to help some people better their lives, for the POD victim, and one can only feel some degree of sorrow for them, psychotherapy is an *end*, an activity to be practiced for its own sake, with scarcely any regard for what might be accomplished by it.

At least without regard to what might be accomplished beyond making a living. Once when I was (Neil Jacobson, shut your eyes) on APA Council, a prominent practitioner then (actually perpetually) a member, said during a debate something to the effect of "I like to do psychotherapy. I don't care whether it is effective or not. As long as people are willing to pay me to do it, I am going to continue." (I am pretty sure that was the practitioner someone once described as "the one who looks like a high school shop teacher.") I can well imagine psychotherapy researchers saying something quite similar, or thinking it anyway. (There is some similarity of POD to hysteria, I think, and the concept of secondary gain is clearly applicable to this disorder.)

POD also has an element of delusion associated with it, for part of the syndrome is a conviction, not well substantiated by any actual evidence, of the ultimate effectiveness of psychotherapy. It has often been said that behind every paranoid delusion is an element of truth (hasn't that often been said?), and the conviction of the fundamental effectiveness of psychotherapy appears to lie in a smattering of small signs, subtle things you see, lying here and there in the literature. PODs are like to say that "it has now been shown that psychotherapy is effective," and things of that sort, even though, as D. Klein has repeatedly insisted, not a single well designed and executed study shows psychotherapy to be better than a placebo treatment. In any case, it is also true that not a single study shows psychotherapy to be *sufficiently* effective that one ought to care much about it, let alone invest all one's professional capital in it. Yet that is just what PODs do.

The response of PODs to challenges to the effectiveness of psychotherapy is to redouble their commitment to it. They insist that: a. since psychotherapy is effective, even a weensy bit, b. there must be something there, and c. *it*, the essence, can be discovered if enough effort is put into it. Therefore, they commit themselves, and wish to commit considerable public resources, to a program of research aimed at discovering just what it is about psychotherapy "that works," with the almost equally futile assumption that if they could just discover it, they could improve it a great deal and make psychotherapy "really work." Poor PODs! I am writing this and the parts that follow in an effort to

provide some cognitive therapy for PODs and maybe some preventive intervention for younger colleagues and students who may have some of the risk factors for POD. I imagine that many of you who are, seemingly, immune to this strange disease will be able to imagine what those risk factors are.

One final note: POD is not yet in the DSM, but I have no doubt that with modest effort we can get it included in DSM-IV-R or at least in DSM-V since virtually any aberration known to man counts as a mental disturbance. The advantage will be that POD can then be treated. By psychotherapy, no doubt, which will result in ironic justice--interminable POD.

Lee Sechrest, Univ. of Arizona

Date: Mon, 16 Oct 1995 09:12:43 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: Lee B Sechrest <sechrest@U.ARIZONA.EDU> Subject: Re: Howard's End
--

Dear Lee and Ken,

Two words: group therapy. I know someone who would treat you conjointly via conference call, with an empirically validated treatment-which, by the way, doesn't work. Good luck: I don't think these relationship problems are insurmountable.

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

Date: Mon, 16 Oct 1995 14:27:29 -0400 (EDT)
From: "Robert C. Carson" <rcarson@acpub.duke.edu>
To: Lee B Sechrest <sechrest@U.ARIZONA.EDU>
Subject: Re: Howard's End

Lee,

Well, there you've gone and blown my good opinion of your advanced-age wisdom almost as soon as I expressed it. If POD makes it into DSM-IV-R (or V), the recommended treatment will certainly not be that of any form of psychosocial intervention. More likely, it will be that of pharmacological obliteration of the offending idea with a compound that will, by then, be POD-specific--much as the SSRIs now counter the several dozen diseases known to be caused by serotonin deficit, *and that alone.* Indeed, with scientific advance in this decade of the brain we may even be able to locate precisely the anatomical locus of POD, eradicating it forever by heating, freezing, lasering, cytotoxic lysis, or surgical extirpation by very fine cannula. In an unique twist of fate, these latter procedures might even fall within the purview of the mass-produced graduates of doctoral factories, our own professional brethren (or sistren, in p.c.). The possibilities are breathtaking, and both of us may yet survive to see it!

Bob

Date: 16 Oct 1995 14:52:16 -0700
From: "Richard Suinn" <richard_suinn@cnsmail.natsci.colostate.edu>
To: "Ken Howard" <k-howard@nwu.edu>,
"Donald K. Routh" <drouth@umiami.ir.miami.edu>,
Subject: Re: Invalidated Treatments

Reply to: RE>>Invalidated Treatments

Reply to: RE>Invalidated treatments

What an interesting question from Howard and from Routh! The questions might stir the perspective of philosophy of science-type orientation: if we set out to invalidate the belief that the world is flat, we do this by testing the hypothesis that contradicts this belief (and prove instead that the world is round). So we have a definition and procedure. Applying this approach, if we set out to identify the "invalid" treatments, what is the obverse: is it showing that the treatment is no better than no treatment or placebo treatment...or is it showing that the treatment is worse than, ie harmful. It is intriguing that researchers tend not to set out to determine (prove) if a psychological therapy to be harmful as a main goal; on the other hand, regulations over new drug therapies must ultimately pass the test of not doing harm, otherwise FDA will not permit release of the drug. Another issue is the complicated one of research design. We generally assume that an effective therapy is reasonably tested provided that the research says that the therapists are 'trained therapists', and that there is a reasonable matching regarding the

characteristics of the patient sample (e.g., not on medication, of same age group, meeting certain criteria of severity, etc.) However, an intervention might not show up in a study as leading to significant improvements, but the issue is the design. I am referring to the possibility that the effectiveness of a therapy is influenced by matching, ie matching the therapy to the type of disorder, but also matching the therapy to the stage or severity of the disorder, the existence of additional disorders, the experience/expertise of the therapist, etc. Put it another way, a therapy might not help clients with a disorder of high severity, but might prove useful during the early stages - if the study does not offer this type of analysis, it can be misleading. So, perhaps the simple finding that one treatment is not as good as another treatment for a specific disorder only establishes the power of the one but does not invalidate the other until a more sophisticated series of studies is conducted. (As an aside, it is interesting to note that licensing tends to determine that the individual 'will do no harm' rather than that the individual will be effective. The diplomate is oriented on the other hand to determining competency at a high level).

It would seem to me that defining the criterion to conclude a treatment is "invalid" is the first step, and a step we have seldom if ever taken for studies of psychological therapies. Showing that a treatment is no better than wait list or placebo does not necessarily invalidate the treatment if we use the definition of doing harm. If we do adopt "doing harm" as the criterion to answer your student's question, then I suspect we have very few, if any, studies leading to such a conclusion. First, because of a lack of definition of harm (long term follow-up is not routine in studies), and next, because of a reluctance to make this drastic statement - I think I can only recall one such conclusion made in an article about a therapy...perhaps others have seen more). I am not proposing "doing harm" as the definition of choice for "invalid". I am suggesting that we consider that we have several categories: an effective treatment for certain disorders (perhaps regardless as to severity level), an effective treatment for certain disorders (perhaps if matching criteria is identified, e.g., valid for mild severity but not for high severity), non-effective for certain disorders (in the sense of not being better than wait list or placebo), and contra-indicated (where data suggest possible harm).

-Dick Suinn

Date: Mon, 16 Oct 1995 22:10:51 -0400
From: Donald Klein@
To: richard_suinn@cnsmail.natsci.colostate.edu
Subject: Re: Invalidated Treatments

dear all

i have difficulties following suinn's discussion about the definition of harm. psychotherapies are necessarily(if implicitly) held out as more effective than placebo or noone would buy them. if ,in fact, they are no better than placebo ,then the patient has wasted time and money on a complicated professional endeavor when they could have received a cheaper/more easily administered nontoxic pill that could be truly described as equivalently effective. i view being induced to waste time and money as harmful. does anyone disagree?

that psychotherapies would have to do worse than placebo to be considered harmful equates "harm" with "specifically toxic" which ignores the other real harms defined above. if toxicity was shown it would be overkill but perhaps necessary to deal with that therapy's aficianadoes who will surely find reason to fault a mere lack of null hypothesis invalidation. they might not be put off by toxicity either.

cordially
don klein

Date: Tue, 17 Oct 95 06:50:55 PDT
From: beutler@edstar.gse.ucsb.EDU (Larry Beutler)
To: DonaldK737@aol.com
Subject: Re: Invalidated Treatments

The concept of "placebo" as applied to psychotherapy is particularly difficult. Most believe that a credible placebo is unavailable and that the medical analogy in which 'PLACEBO' is applied is awkward at best and useless at worst. Don Klein's comments suggest that all placebos are equivalent, another fallacy. If one could substitute an inert pill for psychotherapy and get the same effect, fine, but the consensus of evidence when pill placebos are compared to psychotherapy is that there is a difference. Even different types of placebos have different effects. The concern derives in part from the fact that the consequences of placebo, in the form of hope, expectation, etc., are similar to the mediators of much of psycho- therapy. That is not to say that they are equivalent, however, either in effect or mechanism. If the so-called placebo effects are true mediators and moderators of psychotherapy, they then are active ingredients, not inert. The question as to whether there are interpersonal processes that can be called and identified as controllable facets of psychotherapy can activate these qualities better, faster, or more predictably than other interventions, becomes very viable. The task, I believe, is to be able to identify these qualities, and to establish some control over

them so that their effects are both maximized and predictable. At present we suffer from the fact that their effect is inconsistent (within cell variance) and not precisely controllable. But, it remains, that if these are the active ingredients of a treatment, they are by definition, not a placebo.

LEBeutler

Date: Tue, 17 Oct 1995 11:47:51 -0500 (CDT) From: HOUTSAC@MSUVX1.MEMPHIS.EDU (ART HOUTS University of Memphis) Subject: Re: Invalidated Treatments To: sscpnet@bailey.psych.nwu.edu
--

Let's suppose that pill placebo as implemented in randomized trials is an effective psychosocial "treatment" compared to seeking treatment but having to wait. Further, suppose that pill placebo of some variations is just as effective as psychotherapy of some variations. I have a practical question. Is it feasible to offer pill placebo interventions outside the context of a study? Is it ethical to transport the pill placebo manipulation from the laboratory to practice? If pill placebo cannot be implemented, what is the status of arguments about psychotherapy versus placebo?

Date: Tue, 17 Oct 1995 17:56:52 -0500 (CDT) From: David Tolin <dtolin@comp.uark.edu> To: ART HOUTS University of Memphis
--

On Tue, 17 Oct 1995, ART HOUTS University of Memphis wrote:

> see above >

I have wondered about the ethical permissibility of placebo treatments as well, but primarily from the standpoint of non-validated psychological treatments. If we cannot demonstrate that a treatment has any effect over and above that of a placebo, and yet no validated interventions have been found for the presenting problem, is it ethical to administer the treatment? I would especially like to know if the AMA (having wrestled with the placebo issue longer than we have) has taken a position on the use of pill placebos, and if so, whether we might take some guidance from them?

Regards,
David

David F. Tolin

email: dtolin@comp.uark.edu

Department of Psychology
University of Arkansas
216 Memorial Hall
Fayetteville, AR 72701

Phone: (501) 575-4256
Fax: (501) 575-3219

Date: Tue, 17 Oct 1995 12:55:11 -0600 (MDT) From: miklowitz@clipr.Colorado.EDU Subject: RE: Invalidated Treatments To: HOUTSAC@msuvx1.memphis.edu
--

On Tue, 17 Oct 1995 HOUTSAC@MSUVX1.MEMPHIS.EDU wrote:

> see above >

Do you mean transporting pill placebo, or do you mean transporting the "active clinical management" or "warm pill" approach that, in randomized trials, usually accompanies a pill placebo condition? It seems to me there are separate therapeutic benefits attributable to the placebo pill versus the therapeutic context in which it is administered. I can't imagine transporting the one without the other.

David Miklowitz PhD
U Colorado, Boulder

Date: Tue, 17 Oct 1995 13:57:04 -0500 (CDT) From: HOUTSAC@MSUVX1.MEMPHIS.EDU (ART HOUTS University of Memphis) To: miklowitz@clipr.colorado.edu Subject: RE: Invalidated Treatments
--

David Miklowitz PhD wrote:
< see above >

I mean transporting the whole thing. Is this even possible from practical and ethical standpoints?

Hout

Date: Tue, 17 Oct 1995 19:17:43 -0400 (EDT)
From: "David H. Barlow" <MB399@cnsvox.albany.edu>
To: beutler@edstar.gse.ucsb.edu
Subject: Re: Invalidated Treatments

In addition to Larry's interesting points, it seems to me that patients, on the face of it, are most likely to respond differentially to pill placebo as a function of their model of disorder. If they believe in a chemical imbalance (based perhaps on something they recently read) they might respond more favorably than if they believed in a psychological cause. We're testing this in our collaborative panic study. But this, if it's true, would make pill placebo little more than a crude benchmark for comparing psychosocial treatment. Psychosocial placebos are needed. Returning to an earlier issue on the net, it seems to me that George Stricker is correct that the development of most psychosocial treatments begins with studies of Effectiveness, or what the APA Task Force calls Clinical Utility. Most often these are in the form of clinical replication series that we have described at some length (Barlow Hayes & Nelson 1984) Witness the Wolpe series on fear reduction that sparked so much efficacy research, or Masters and Johnson's series. The RCTs evaluating clinical utility proposed on the net are elegant, but they are unlikely to be carried out or funded as anyone knows who's been involved with the services IRG lately. I think that the overwhelming preponderance of data on clinical utility will come from well conducted clinical replication series, which are about to get an enormous boost from the creation of practice research networks "on line" to a central data base in professional associations or large MCOs At least 3 are in the works but A Psychiatric Assoc. deserves credit for being the first, lining up over 1000 clinicians.

Dave Barlow

Date: Tue, 17 Oct 1995 19:40:44 -0700
From: maxm2@ix.netcom.com (Max Molinaro)
Subject: Re: Invalidated Treatments and active placebos
To: beutler@edstar.gse.ucsb.edu, sscpnet@Bailey.psych.nwu.edu

Larry Beutler wrote:

> see above >.<

In fact, the 1992 article "A Meta-Analysis of Antidepressant Outcome under 'Blinder' Conditions" (Greenberg, Bornstein, Greenberg, & Fisher, JCCP, 60, 664-669) points out that the use of an "active" medication placebo (one that induces some kind of bodily sensation such as dry mouth) makes a significant difference in results of studies of medication, in this case, antidepressants. To wit: "Reviews of antidepressant outcome studies using active-placebo controls

suggest that the supposed superiority of antidepressants to placebos virtually disappear under these conditions." (p.667)

Even in the more medical side of treatment, it seems wise to choose the placebo carefully because of the difficulty of finding something which is only inactive in the crucial variable.

--Max Molinaro
maxm2@ix.netcom.com

Date: Wed, 18 Oct 1995 09:03:23 -0700 (PDT) From: Jake Jacobs <jakej@alnitak.usc.edu> To: sscp <sscpnet@bailey.psych.nwu.edu> Subject: Nonspecific factors and valid treatment

I have extracted three paragraphs from an article In today's New York Times:

"Just as talk psychotherapists vary widely in their techniques, dance/movement therapy is a diverse field, with each practitioner providing an individualized approach. Often the techniques of dance/movement therapy are combined with other psychotherapeutic methods, including art therapy and music therapy.

<snip>

Dance/movement therapy has helped those afflicted with a wide range of problems: physical and mental limitations, speech and learning disabilities, a history of sexual abuse, autism, family discord and personal psychological problems associated with feelings or acts the may be difficult to express in words.

<snip>

Dance/movement therapy has proved especially helpful for the elderly, whose outlets for relieving tension and loneliness are often limited. It is usually practiced in a group setting, where participants can benefit from touching and being touched. The physical contact alone can help to relieve feelings of loneliness and isolation and can provide much needed sensory stimulation."

(Thanks to Michael Gann for pointed this article out to me)

The article makes reference to two books (\$18.95 and \$29.95) and gives a way to find a trained dance/movement therapist in "your area." It occurred to

me that more detailed information may not be useful to members of this newsgroup, so I did not include it. But, if someone is interested, you can find the article on page B9 of the 18 October 1995 New York Times.

As our friend and colleague, Jerry Davison, suggested in an earlier post, we have already tried to train our students in the Boulder model (the experiment has been done.). He implied the experiment has failed. Given that, the question arises, "What real world contingencies (environmental, professional pressures) move our students from scientifically based psychotherapy to those that make such claims?" Given an answer to that question, the next one becomes, "What can a university program do to inoculate graduating students from such movement?" A third question arises from the last one, "If we discovered them, would it be appropriate to use effective inoculation procedures?"

W. Jake Jacobs

Date: Wed, 18 Oct 1995 14:07:03 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: Jake Jacobs <jakej@alnitak.usc.edu> Subject: Re: Nonspecific factors and valid treatment

Lee Sechrest, Univ. of Arizona

On Wed, 18 Oct 1995, Jake Jacobs wrote:
> see above >

REPLY: This posting makes two absolutely crucial points that all of us must understand.

First, the same sort of evidence of effectiveness that people want to adduce as probative for psychotherapy will support the effectiveness of dance therapy.

Second, you see what dancing leads to.

> As our friend and colleague, Jerry Davison, suggested in an earlier post, etc
> see above>

REPLY: I would suggest for starters that we spend more time with students explaining the difference between evidence and wish fulfilling fantasy. I would suggest, further, that we make clear to students our respect for practitioners who make an honest effort to make a scientifically honest living and our scorn for those who are mush-brained, muddle-headed, hypocritical, and/or dishonest. (I believe in the power of cognitions, and I believe cognitions can be implanted.) Finally, I suggest that we do our best to

diagnose the foregoing undesirable states in our otherwise splendid trainees and then set about trying to set them right or invite them to find other things to do if their states prove, instead, to be deep-seated traits. Psychology is pretty good at changing states but ineffective at changing traits.

Date: Wed, 18 Oct 95 12:43:46 PDT From: beutler@edstar.gse.ucsb.EDU (Larry Beutler) To: jakej@alnitak.usc.edu Subject: Re: Nonspecific factors and valid treatment

A comment on Jake Jacobs concerns with dance and related therapies.

If we think we have problems supporting the value of psychotherapy, you ought to review the literature on dance therapy and art therapy. It is devoid of scientific standards at the meagerist levels. I'll beg Jake's first question and move to the second one--WHAT CAN WE DO?

Frankly, we can do something. Most of the certificates to these programs are offered through extension programs, continuing education, and some university (medical school) related outreach programs. The first and last of these require some agency approval within the purviews of the university. I, for example, have to approve all extension courses in the mental health area in order for it to be offered as a University of California Santa Barbara program. I simply require that the program instructor, upon submitting the course plan/proposal, include references to at least two empirical studies showing that the procedures are effective (the Div 12 criteria). You'd be surprised what I get--the Bender-Gestalt validity studies are submitted in support of Art Therapy, etc. I have identified my criteria and have approval from the University and can stop any instruction in procedures that do not possess this minimal evidence, in my judgment. I think that our system is similar to most and it is simply a matter of reaching the people who must sign off on these proposals.

LEBeutler

Date: Wed, 18 Oct 1995 13:28:34 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Invalidated Treatments
To: SSCPNet@Bailey.Psych.nwu.edu

In response to David Barlow's query, I recently attended a lecture by Leon Eisenberg in which he talked about the early days of controlled clinical psychopharmacology research--in this case, the use of stimulant medications with hyperactive children. He mentioned that the FDA had a general category for those drug treatments that were well accepted in the professional community but had never been subjected to controlled trials. These compounds were collectively referred to as "GRAS" pronounced "grass" of course, meaning "generally regarded as safe." This had the effect of a sort of grandparent clause in the approval of these agents for routine clinical use.

Best wishes,
Don Routh

Date: Fri, 20 Oct 1995 15:00:00 -0500
From: k-howard@nwu.edu (Ken Howard)
To: sscpnet@bailey.psych.nwu.edu
Subject: POD

Anyone seeking reinforcement for their POD should read the November Consumer's Report. Psychotherapy is alive and well.

Ken Howard

Date: Fri, 20 Oct 1995 19:28:30 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: sscpnet@bailey.psych.nwu.edu, k-howard@nwu.edu (Ken Howard)
Subject: Re: POD

Ken,

Yes, some forms are alive and well--other forms are just alive, yet other forms are alive and should be offed (see Frontline Oct. 24). (It is one thing to talk to someone who is very distressed, another to talk to someone whose distress can be directly traced to our "hands-off" professional attitudes, as I did recently to the patient of Dr. B. who will most likely be on that program. It's the same feeling you get talking to "refusnicks" in 1979 in the USSR whose

children have been taken from them--only this time it's "us," not the USSR.)
But a survey.

1. How many therapists self-consciously use a technique discussed approvingly by CU/Marty S?
2. How many patients believe they are receiving one of these types of therapies?
3. And now here's the interesting part, but one that could involve some deception/violation of privacy. If some randomly selected therapist and patient pairs are asked independently what type of therapy is being given/received, how often would they agree?

Robyn

PS Here's a "subjective probability" question. How many people on this net would take an equal odds bet that each of these percentages is greater than 50%? less than 50%?

Date: Sat, 21 Oct 1995 10:23:31 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: Max Molinaro <maxm2@ix.netcom.com> Cc: sscpnet@Bailey.psych.nwu.edu, beutler@edstar.gse.ucsb.edu

On Tue, 17 Oct 1995, Max Molinaro wrote:

--stuff deleted--

> Even in the more medical side of treatment, it seems wise to choose the placebo carefully because of the difficulty of finding something which is only inactive in the crucial variable.<

Your point may be well taken insofar as the aim of research is to provide enlightenment about mechanisms by which a treatment works. But at a more pragmatic level, it seems clear that if it is so difficult to find a presumably cheap placebo that will allow the treatment to show some marginal effect, then we ought simply to use the placebo, including its "active" and "crucial" element.

Lee Sechrest, Univ. of Arizona

Date: Fri, 20 Oct 95 14:28 EDT
From: SPRINKS@CLEMSON.EDU
To: practice@LISTS.APA.ORG, div12@VM1.NODAK.EDU,
clinpsy@VM1.NODAK.EDU
Subject: practice: Div 12's "Manuals for Empir Treatment"

What follows comes from APA's Div 12, which has established a "Task Force on Psychological Interventions." The most recent issue of the *APA Monitor* carried a small item that indicated anyone could e-mail William Sanderson and receive a "Manuals List for Empirically Validated Treatments." Sanderson e-mailed the list to me, and indicated that I could re-post it. The task force is apparently open to feedback--see below on how to reach them. Also, I'm curious about other list member's impressions. --Steve Sprinkle
sprinks@clemson.edu

----- Text of forwarded message -----

MANUALS FOR EMPIRICALLY VALIDATED TREATMENTS

A Project of the Task Force on Psychological Interventions
Division of Clinical Psychology, American Psychological Association

William C. Sanderson, PhD & Sheila Woody, PhD (Editors)
June 1995

Development of this resource

The Division 12 (Clinical Psychology) Task Force on Promotion and Dissemination of Psychological Interventions issued a report in October of 1993 establishing criteria for judging whether a treatment may be considered to be empirically validated as well as forming an initial list of well-established and probably efficacious treatments. Please see that report (Task Force on Promotion and Dissemination of Psychological Procedures, 1995) for complete details. In order to facilitate the dissemination of these treatments, the Task Force on Psychological Interventions has now collected a list of manuals detailing the well-established treatments. (One exception will be noted by careful readers: Cognitive Behavior Therapy for Bulimia was inadvertently omitted from the initial Task Force list but is included in this resource.) Of course, this list will be updated as new treatments are empirically evaluated. In addition, several treatments that are in fact empirically validated but were overlooked in the initial report will be added to the updated version. We want to emphasize that this list is inclusive, not exclusive.

In order to compile this resource listing of manuals, we wrote to leading investigators in the respective areas of treatment research, particularly those whose work formed the basis for judging a particular treatment to be efficacious. These investigators provided citations for those published manuals. Many of them offered to provide copies of unpublished manuals to

other clinicians, although requests must be accompanied by a check to cover the costs of photocopying and postage.

What qualifies as a manual? In building this resource, we attempted to locate materials that provide sufficient detail to allow a trained clinician to replicate the treatment. Of course, no treatment manual is adequate in the absence of solid theoretical grounding and supervised training in the particular approach. Recognizing this, we have also included, when available, information about training in these approaches. We specifically excluded conference workshops as a training resource, because these workshops typically do not offer the opportunity for supervised experience.

Comments and feedback This listing is intended as a growing resource for researchers and clinicians. We welcome your feedback on the materials listed here, as well as your suggestions for materials we have overlooked. Please contact either of us:

William C. Sanderson, Ph.D.
Department of Psychiatry
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, NY 10467-2490
e-mail: SANDERSO@aecom.yu.edu

Sheila Woody, Ph.D.
Department of Psychology
Yale University
P.O. Box 208205
New Haven, CT 06520-8205
e-mail: SWOODY@minerva.cis.yale.edu

Reference:

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *the Clinical Psychologist*, 48, 3-23.

Suggested Citation:

Sanderson WC, Woody S. (1995). *Manuals for Empirically Validated Treatments: A Project of the Task Force on Psychological Interventions*. Division of Clinical Psychology, American Psychological Association. A hardcopy is available from Division 12 Central Office. Send stamped, self-addressed envelope and \$1.50 for handling to: PO Box 22727, Oklahoma City, OK 73123

BULIMIA

Cognitive Behavioral Therapy

Treatment References/Manuals:

Fairburn, C.G. (1985). Cognitive-behavioral treatment for bulimia. In D.M. Garner and P.E. Garfinkel (Eds.) *Handbook of Psychotherapy for Anorexia Nervosa and Bulimia*. New York: Plenum Press.

Fairburn, C.G., Marcus, M.D., & Wilson, G.T. (1993). Cognitive-behavioral therapy for binge eating and bulimia nervosa. In C.G. Fairburn & G.T. Wilson (Eds.) *Binge Eating: Nature, Assessment, and Treatment*. New York: Guilford Press.

Interpersonal Therapy

Treatment References/Manuals:

Fairburn, C.G. (1993). Interpersonal psychotherapy for bulimia nervosa. In G.L. Klerman & M.M. Weissman (Eds.) *New Applications of Interpersonal Therapy*. Washington, DC: American Psychiatric Press.

CHRONIC HEADACHE

Behavioral Treatment

Treatment References/Manuals:

Blanchard, E.B., & Andrasik, F. (1985). *Management of Chronic Headache: A Psychological Approach*. Elmsford, NY: Pergamon Press.

CHRONIC PAIN

Cognitive Behavioral Treatment

Treatment References/Manuals:

Cognitive Behavioral Treatment for Arthritis Pain (Contact: Francis Keefe, PhD, Pain Management Program, Duke Medical Center, Box 3159, Durham, NC 27710. Cost=\$30.00).

Turk, D.C., Meichenbaum, D., & Genest, M. (1983). *Pain and behavioral medicine: A cognitive-behavioral perspective*. New York: Guilford Press.

Training Available:

Francis J. Keefe, PhD
Director, Pain Management Program
Duke Medical Center - Box 3159
Durham, NC 27710
Telephone: 919-684-6212

Pain Evaluation and Treatment Institute
University of Pittsburgh School of Medicine
4601 Baum Boulevard
Pittsburgh, PA 15213

CHRONICALLY MENTALLY ILL

Token Economy Programs

Treatment References/Manuals:

Ayllon, T., & Azrin, N. (1968). The token economy: A motivational system for therapy and rehabilitation. New York: Appleton-Century-Crofts.

DEPRESSION

Cognitive Therapy

Treatment References/Manuals:

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). Cognitive Therapy of Depression. New York: Guilford.

Training Available:

Cory Newman, PhD
Center for Cognitive Therapy
University of Pennsylvania Medical School
3600 Market Street, Room 754
Philadelphia, PA 19104
Telephone: 215-898-4100

Judy S. Beck, PhD
Beck Institute for Cognitive Therapy
GSB Building - Suite 700
City Line & Belmont Avenues
Bala Cynwyd, PA 19004
Telephone: 610-664-3020

Cognitive Therapy Training Program
Cognitive Therapy Center of New York
3 East 80th Street
New York, NY 10021
Telephone: 212-717-1052

Interpersonal Therapy

Treatment References/Manuals:

Klerman, G.L., Weissman, M.M., Rounsaville, B.J., & Chevron, E.S. (1984). *Interpersonal Psychotherapy of Depression*. New York: Basic Books.

Training Available:

Cleon Cornes, MD & Ellen Frank, PhD
Western Psychiatric Institute and Clinic
3811 O'Hara Street
Pittsburgh, PA 15213
Telephone: 412-624-2211

Myrna Weissman
Psychosocial Therapeutic Systems
Graywind Publications
Stuyvesant Plaza
Executive Park Drive
Albany, NY 12203
Telephone: 518-438-3231

Videotape: *Interpersonal Therapy of Depression*
IPT Educational Foundation
5307 Cherokee
Houston, TX 77005
Telephone: 800-782-0015

DISCORDANT COUPLES

Behavior Therapy

Treatment References/Manuals:

Jacobson, N.S., & Margolin, G. (1979). *Marital therapy: Strategies based on Social Learning and Behavior Exchange Principles*. New York: Brunner/Mazel.

Baucom, D.H., & Epstein, N. (1990). *Cognitive-Behavioral Marital Therapy*. New York: Brunner/Mazel.

ENURESIS

Behavioral Treatment

Treatment References/Manuals:

Azrin, N.H. & Besalel, V.B. (1979). A Parent's Guide To Bedwetting Control. New York: Pocket Books.

Full Spectrum Home Training for Nocturnal Enuresis. (Contact: Arthur C. Houts, Department of Psychology, University of Memphis, Memphis, TN 38152. Cost=\$5.00).

GENERALIZED ANXIETY DISORDER

Anxiety Management

Treatment References/Manuals:

Anxiety Managment for Generalized Anxiety. (Contact: Secretary, Department of Psychology, Warnerford Hospital, Headington, Oxford, OX3 7JX. Cost = 2 pounds, prepayment required).

Cognitive Behavior Therapy

Treatment References/Manuals:

Controlling Anxiety. (Contact: Secretary, Department of Psychology, Warnerford Hospital, Headington, Oxford, OX3 7JX. Cost = 2 pounds, prepayment required).

Brown, T., O'Leary, T., & Barlow, D.H. (1994). Generalized anxiety disorder. In D.H. Barlow (Ed.), Clinical Handbook of Psychological Disorders. New York: Guilford.

Training Available:

Cory Newman, PhD

Center for Cognitive Therapy
University of Pennsylvania Medical School
3600 Market Street, Room 754
Philadelphia, PA 19104
Telephone: 215-898-4100

Judy S. Beck, PhD

Beck Institute for Cognitive Therapy
GSB Building - Suite 700
City Line & Belmont Avenues
Bala Cynwyd, PA 19004
Telephone: 610-664-3020

OBSESSIVE COMPULSIVE DISORDER

Behavioral Treatment

Treatment References/Manuals:

Steketee, G. (1993). *Treatment of Obsessive Compulsive Disorder*. New York: Guilford Press.

Riggs, D.S. & Foa, E.B. Obsessive compulsive disorder. In D.H. Barlow (Ed.), *Clinical Handbook of Psychological Disorders*. New York: Guilford.

PANIC DISORDER

Cognitive Therapy

Treatment References/Manuals:

Barlow, D.H., & Cerny, J.A. (1988). *Psychological Treatment of Panic*. New York: Guilford Press.

Barlow, D., & Craske, M. (1994). *Mastery of Your Anxiety and Panic - II*. Albany, NY: Graywind Publications. (Both therapist and client versions are available - To order call 518-438-3231)

Clark, D.M. (1989). Anxiety states: Panic and generalized anxiety. In K. Hawton, P. Salkovskis, J. Kirk, & D.M. Clark (Eds.) *Cognitive Behavior Therapy for Psychiatric Problems*. Oxford: Oxford University Press.

Salkovskis, P.M., & Clark, D.M. (1991). Cognitive treatment of panic disorder. *Journal of Cognitive Psychotherapy*, 3, 215-226.

Training Available:

Graywind Publications
Executive Park Drive
Albany, NY 12203
Telephone: 518-438-3231

POST TRAUMATIC STRESS DISORDER

Treatment References/Manuals:

Clinical Handbook/Therapist Manual on PTSD. (Contact: Donald Meichenbaum, University of Waterloo, Department of Psychology, Waterloo, Ontario, Canada N2L 3G1, Phone: 519-885-1211, ext. 2551, cost: \$40 plus \$5 shipping).

SOCIAL PHOBIA

Cognitive Behavioral Group Therapy

Treatment References/Manuals:

Cognitive Behavioral Group Therapy for Social Phobia by R. Heimberg (Contact: Karen Law, Center for Stress and Anxiety Disorders, Pine West Plaza, Building 4, Washington Avenue Extension, Albany, NY 12205. Cost = \$20.00).

Social Effectiveness Therapy: A Program for Overcoming Social Anxiety and Phobia (Contact: Samuel M. Turner, Ph.D. or Deborah C. Beidel, Ph.D., Turndel Inc., Suite 200, 615 Wesley Drive Charleston, SC 29464. Cost = \$39.00).

Training Available:

Richard Heimberg, Ph.D. & Harlan Justen, Ph.D.
Center for Stress and Anxiety Disorders
Pine West Plaza, Building 4
Washington Avenue Extension
Albany, NY 12205
Telephone: (518) 464-0241 or 869-2033
E-mail: rh188@albnyvms.bitnet

SPECIFIC PHOBIA

Systematic Desensitization

Treatment References/Manuals:

Wolpe, J. (1990). Practice of Behavior Therapy (4th Edition). New York: Pergamon Press.

Exposure Therapy

Treatment References/Manuals:

Marks, I. (1978). *Living with Fear*. New York: McGraw Hill.

END-----END

The Consumer Report

letters from sscpnet 6/95

This is the sixth report extracted from the network communications of the Society for Scientific Clinical Psychology (SSCP). It deals with a truly controversial topic: The scientific merits of a marketing device few very of us scientists engaged in psychotherapy research would have the courage to sell to his peers.

Well, read the thoughtful discussions and enlightenments on the pro and cons will ensure.

Date: Sun, 22 Oct 1995 12:05:46 -0700 (PDT)
From: Bill Follette <follette@unr.edu>
To: Ken Howard <k-howard@nwu.edu>
Subject: Re: POD

Say, Ken-

Do you see any minor sampling flaws in the Consumer's Report article?
Bill

On Fri, 20 Oct 1995, Ken Howard wrote:

>Anyone seeking reinforcement for their POD should read the November
Consumer's Report. Psychotherapy is alive and well.<

>Ken Howard

From: Howard Berenbaum <hberenba@s.psych.uiuc.edu>
Subject: Consumer Reports
To: sscpnet@bailey.psych.nwu.edu
Date: Sun, 22 Oct 1995 17:02:34 -0500 (CDT)

Although the data reported by Consumer Reports (CR) are probably interesting and have some useful things to teach us, they seem to not be capable of answering some of the questions that I would think we as clinical scientists should be most interested in. First, the data do not directly address the issue of the degree to which the problems that led folks to seek help were successfully addressed. It is my anecdotal impression that lots of folks report being very satisfied with their therapy experience despite little evidence of their problems having been overcome (would Woody Allen report that therapy has been helpful and he has been satisfied with it?). Second, the data seem to address the question 'does therapy work' -- haven't most clinical scientists agreed for a relatively long time that this is not the question we should be asking?

Howard Berenbaum
Department of Psychology
University of Illinois at Urbana-Champaign

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
Subject: Re: POD
To: follette@unr.edu (Bill Follette)
Date: Sun, 22 Oct 1995 17:03:37 -0400 (EDT)

I have looked at the sampling in the CR article rather carefully-- in fact I've written a 30 page article about the methodology.

What sampling flaws do you find, Bill?

Marty

Date: Sun, 22 Oct 1995 17:11:14 -0400
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: div12@vm1.nodak.edu, sscpnet@bailey.psych.nwu.edu
Subject: Sampling and the CR survey

Bill Follette wonders if there are sampling problems in the Consumer Reports article. Lest one think that this study was anything other than scrupulously done, I want to present what I could detect in the way of sampling flaws.

Marty

Consumer Reports (CR henceforward) included a supplementary survey about psychotherapy and drugs in one version of its 1994 annual questionnaire, along with its customary inquiries about appliances and services. One-hundred eighty-thousand readers received this version which included approximately 100 questions about automobiles and about mental health. CR asked readers to fill out the mental health section "if at any time over the past three years you experienced stress or other emotional problems for which you sought help from any of the following: friends, relatives, or a member of the clergy; a mental health professional like a psychologist or a psychiatrist; your family doctor; or a support group." Twenty-two thousand readers responded. Of these, approximately 7000 subscribers responded to the mental health questions. Of these 7000, about 3000 had just talked to friends, relatives or clergy. 4100 went to some combination of mental health professionals, family doctors, and support groups. Of these, 2900 saw a mental health professional: psychologists (37%) were the most frequently seen mental health professional, followed by psychiatrists (22%), social workers (14%) and marriage counselors (9%). Other mental health professionals made up 18%. 1300 joined self-help groups and about 1000 saw family physicians. The respondents, as a whole, were highly educated, predominantly middle class,

about half were women, and they had a median age of 46. Sampling. This survey is, as far as I have been able to determine, the most extensive study of psychotherapy effectiveness on record. The sample is not representative of the United States as a whole, but my guess is that it is roughly representative of the middle class and educated population who make up the bulk of psychotherapy patients. Importantly, the sample represents people who chose to go to treatment for their problems, not people who do not "believe in" psychotherapy or drugs. The CR sample, moreover, is probably weighted toward "problem solvers," people who actively try to do something about what troubles them. Sampling. Is there a bias such that those respondents who succeed in treatment selectively return their questionnaires? CR, not surprisingly, has gone to considerable lengths to find out if its reader's surveys have sampling bias. The annual questionnaires are lengthy and can run to one hundred questions or more. Moreover, the respondent not only devotes a good deal of her own time to filling these out, but also pays her own postage and is not compensated. So the return rate is rather low, although the 12% return rate for this survey was normal for the annual questionnaire. But it is still possible that respondents might differ systematically from the readership as a whole. For the mental health survey (and for their annual questionnaires generally) CR conducted a "validation survey," in which postage was paid and the respondent compensated. This resulted in a return rate of around 40% as opposed to the 13% uncompensated return rate, and there were no differences between data from the two samples.

The possibility of two other kinds of sampling bias, however, is notable, particularly with respect to the remarkably good results for AA. First, since AA encourages lifetime membership, a preponderance of successes, rather than dropouts would be more likely in the three year time slice ("have you had help in the last three years?"). Second, AA failures are often completely dysfunctional and thus much less likely to be reading Consumer Reports and filling out extensive readers' surveys, than, say, psychotherapy failures who were unsuccessfully treated for anxiety.

A similar kind of sampling bias, to a lesser degree, cannot be overlooked for other kinds of treatment failures. At any rate, it is quite possible that there was a large oversampling of successful AA cases and a smaller oversampling of successful treatment for problems other than alcoholism.

Could the benefits of long-term treatment be an artifact of sampling bias? Suppose that people who are doing well in treatment selectively remain in treatment and people who are doing poorly drop out earlier. In other words, the early drop-outs are mostly people who fail to improve, but later drop-outs are mostly people whose problem resolves. CR disconfirmed this possibility empirically: Respondents reported not only when they left treatment but why, including leaving because their problem was resolved. The drop-out rates because the problem was resolved were uniform across duration of treatment (less than one month, 60%; 1-2 months, 66%; 3-6 months, 67%, 7-11 months, 67%; 1-2 years, 67%; over two years, 68%).

A more sweeping limit on generalizability comes from the fact that the entire sample chose their treatment. To one degree or another, each person believed that psychotherapy and/or drugs would help him. To one degree or another, each person acknowledged that he had a problem and believed that the particular mental health professional he saw and the particular modality of treatment he chose would help him. One cannot argue compellingly from this survey that treatment by a mental health professional would prove as helpful to troubled people who deny their problems and who do not believe in and do not choose treatment.

Date: Sun, 22 Oct 1995 18:27:35 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: div12@vm1.nodak.edu, sscpnet@bailey.psych.nwu.edu Subject: multivariate CR
--

Howard Berenbaum suggested that the CR survey was a "reader satisfaction" survey and did not address the question of whether the specific problem--disorder--was ameliorated.

Not so:

CR's analysts decided that no single measure of therapy effectiveness would do and so created a multivariate measure. This composite had three subscales consisting of:

a) specific improvement ("how much did treatment help with the specific problem that led you to therapy: made things a lot better; made things somewhat better; made no difference; made things somewhat worse; made things a lot worse; not sure"),

b) satisfaction ("Overall how satisfied were you with this therapist's treatment of your problems: completely satisfied; very satisfied; fairly well satisfied; somewhat satisfied; very dissatisfied; completely dissatisfied"), and

c) global improvement (how respondents described their "overall emotional state" at the time of the survey compared to the start of treatment: "very poor: I barely managed to deal with things; fairly poor: Life was usually pretty tough for me; so-so: I had my ups and downs; quite good: I had no serious complaints; very good: Life was much the way I liked it to be"). Each was transformed and weighted equally on a 0-100 scale, resulting in a 0-300 scale for effectiveness. The statistical analysis was largely multiple regression, with initial severity and duration of treatment (the two biggest effects) partialled out. Stringent levels of statistical significance were used.

All of the major results held for each of these subscales separately and the composite.

Date: Mon, 23 Oct 1995 08:56:26 -0500
To: kaechele@sip.medizin.uni-ulm.de
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: consumer report

Hi:

Consumer Reports is a monthly magazine published in America. It surveys cars, T.V.'s, coffee, etc. This is their first survey on psychotherapy. I suppose you can get the magazine a place that sells American periodicals. If you can't get it, let me know and I will send you one.

Love

Ken Howard

Date: Mon, 23 Oct 1995 13:30:05 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: follette@unr.edu (Bill Follette),
seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman)
Subject: Re: POD

Marty,

My colleague Paul Fischbeck has an interesting observation about investigating planes for damage at the end of missions in order to determine which parts to reinforce. (Paul is/was an airforce fighter pilot in addition to being a behavioral decision making person.)

Reinforce the parts on which no damage is found.

Once you realize why, you appreciate at least one problem with the type of retrospective sampling CR engaged in.

The AA example perhaps illustrates it. Random assignment studies indicate that AA does no better (or worse) than other treatments--overall. But retrospectively it does better. Why? It's a total life commitment program. Becoming dissatisfied but staying in it anyway is not much of an option (and incidentally, where is your evidence that dropping out indicates a lower probability of giving up alcohol than does staying in it? I hope not from AA sources--which regard dropping out but merely not drinking as not being "cured," or "recovering," or whatever.)

Robyn

Date: Mon, 23 Oct 1995 14:15:34 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: sscpnet@bailey.psych.nwu.edu Subject: retrospection and CR
--

Robyn Dawes seems to think that the CR sampling method is seriously compromised by being retrospective. He also suggests that the problems with oversampling successful AA cases also applies in some unspecified (perhaps Robyn would be good enough to make this argument explicit) way to the CR conclusions as a whole. While I think retrospection is a mild flaw, and the AA results are highly suspect, I do not think either issues affect the validity of the main CR conclusions:

Retrospective. The CR respondents reported retrospectively on their emotional states. While a one-time survey is highly cost-effective, it is necessarily retrospective. Retrospective reports are less valid than concurrent observation, although an exception is worth noting: waiting for the rosy afterglow of a newly completed therapy to dissipate, as the CR study does, may make for a more sober evaluation. The retrospective method does not allow longitudinal observation of improvement across time in the same individuals. Thus the benefits of long term psychotherapy are an inference from comparing different individuals' improvements cross-sectionally. A prospective study would allow comparison of the same individuals' improvements over time. Retrospective observation is a flaw, but it may introduce random rather than systematic noise in the study of psychotherapy effectiveness. The distortions introduced by retrospection could go either in the rosier or more dire direction, but only further research will tell us if the distortions of retrospection are random or systematic.

It is noteworthy that Consumer Reports generally uses two methods: one is the laboratory test, in which, for example, a car is crashed into a wall at five miles per hour and damage to the bumper is measured. The other is the reader's survey. These two methods parallel the efficacy study and the effectiveness study, respectively, in many ways. If retrospection was a fatal flaw, CR would have given up reader's survey method long ago, since reliability of used cars and satisfaction with airlines, physicians, and insurance companies depends on retrospection. Regardless, the survey method could be markedly improved, however, by being longitudinal, in the same way as an efficacy study is. Both self-report and diagnosis could be done before and after therapy as well as a thorough follow up carried out. But retrospective reports of emotional states will always be with us, since even in a prospective study that begins with a diagnostic interview, the patient retrospectively reports on her (presumably) less troubled emotional state before the diagnosis.

AA sampling bias. The possibility of two other kinds of sampling bias, however, is notable, particularly with respect to the remarkably good results for AA. First, since AA encourages lifetime membership, a preponderance of successes, rather than dropouts would be more likely in the three year time slice ("have you had help in the last three years?"). Second, AA failures are often completely dysfunctional and thus much less likely to be reading Consumer Reports and filling out extensive readers' surveys, than, say, psychotherapy failures who were unsuccessfully treated for anxiety.

A similar kind of sampling bias, to a lesser degree, cannot be overlooked for other kinds of treatment failures. At any rate, it is quite possible that there was a large oversampling of successful AA cases and a smaller oversampling of successful treatment for problems other than alcoholism.

Could the benefits of long-term treatment be an artifact of sampling bias? Suppose that people who are doing well in treatment selectively remain in treatment and people who are doing poorly drop out earlier. In other words, the early drop-outs are mostly people who fail to improve, but later drop-outs are mostly people whose problem resolves. CR disconfirmed this possibility empirically: Respondents reported not only when they left treatment but why, including leaving because their problem was resolved. The drop-out rates because the problem was resolved were uniform across duration of treatment (less than one month, 60%; 1-2 months, 66%; 3-6 months, 67%, 7-11 months, 67%; 1-2 years, 67%; over two years, 68%).

Date: Mon, 23 Oct 1995 16:08:19 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: sscpnet@bailey.psych.nwu.edu, seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman) Subject: Re: retrospection and CR

Marty,

My point was that there is at least one systematic bias with retrospective analyses of psychotherapy outcome studies. The subjects have to be around to fill out the questionnaires. A bad enough result, and a subject might not be. When we sample prior to therapy, we don't have to worry about that (in a "good" study). If a person is in a hospital, or has dropped out socially, or has dropped out of life altogether, such bad outcomes can be related to group assignment. Not so if such people are not there so that we do not know what treatment they received.

Well, what about the product assessments? They do not suffer from the same POTENTIAL problem, or at least not as much. Except if extreme cases, if your dishwasher fails, that will not imply that you are no longer at home to subscribe to CR. If your treatment for alcoholism, drug usage, depression, or ("counterphobic"?) risk-taking fails, however, you might not be.

If we're dealing with serious stuff, then looking only at the planes that made it back can give a very biased impression. In fact--and I want to make clear that this is just a LOGICAL possibility--any treatment that was so pernicious that a lot of the more severely afflicted died, while many of the rest got better anyway, would look great if you sampled only on a retrospective basis.

Robyn

From: "Jennifer Lish" <JENNIFER@compass.compass-is.com> To: sscpnet@bailey.psych.nwu.edu, seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman) Date: Mon, 23 Oct 1995 16:15:18 EDT Subject: Re: retrospection and CR
--

It would help me to understand the CR study to know exactly the questions that were asked. In particular, as I think has been mentioned on this network, we find that consumer satisfaction assessed with one of Atkinson's questionnaires has zero correlation with outcome (symptoms, functioning and well-being at treatment termination minus the same at treatment initiation). So consumers could be 100% satisfied (a worthwhile goal in itself) even though they did not get better (all their bumpers were destroyed in the crash test). The efficacy-effectiveness analogy only describes the difference between assessing improvement in mental health treatment in a randomized controlled trial and in the CR survey if the CR survey asked the same outcome questions as are asked in RCTs -- "were the patient's target psychiatric symptoms relieved and his functioning restored?" -- not "was the treatment satisfactory. Just as the CR question about my car is "how many times did you have to take it for repairs", not "are you satisfied with its repair history".

Jennifer D. Lish, Ph.D.
Research Scientist
Compass Information Services
Phone: 610-992-7073
Fax: 610-992-7070
Adjunct Assistant Professor, Department of Psychiatry
Medical College of Pennsylvania and Hahnemann University

Date: Mon, 23 Oct 1995 16:16:05 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: retrospection and CR

Lee Sechrest, Univ. of Arizona

On Mon, 23 Oct 1995, Martin E. P. Seligman wrote:
--much deleted--

>Could the benefits of long-term treatment be an artifact of sampling bias? Suppose that people who are doing well in treatment selectively remain in treatment and people who are doing poorly drop out earlier. In other words, the early drop-outs are mostly people who fail to improve, but later drop-outs are mostly people whose problem resolves. CR disconfirmed this possibility empirically: Respondents reported not only when they left treatment but why, including leaving because their problem was resolved. The drop-out rates because the problem was resolved were uniform across duration of treatment (less than one month, 60%; 1-2 months, 66%; 3-6 months, 67%, 7-11 months, 67%; 1-2 years, 67%; over two years, 68%).<

Those data are just too remarkably consistent to be easily believable. They cry out the warning "artifact!" Dick Bootzin and I were just discussing these data, and here is an alternative take:

About two thirds of people who go into therapy experience problems from which, eventually, they are likely to recover, i.e., the problem will be resolved. They stay in therapy until the problem is resolved, resolves itself, or whatever, and then, "cured," they leave therapy. Some absolutely famous psychologist suggested a long time ago that about 2/3 of people get over their problems no matter what is done for them. Presumably people will not stay in therapy once their problem has gone away.

The coincidence that 2/3 of people leave therapy because they have been cured by therapy, no matter how long the therapy, is just too much of a stretch. That means there is *no* dose-response relationship. Do we know that the people who stayed in longer had more serious problems and, hence, needed more therapy? Note that the data pertain to those people whose problem was *resolved*, not just better.

Something does not add up here.

Date: Mon, 23 Oct 1995 16:20:38 -0500 (CDT) From: HOUTSAC@MSUVX1.MEMPHIS.EDU (ART HOUTS University of Memphis) Subject: Re: Prevalence of mental illness To: k-howard@nwu.edu (Ken Howard)

Ken Howard seems to place a lot of faith in the validity of the ECA studies to bolster arguments for why the US needs more mental health treatment. For some time, I have been skeptical about some things regarding these studies. (1) Given the published sensitivities and specificities of the DIS, it seems highly likely that DIS will overestimate disorders, especially if the theoretical base rate of the disorder is low. In fact, does anyone know how the ECA

studies manage to get some of the more believable incidence figures they do? How do they eliminate false positives in particular. I have not seen this addressed in all of the reports I have read. What did I miss?

(2) The utilization figures cut both ways on the validity question about the DIS. Apparently, two thirds of the people with a DIS diagnosis get by without treatment. Also, 50% of the people who seek treatment do not have a DIS diagnosis (Regier et al AGP, 50, 85-94). Now, something is terribly wrong if the DIS is to be used as a gauge of the country's mental health. Granted, seeking treatment may not be the best criterion measure for mental "illness" but if two thirds of the putative mentally do not go for treatment, maybe many of them are not mentally ill after all.

Date: Mon, 23 Oct 1995 16:25:20 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: multivariate CR
--

Dear Marty,

One of my favorite quotes on research methodology comes from Alan Kazdin's textbook, where he defines "face validity" as the embarrassment one feels when one has to face one's colleagues with an instrument devoid of psychometric properties. What can you tell us about the reliability and validity of the outcome measures used in the CR study?

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

Date: Mon, 23 Oct 1995 18:40:42 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: div12@vm1.nodak.edu, sscpnet@bailey.psych.nwu.edu Subject: "Satisfaction" and Cr
--

Jennifer Lish wonders if this was a reader "satisfaction" survey. At the risk of repeating my previous posts: It was not. Satisfaction was one peripheral question, but the heart of the survey consisted of "How much therapy helped (from "made things a lot better" to "made things a lot worse") and in what areas (specific problem that led to therapy, relations to others, productivity, coping with stress, enjoying life more, growth and insight, self-esteem and confidence, raising low mood)." Here are some details (If you want the original questionnaire I suggest phoning Mark Kotkin 914 378 2253, the principal CR analyst, since I assume it is proprietary.)

Twenty-six questions were asked about mental health professionals, and parallel, but less detailed questions were asked about physicians, medications and self-help groups:

- What kind of therapist
- What presenting problem (e.g., general anxiety, panic, phobia, depression, low mood, alcohol or drugs, grief, weight, eating disorders, marital or sexual problems, children or family, work, stress)
- Emotional state at outset (from very poor to very good)
- Emotional state now (from very poor to very good)
- Group versus individual therapy
- Duration and frequency of therapy
- Modality (psychodynamic, behavioral, cognitive, feminist)
- Cost
- Health care plan and limitations on coverage
- Therapist competence
- How much therapy helped (from "made things a lot better" to "made things a lot worse") and in what areas (specific problem that led to therapy, relations to others, productivity, coping with stress, enjoying life more, growth and insight, self-esteem and confidence, raising low mood)
- Satisfaction with therapy
- Reasons for termination (problems resolved or more manageable, felt further treatment wouldn't help, therapist recommended termination, a new therapist, concerns about therapist's competence, cost, and problems with insurance coverage).

CR's analysts decided that no single measure of therapy effectiveness would do and so created a multivariate measure. This composite had three subscales consisting of:

- a) specific improvement ("how much did treatment help with the specific problem that led you to therapy: made things a lot better; made things

somewhat better; made no difference; made things somewhat worse; made things a lot worse; not sure"),

b) satisfaction ("Overall how satisfied were you with this therapist's treatment of your problems: completely satisfied; very satisfied; fairly well satisfied; somewhat satisfied; very dissatisfied; completely dissatisfied"), and

c) global improvement (how respondents described their "overall emotional state" at the time of the survey compared to the start of treatment: "very poor: I barely managed to deal with things; fairly poor: Life was usually pretty tough for me; so-so: I had my ups and downs; quite good: I had no serious complaints; very good: Life was much the way I liked it to be")

General Functioning. The CR study measured self-reported changes in well-being, insight, and growth, in addition to improvement on the presenting problem. The main findings held for these "general functionality" measures as well as for symptom relief: for example, long term treatment produced better quality of life scores than short term treatment, and mental health professionals did better than family doctors on general function scores as well as symptom reduction for treatment which lasted longer than six months. Since improvement in general functioning, as well as symptom relief, is almost always a goal of actual treatment, but rarely of efficacy studies, the CR study adds to our knowledge of how treatment does beyond merely eliminating symptoms.

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: retrospection and CR To: sechrest@U.ARIZONA.EDU (Lee B Sechrest) Date: Mon, 23 Oct 1995 19:14:18 -0400 (EDT)

Those who had more therapy, also had significantly more severe problems at the outset.

These data do cry out for a careful scrutiny. It is my hope that CR will allow us, after their lawyers scrutinize all contingencies, to form groups to analyze and reanalyze the original data set. I judge it to be worth quite a lot of careful scrutiny.

If you are interested let me know. Particularly the subset of data that interest you.

Date: Tue, 24 Oct 1995 02:27:06 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: div12@vm1.nodak.edu, sscpnet@bailey.psych.nwu.edu Subject: face validity and CR

Neil Jacobson worries about the reliability and face validity of the measures in the CR survey. The reliabilites were pretty good as I recall (I don't have the data set in my possession), but it cannot be denied that the items were primarily "face valid." I think there is something to such a worry, but I don't see how any large-scale effectiveness study can be done without a preponderance of merely face valid items--and still be short enough to get a reasonable response rate.

It is possible to go entirely overboard, as Kazdin is represented as doing by Neil, in denying the usefulness of some face valid items. It would be a fool's errand to bother to get more than face validity on such CR items as "Was your therapist easy to confide in?" or "Did you check out other therapists before selecting this one?" or even "I felt my therapy helped being productive with work (from made things a lot better to made things a lot worse)." I suppose I can see a generation of research assistants finding that this actually correlates about .55 with ingeniously measured behavioral indices, but I would not be enthusiastic about setting anyone such a task.

There is some "beef" in Neil's worry, but I would put the cavil slightly differently:

Inadequate Outcome Measures. CR's indexes of improvement were molar. Responses like "made things a lot better" to the question "How much did therapy help you with the specific problems that led you to therapy?" tap into gross processes. More molecular assessment of improvement, for example, "how often have you cried in the last two weeks," or "how many ounces of alcohol did you have yesterday" would increase the validity of the method. Such detail would, of course, make the survey more cumbersome.

A variant of this objection is that the outcome measures were insensitive. This objection looms large for the failure to find that any modality of therapy did better than any other modality of therapy, or any drug for that matter, for any disorder. Perhaps if more detailed, disorder-specific measures were used, the Dodo bird hypothesis would have been disconfirmed. A third variant of this objection is that the outcome measures were poorly normed. Questions like "How satisfied were you with this therapist's treatment of your problem? Completely satisfied, very satisfied, fairly well satisfied, somewhat dissatisfied, very dissatisfied, completely dissatisfied," and "How would you describe your overall emotional state? very poor: I barely managed to deal with things; fairly poor: Life was usually pretty tough for me; so-so: I had my ups and downs; quite good: I had no serious concerns; very good: Life was much the way I wanted it to be" are seat-of-the-pants items which depend almost entirely on face validity, rather than several generations of norming. So the conclusion that 90% of those people who started off "very poor" or "fairly poor" wound up in the "very good," "fairly good," or "so-so" categories does not guarantee that they had returned to normality in any

strong psychometric sense. The addition of extensively normed questionnaires like the Beck Depression Inventory would strengthen the survey method (and make it more cumbersome).

Date: Thu, 26 Oct 1995 17:12:53 -0600 (MDT) From: JAMES MICHAEL WOOD <jwood@utep.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: CR and Control Groups
--

Jim Wood
University of Texas at El Paso

Dear Dr. Seligman,

You have generated a very interesting series of exchanges on the net. Could I ask just a couple questions related to the Consumer Reports (CR) study?

1. You mention that you are trying to persuade the CR attorneys to release the study data. Does that mean that American Psychologist has agreed to publish your methodological analysis of the study even though the data are unavailable for examination by other researchers? Isn't this a switch from normal APA policy regarding public availability of data?

2. You have said that you consider the CR report the best (or perhaps the "most important") study ever conducted on the "effectiveness" of psychotherapy. Does this mean that you consider its findings more reliable or important than the findings of effectiveness studies that have used (a) control groups, (b) random assignment, (c) standardized measures, of known reliability and validity, and (d) contemporaneous, rather than retrospective, data collection? Or would you say that any "effectiveness" study with these characteristics is therefore really an "efficacy" study? It's my impression, at least in medicine, that "effectiveness" studies can have all these characteristics, which I'd consider pretty desirable.

I'm looking forward to your clarification. Thanks again for introducing "life" into the net.

Jim Wood
Department of Psychology
University of Texas at El Paso

Date: Thu, 26 Oct 1995 17:21:10 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Consumer Reports article
To: seligman@CATTELL.PSYCH.UPENN.EDU

Dear Marty:

I have been trying for a couple of days to find a copy of the November Consumer Reports to read the article you and others were discussing. Finally, last night, I found the magazine in a bookstore on my way home.

The article was pleasant to read and made me feel good that these sophisticated consumers have such a high opinion of psychologists and other mental health specialists, and of psychotherapy as a worthwhile service. However, in the absence of the SSCP net discussion, I don't think I it would have occurred to me to consider it as any kind of evidence relevant to the issue of the efficacy of psychotherapy.

Neil Jacobson's humorous citation of Kazdin's definition of face validity, however, kept reverberating in my thoughts. For those who did not read Jacobson or Kazdin, this was something like 'the embarrassment one feels in facing colleagues without evidence on the psychometric properties of a measure.' I don't think Consumer Reports bothers with the psychometric properties of its questions about car bumpers, and it would seem like an almost impossible chore to convince them that this is necessary when asking consumers questions about psychotherapy. Did you try to persuade them to use more rigorous scientific methods in their survey? If so, with what results?

Don

Date: Thu, 26 Oct 1995 17:24:29 -0400
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: div12@vm1.nodak.edu, doris@psych.nyu.edu, sscpnet@bailey.psych.nwu.edu
Subject: CR and Control Groups

Doris Aaronson raises the important objection to the CR article that there are no control groups. Not unsurprisingly, CR (and I) have thought quite a lot about this issue:

No Control Groups. The overall improvement rates were strikingly high across the entire spectrum of treatments and disorders in the CR study. The vast majority of people who were feeling very poorly or fairly poorly when they entered therapy made "substantial" (now feeling fairly or very good) or "some" (now feeling so-so) gains.

Perhaps the best news for patients was that those with severe problems got, on average, much better. While this may be a ceiling effect, it is a ceiling effect

with teeth. It means that if you have a patient with a severe disorder now, the chances are quite good that he will be much better within three years. But methodologically, such high rates of improvement are a yellow flag, cautioning us that global improvement over time alone, rather than with treatment or medication, may be the underlying mechanism.

More generally because there are no control groups, the CR study cannot tell us directly whether talking to sympathetic friends or merely letting time pass would have produced just as much improvement as treatment by a mental health professional. The CR survey, unfortunately, did not ask those who just talked to friends and clergy to fill out detailed questions about the results.

This is a serious objection, but there are internal controls which perform many of the functions of control groups: First, marriage counselors do significantly worse than psychologists, psychiatrists, and social workers, in spite of no significant differences in kind of problem, severity of problem, or duration of treatment. Marriage counselors control for many of the nonspecifics such as therapeutic alliance, rapport, and attention, as well as for passage of time. Second, there is a dose-response curve with more therapy yielding more improvement.

The first point in the dose-response curve approximates no treatment: people who have less than one month of treatment have on average an improvement score of 201, whereas people who have over two years of treatment have a score of 241. Third, psychotherapy does just as well as psychotherapy plus drugs for all disorders, and there is such a long history of placebo controls inferior to these drugs that one can infer that psychotherapy likely would have outperformed such controls had they been run. Fourth, family doctors do significantly worse than mental health professionals when treatment continues beyond six months. It might be objected that since total length of time in treatment, rather than total amount of contact is the covariate, comparing family doctors who don't see their patients weekly to mental health professionals--who see their patients once a week or more--is not fair. It is, of course, possible that if family doctors saw their patients as frequently as psychologists do, the two groups would do equally well. It was notable, however, that there were a significant number of complaints about family doctors: 22% of respondents said the doctor had not "provided emotional support," 15% said that the doctor "seemed uncomfortable discussing emotional issues," and 18% said the doctor was "too busy to spend time talking to me." At any rate, the CR survey shows that long-term family doctoring for emotional problems--as it is actually performed in the field--is inferior to long-term treatment by a mental health professional as it is actually performed in the field.

It is also relevant that the patients attributed their improvement to treatment and not time ("How much do you feel that treatment helped you in the following areas?"), and I conclude that the benefits of treatment are very unlikely to be caused by the mere passage of time. But I also conclude that the CR study could be improved by control groups matched for severity and kind of problem, that are not treated by mental health professionals (but beware of

the fact that random assignment will not occur). This would allow the Bayesian inference that psychotherapy does better than talking to friends, seeing an astrologer, or going to church to be made more confidently.

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: Consumer Reports article To: drouth@umiami.ir.miami.edu (Donald K. Routh) Date: Thu, 26 Oct 1995 17:32:41 -0400 (EDT)
--

Don:

As you can see from my verbose responses about control groups, face validity, retrospection, etc., I consider the CR piece to be quite rigorously done. In fact, I consider it to be the most extensive and best carried out study on the *effectiveness*--not the *efficacy*--of psychotherapy ever published.

I have a long methodological article coming out in the December American Psychologist about this topic.

Perhaps you can make explicit any thoughts you have on why it is not rigorous and we can all debate them.

best wishes,
Marty

Date: Thu, 26 Oct 1995 17:48:40 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: CR Article To: seligman@CATTELL.PSYCH.UPENN.EDU
--

Dear Marty:

To begin with, what information is there about the reliability of the items or scales in the survey (whether this is in the form of test-retest reliability, internal

consistency, or what not)?

Don

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: CR Article To: drouth@umiami.ir.miami.edu (Donald K. Routh) Date: Thu, 26 Oct 1995 18:03:03 -0400 (EDT)
--

Internal consistency was analyzed for, and as I recall--I do not possess the dataset--it is owned by CR, they were quite good. CR's statistical analysts are a serious and sophisticated lot.

My long term hope is that I can convince CR's lawyers--I am now trying--to release the entire dataset to responsible investigators, so that we can document such matters ourselves.

Marty

>Dear Marty:

To begin with, what information is there about the reliability of the items or scales in the survey (whether this is in the form of test-retest reliability, internal consistency, or what not)?<

>Don<

Date: Thu, 26 Oct 1995 19:13:45 -0600 (MDT) From: JAMES MICHAEL WOOD <jwood@utep.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: CR and Control Groups
--

Jim Wood
Department of Psychology
University of Texas at El Paso

Dear Dr. Seligman,

Thanks for your quick reply. One of the great things about Internet is that it allows discussions like this to occur quickly, and with input and observations from many other researchers and clinicians.

My understanding of your postings to the net is as follows:

(a) The Consumer Reports (CR) study used retrospective self-reports from subjects, (b) the measures were not well-accepted standardized measures, with established reliability and validity, (c) there were no control groups, (d) there was no randomization, (e) the data of the CR study are not available for examination by other psychologists, (e) despite these limitations, you consider the CR study to be the "best" or "most important" effectiveness study ever conducted on psychotherapy outcome, (f) despite these limitations, the _American Psychologist_, under the editorship of Ray Fowler, has accepted your methodological analysis of the CR study for publication.

This is quite a list, I'm sure you will agree. Do I have my facts straight?

As a psychologist concerned about the public accessibility of scientific data, the reliability and validity of outcome measures, and the possibility of regression, selection and reporting artifacts, I must say that the CR study does not seem very convincing to me, even considering your very intelligent and spirited defense.

Again, with appreciation for your willingness to participate in dialogue.

Jim Wood
Department of Psychology
University of Texas at El Paso

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: CR and Control Groups To: jwood@utep.edu (JAMES MICHAEL WOOD) Date: Thu, 26 Oct 1995 19:43:24 -0400 (EDT)
--

>Jim Wood
>University of Texas at El Paso<

>Dear Dr. Seligman,<

>You have generated a very interesting series of exchanges on the net. Could I ask just a couple questions related to the Consumer Reports (CR) study?<

>1. You mention that you are trying to persuade the CR attorneys to release the study data. Does that mean that _American Psychologist_ has agreed to publish your methodological analysis of the study even though the data are unavailable for examination by other researchers? Isn't this a switch from normal APA policy regarding public availability of data?<

I have no idea what APA policy is on methodological critiques of such studies. I seem to recall that as an author of a research report, I must make my data public if asked. MY AP article is not a research report. You should direct your question to Ray Fowler.

>2. You have said that you consider the CR report the best (or perhaps the "most important") study ever conducted on the "effectiveness" of psychotherapy. Does this mean that you consider its findings more reliable or important than the findings of effectiveness studies that have used (a) control groups, (b) random assignment, (c) standardized measures, of known reliability and validity, and (d) contemporaneous, rather than retrospective, data collection? Or would you say that any "effectiveness" study with these characteristics is therefore really an "efficacy" study? It's my impression, at least in medicine, that "effectiveness" studies can have all these characteristics, which I'd consider pretty desirable.<

The CR article is IMHO a very serious piece of research which merits careful scrutiny by our community. Because I know of no psychotherapy effectiveness studies which meet Jim Wood's desiderata, or even comes close, the CR study is about the best we have that I know of. In the AP article I outline desiderata for such a future study very much like Jim's. I hope the CR study will inspire our community to carry out such an ideal study of psychotherapy effectiveness.

>I'm looking forward to your clarification. Thanks again for introducing "life" into the net.<

>Jim Wood
Department of Psychology
University of Texas at El Paso<

Date: Thu, 26 Oct 1995 20:56:12 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Chrysler Corporation and the Psychotherapy Industry
--

Marty, and all,

CR has a very firm rule against use of its product evaluations by manufacturers. Chrysler, for example, could not claim in its ads that it was rated "best" (should that ever happen) by CR. In fact, CR has gone to court to prevent just such advertising.

I like the CR study, I really do. I do not regard it as quite so unambiguously interpretable as you do, but it is a good contribution to our entire array of empirical inquiries into the effectiveness of psychotherapy. It provides knowledge that may prove quite helpful in deciding just how much good psychotherapy might be counted on to do.

My reservations about the CR study as a contribution to the field would virtually vanish if I thought that the psychotherapy industry would be prohibited (or would exercise self-restraint) from using it for PR/advertising purposes. Chrysler has survived; why not the psychotherapy industry?

As useful as the CR study may be when put in the context of all the other (500+ Ken Howard says) psychotherapy outcome studies. Even if the CR study is the best ever, as you contend, it is a weak reed, if used alone, as a basis for public policy.

Lee

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
Subject: Re: CR and Control Groups
To: jwood@utep.edu (JAMES MICHAEL WOOD)
Date: Thu, 26 Oct 1995 21:33:14 -0400 (EDT)

>Jim Wood
Department of Psychology
University of Texas at El Paso<

>Dear Dr. Seligman,<

>Thanks for your quick reply. One of the great things about Internet is that it allows discussions like this to occur quickly, and with input and observations from many other researchers and clinicians.<

I agree

Without putting my whole AP article on the net now, I cannot say too much more than I have already said re control groups, randomization, sampling, self-report, reliability, and retrospection-- this is much of the substance of my article. But I'll pipe you a preprint.

>My understanding of your postings to the net is as follows:<

>(a) The Consumer Reports (CR) study used retrospective self-reports from subjects, (b) the measures were not well-accepted standardized measures, with established reliability and validity, (c) there were no control groups, (d) there was no randomization, (e) the data of the CR study are not available for examination by other psychologists, (e) despite these limitations, you consider the CR study to be the "best" or "most important" effectiveness study ever conducted on psychotherapy outcome, (f) despite these limitations, the _American Psychologist_, under the editorship of Ray Fowler, has accepted your methodological analysis of the CR study for publication.<

>This is quite a list, I'm sure you will agree. Do I have my facts straight?<

>As a psychologist concerned about the public accessibility of scientific data, the reliability and validity of outcome measures, and the possibility of regression, selection and reporting artifacts, I must say that the CR study does not seem very convincing to me, even considering your very intelligent and spirited defense.<

I have the same concerns, and it would be repititious for me to say that I consider the CR article and its database as serious science indeed. I am doing my best to persuade them to open the dataset for public examination.

>Again, with appreciation for your willingness to participate
in dialogue.<

I am sure we will continue this after the December AP comes out.

>Jim Wood
Department of Psychology
University of Texas at El Paso<

Date: Fri, 27 Oct 1995 08:17:36 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: CR article (2) To: seligman@CATTELL.PSYCH.UPENN.EDU
--

Dear Marty:

It is good to know that the CR statisticians looked at the internal consistency of the scales. An unfortunate feature of journalistic articles is that they do not provide such information or even indicate where an interested reader could obtain it. I think the scientific credibility of such work depends upon the public availability of such technical information. Don't get me started about the lawyers who are involved in the process.

In order to come up with my next query, I brought my copy of the November issue of CR to the office with me this morning. Placing myself in the role of consumer, I tried to respond to some of the questions (referring to my own most significant encounter with a therapist, between the ages of 21 and 22). I would classify my emotional state at the time of entering treatment as "fairly poor" ("Life was usually pretty tough"). Referring to the figures on p. 736 of the article, my experience would place me in the category of people treated more than six months who were "helped a lot." Referring to the figure on p. 738, I would classify this treatment episode as one involving a mental health specialist rather than a family doctor and the use of both drugs and therapy.

Now, my actual experience was that I did a lot better after completing therapy. Life has hardly ever seemed that tough to me again (it only seemed so during the time after our adult daughter was killed in an automobile accident, and that is quite understandable as bereavement rather than depression). However, I don't think I will ever be sure that "doing a lot better" was in fact due to either the therapy or the drugs. After all, there is a real possibility (given the absence of a control condition in the "study") that the natural course of my suffering would have been self-limiting, and I would have gotten better anyway. This interpretation brings to light some of the problems in interpreting surveys such as that described in the CR article. Don

Date: Fri, 27 Oct 1995 11:50:32 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: CR and Control Groups

Lee Sechrest, Univ. of Arizona

On Fri, 27 Oct 1995, Martin E. P. Seligman wrote:

>Surely a prospective effectiveness study would be superior to a retrospective effectiveness study. I see the CR study as a serious beginning that might justifiably inspire our scientific community and funding agencies, as well as apa, to undertake more ideal versions: prospective, blind diagnosis, extensive validated measures, control groups, and the like.<

Marty, I do not think anyone would take issue with the foregoing statement. The problems arise from previous statements that make it seem that you may take the CR study as finally definitive evidence that psychotherapy really does a lot of good.

Date: Fri, 27 Oct 1995 12:00:57 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>,
JAMES MICHAEL WOOD <jwood@utep.edu>
Subject: Re: CR and Control Groups

Jim, Marty

Am I off base? I have mentioned more than once the problem of looking at where the bullet holes **aren't** on the planes that come back, rather than looking at where they are. (Lee Sechrest tells me this original observation is due to Abraham Wald in WWII.) If your microwave fails, that problem most likely won't affect your probability of continuing a subscription to CR and being around to answer questionnaire in it. If you have a serious psychology problem and your therapy fails, you might not be around or interested in answering. But the whole data set is based on sampling people who HAVE BEEN in therapy and --like the returning planes with bullet holes--are available for analysis.

If we work prospectively rather than retrospectively, we don't have that problem. Sure, it might be hard to contact drop-outs (of therapy, of their previous role, or even of life), but that is not a problem with the **logic** of the

design. It's a problem of implementation. In contrast, the retrospective design has that problem intrinsically.

Is my concern misplaced?

Robyn

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: CR and Control Groups To: rd1b+@andrew.cmu.edu (Robyn Dawes) Date: Fri, 27 Oct 1995 12:27:05 -0400 (EDT)

Surely a prospective effectiveness study would be superior to a retrospective effectiveness study. I see the CR study as a serious beginning that might justifiably inspire our scientific community and funding agencies, as well as APA, to undertake more ideal versions: prospective, blind diagnosis, extensive validated measures, control groups, and the like.

But as I suggest in my AP article and so far on the net, these less-than-ideal conditions do not easily explain away the major, and robust, results of the CR study as it stands.

From: "Jennifer Lish" <JENNIFER@compass.compass-is.com> To: sscpnet@bailey.psych.nwu.edu Date: Fri, 27 Oct 1995 14:09:28 EDT Subject: Re: CR and Control Groups
--

I mentioned before that patient satisfaction measured at the end of treatment is not correlated ****at all**** with change in clinical status (well-being, symptoms and functioning) determined by measuring clinical status at treatment initiation and termination. This may be because what determines patient satisfaction is things like the waiting time for an appointment, attractiveness of the waiting room, or warmth of the therapist, rather than whether symptoms are relieved and functioning restored. However, it may alternatively be because people do not have accurate memories of exactly their clinical state at treatment initiation and termination and the ability to compare these and state whether they improved, and this is what the CR question asked them to do.

I also agree with the person who said he cannot say whether treatment helped him because he does not know how quickly he would have gotten better spontaneously. This is one of many reasons why efficacy studies and

effectiveness studies are used instead of retrospective surveys of patients to make public health policy decisions in medical-surgical areas. Would you use a survey of patients to decide whether hypertension should be medicated? No, because patients would be unable to vote "yes, it should, because I am dead and my death could have been prevented by treatment of my high blood pressure", and because a lot of patients would vote "no, because the side effects are really uncomfortable", whereas we have a legitimate societal interest and a legitimate interest since his medicare or other health insurance premiums are taken out of our paychecks, in seeing that his doctor makes a big effort to persuade him to take blood pressure medication.

Another issue I have with this whole survey is that what I want from consumer reports is for them to try to answer the question "What is the best way to heat houses in the Northeastern United States?" and then tell me what their tests (including objective studies and surveys of consumers) showed about gas vs. electric vs. coal vs. oil vs. etc., not for them to survey kerosene users who live in Arozona and tell me they are all happy with kerosene. The paralell is to see if somebody has a mental illness (lives in NE) and ask them what they do about it and how satisfactory the different options are. CR readership surely underrepresents people with serious mental illnesses, and people who feel that they had a mental illness and chose to not get treatment of any sort were not included in the survey. If they all feel that they got better just like the people who got therapy, that sure undermines the argument that the CR study shows that therapy is effective. And under-representing people with serious mental illnesses makes the data in the CR report of questionable applicability to the average patient walking into the average therapist's office.

Jennifer D. Lish, Ph.D.
Research Scientist
Compass Information Services
Phone: 610-992-7073
Fax: 610-992-7070
Adjunct Assistant Professor, Department of Psychiatry
Medical College of Pennsylvania and Hahnemann University

Date: Fri, 27 Oct 1995 17:01:34 -0500 (CDT)
From: HOUTSAC@MSUVX1.MEMPHIS.EDU (ART HOUTS University of Memphis)
Subject: Re: CR and Control Groups
To: Robyn Dawes <rd1b+@andrew.cmu.edu>

The late Milt Trapold told me the following story. John Lily (?Lilly), the neurophysiologist who did flotation and LSD research, also became convinced that it might be possible to communicate with dolphins. Late 60s or early 70s, Lily was giving a talk about his dolphin research to a packed house at the University of Minnesota. He waxed eloquently about how dolphins must be of superior intelligence and recounted all of the stories about how dolphins had rescued humans who had been shipwrecked at sea by carrying the humans to shore. At the end of the talk, a rather meek and unassuming Ken MacCorquodale arose and said, Very interesting Dr. Lily, but what about all the humans the dolphins carried the other way?

Date: Fri, 27 Oct 1995 16:33:03 -0700 (PDT)
From: Bill Follette <follette@unr.edu>
To: ART HOUTS University of Memphis <HOUTSAC@msuvx1.memphis.edu>
Subject: Re: CR and Control Groups

Art-
Those dolphins had MPD.

Bill

On Fri, 27 Oct 1995, ART HOUTS University of Memphis wrote:

>see above<

Date: Fri, 27 Oct 1995 17:17:42 -0400
From: stricker@sable.adelphi.edu
To: sscpnet@bailey.psych.nwu.edu
Subject: The CR study

This past week, two different students told me that clinics at which they were placed had duplicated the CR report and left it in the waiting room for patients to read. I think this pattern will be repeated and, as a result, that project may prove to be the most influential psychotherapy study ever done - not the best or the most important, but the most influential, because it will reach the consumers in a way that no article in a peer review forum ever does. The responses to that article often confirm an observation I have long noted about the negative correlation between our exercise of our methodological skills and

our comfort with the findings of a study. Those who are happy with the findings make little of it being a retrospective study based on self-report data from the standpoint of a single observer, with no controls. Those who are unhappy with it see those issues clearly, but not that it provides remarkable evidence of the subjective satisfaction of consumers with a procedure addressed to what often is their initial subjective discomfort. Rather than leave Marty by himself on this one, I would like to echo his recognition of the importance of these findings and the internal analyses that add to their weight, while also echoing his wish that future work be constructed in a more scientifically rigorous manner.

George Stricker

Internet address (STRICKER@SABLE.ADELPHI.EDU)

Telephone (516) 877-4803 Fax (516) 877-4805

Date: Mon, 30 Oct 1995 11:55:51 -0400 (EDT) From: "Al Lang" <allang@darwin.psy.fsu.edu> To: seligman@CATTELL.PSYCH.UPENN.EDU, sscpnet@bailey.psych.nwu.edu Subject: CR, psychotherapy, and other folk remedies

Dear Dr. Seligman et al.,

From time to time, discussions on the net push me over the edge and I cannot help but enter the fray even though I probably should be doing other things [perhaps I need therapy]. This is one of those times. I am not a psychotherapy outcome researcher, I am not hawking a book for, against, or about psychotherapy. I am just a concerned and puzzled observer seeking a few answers. Maybe someone can help.

I've subscribed to and read Consumer Reports for many years. Their apparently thorough, objective tests help me buy good toasters and stuff like that. I never fill out their annual survey and have sometimes wondered who does and why. It must be a fairly select group. I might trust their opinions on matters of clearly observable fact, but I have grave doubts about their ability to evaluate change in their psychological functioning, let alone give valid reasons for what caused it, if it occurred at all. The scientific literature is full of studies suggesting that people are not very good at determining what caused their behavior even in controlled experimental settings (see review by Nisbett & Wilson in Psych Rev about 15 years ago). Why should one expect them to do better when operating out in the wild, under circumstances fraught with confounds such as expectancies, cognitive dissonance, regression to the mean, etc. Then, of course, there is the lack of control groups, problems of attrition, and so on to which others have alluded. But, I disgress. What I really want to call attention to is this:

The cover story of the Nov issue of CR in which the therapy study appears is on "mystery cures," apparently in reference to the article starting on p. 698 which deals with herbal remedies, although it is somehow appropos that the MH therapy article appears in the same issue. I am going to insert in brackets some substitute wording for comments in the herbal article and ask you to consider their implications for the therapy article:

"...if you decide you want to give herbal medicine [MH therapy] a try, you face a formidable obstacle: The supplement [therapy] marketplace is a shambles. There is no guarantee that the pills [treatments] are what they say they are--in most cases, no one really knows what will happen if you take them. You have no way to be sure:

- *Whether a plant's [treatment's] active ingredients, whatever they might be, have actually ended up in the herbal pills [interventions] you buy.
- *Whether a supplement's [therapy's] ingredients are in a form your body [mind] can use.
- *Whether the dosage makes any sense.
- *What else is in the pills [therapy].
- *Whether the pills [treatments] are safe.
- *Whether the next bottle [purveyor] of the same pills [treatments] will have the same ingredients."

The article goes on to provoke skepticism about typically unsubstantiated claims of success using herbs, reports on some in-house tests of the claims, and recommends some books that give a critical appraisal of herbal remedies. In general, the key issues are raised forcefully, the advice is sound, and the tenor of the report is one of consumer protection and the suggestion that people should think twice about spending a lot of money on herbal medicines. This is what I expect of CR. I note that CR did not bother to survey its readership about their experiences with herbal remedies, perhaps for fear that they would get lots of glowing endorsements from true believers.

Now, my first question is: Why didn't MH therapy get the same careful critique? Surely, it wasn't for lack of qualified consultants, for I doubt that anyone would question Dr. Seligman's stature among clinical scientists. But, where is the balance (e.g., the reference to Jacobsen's *Who or what can do psychotherapy* article, or to Dawes' *House of Cards* book)? Psychotherapy appears to have been held to a much more lenient standard than herbal treatments and, partly as a result, was presented in a much more favorable light. This is embarrassing and now that the uncritical presentation has made its way into the popular media and will certainly be capitalized upon by the sellers of psychotherapy, scientific approaches to clinical psychology seem in even greater peril than ever before and this is very troubling.

While I'm on this diatribe, let me raise a few more issues and invite responses.

An earlier post by Seligman in response to Doris Aaronson suggested that there were a number of internal controls (and related results) in the CR study of therapy to attenuate concern about the lack of a control grp. I was not as impressed by them as he. I don't think, eg, that the poorer outcomes reported for marriage counselors control adequately for either nonspecific therapist effects or the simple passage of time. These therapists are liable (as the CR article itself points out) to encounter a greater proportion of clients with problems occurring in the context of relatively intractable environments and they are faced with the daunting task of pleasing two or more [often hostile] clients simultaneously, a challenge not faced so frequently (or at least not so directly) by other therapists. Then too, resolution of the immediate problems through divorce or separation often introduces a new set of problems with finances, custody, etc.

Neither do I find the poorer outcomes reported for family doctors to be either surprising or adequate to allay concerns about a lack of a control group or for nonspecific effects. Because the incentives for family doctors favor their seeing six patients an hour rather than one, it should not be too remarkable that they get lower marks for listening, empathy, sensitivity, etc. It seems to me that CR should use these data, not to recommend MH "specialists," but rather to examine the incentive system; and in this context might promote the use of low cost counselors given the lack of differences in comparisons among soc wrkrs, psychologists, and psychiatrists. Instead, with no specifics and certainly little empirical basis, CR recommends seeking professionals with "experience and qualifications," a combination that usually translates into higher costs without demonstrably greater benefit.

I am also skeptical of the mechanisms that might underlie apparent dose-response effects, but others have already addressed this point previously. What concerns me more in this context is the suggestion made by CR that challenges to third-party payment for long-term therapy may be misplaced. This, after noting that clients report that treatment helps in three distinct ways, the first being alleviation of presenting symptoms (presumably this effect occurs earliest). I'm willing to accept that my tax dollars and insurance premiums might be well spent in eliminating suffering of this sort, even if the only verification is client report. However, I'm much less willing to pay for improved ability to function, relate, cope, and produce without objective evidence of these alleged improvements; and I am vigorously opposed to spending money on the elective "personal growth" of others under the auspices of essential mental health services.

Finally, for the reasons alluded to above, I have a lot less confidence than Dr. Seligman in the validity of clients' attributions of the causes of any alleged changes in their behavior during or after therapy. People try to make sense of their world, but are often mistaken in the sense they make of it, even if it makes them feel better.

Sorry for the length of this, but I hope there will be continued discussion of the vital issues raised by the CR article. For my part, I can't help but think that if it represents the best MH therapy has to offer in terms of a basis for confidence placed in it, it is on thin ice indeed. Nonetheless, now that the relatively uncritical report is in a mass media outlet which ordinarily deserves the credibility it is given, clinical science had better prepare for the fallout, for surely those with a vested interest in business as usual will try to use it to bury those of us who would urge caution and skepticism. Got to run now, therapy appointment you know...

Alan R. Lang, Ph.D.
Psychology Department
Florida State University
Tallahassee, FL 32306-1051

phone: (904)644-6065
fax: (904)644-0790
allang@psy.fsu.edu

From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: sscpnet@bailey.psych.nwu.edu, div12@vm1.nodak.edu Date: Mon, 30 Oct 1995 17:02:28 -0500 (EST)
--

Forwarded message:

>>From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
Subject: Re: CR article
To: JENNIFER@compass.compass-is.com (Jennifer Lish)
Date: Mon, 30 Oct 1995 12:31:09 -0500 (EST)<<

>>The 3000 people who talked to friends and clergy were not asked the follow up questions. Unfortunately. They were, as I recall, markedly less troubled at outset than the other 4000, so their use as a control group would be questionable.<<

>>Also as I dimly recall, the patient staisfaction measure did correlate pretty noticeably with the variety of measures about how therapy helped for the specific problem, in inter personal relations, in productivity at work, etc, as well as the difference score between how they were feeling at the outset of therapy and how they were feeling at the time fo the survey. Jennifer, you might want to phone, Dr. Mark Kotkin, 914 378 2253, to get these correlations.<<

>Jennifer D. Lish, Ph.D.
Research Scientist
Compass Information Services
Suite 410 -- 1060 First Avenue
King of Prussia, PA 19406
Phone: 610-992-7073
Fax: 610-992-7070

Adjunct Assistant Professor, Department of Psychiatry<
Medical College of Pennsylvania and Hahnemann University

Date: Tue, 31 Oct 1995 16:54:28 -0500 (EST) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman), "Jennifer Lish" <JENNIFER@compass.compass-is.com> Subject: Re: CR article
--

Jennifer

I disagree that friends and clergy would be a "control" group when the distressed people themselves decide from whom to seek help. Consider, for example, the possibility that seeking such people is negatively correlated with the severity of the problem (even WITHIN categories) or financial position.

I agree with Marty that people do in fact choose, but there is no reason that the logic of randomization can't be used within that constraint. Some people are randomly chosen to have the opportunity to choose one or one from a particular group of therapies courtesy the experiment while others are left to their own devices (which will always be in play anyway--just like trying to get a good night's sleep if physically ill).

But I believe it is a mistake to refer to "control" groups in a self-selection nonexperiment.

Robyn

From: "Jennifer Lish" <JENNIFER@compass.compass-is.com>
To: Robyn Dawes <rd1b+@andrew.cmu.edu>
Date: Wed, 1 Nov 1995 13:26:32 EDT
Subject: Re: CR article

>Robyn Dawes said:<

>I disagree that friends and clergy would be a "control" group when the distressed people themselves decide from whom to seek help. Consider, for example, the possibility that seeking such people is negatively correlated with the severity of the problem (even WITHIN categories) or financial position.<

I see the problem you are pointing out -- I was searching for an opportunity to determine what amount of change on the measure used by CR would be expected due to spontaneous remission over time, along the lines of the point Ken Howard made by referring people to the McNeilly and Howard paper. Perhaps one would need to equate subjects for initial severity and length of time since help-seeking, but I think the group that saw friends and clergy would be a useful comparison group to determine whether the improvement that the CR respondents attributed to psychotherapy was truly due to it.

Jennifer D. Lish, Ph.D.
Research Scientist
Compass Information Services

Date: Wed, 1 Nov 1995 13:40:39 -0600
To: "David H. Barlow" <MB399@cns.vax.albany.edu>
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: Balow's comments on Consumer Reports and Clinical Utility

>Like many readers on this net I was delighted to see the CR article, I don't believe this survey can legitimately be called "effectiveness research" or "services research".<

Although I try not to agree with him (he is almost always wrong), Sechrest makes an important point when he states that we should be concerned about the questions we ask -- the particular research question should guide our methodology.

So, e.g., if we ask "Do (specific kinds of) patients who are offered treatment A (in a particular setting) have better average outcomes than do patients who are offered treatment B (regardless of whether any patient actually receives either treatment)?" we could do a randomized clinical trial (RCT).

If we ask "How do readers of Consumers Reports who have had psychotherapy (and bother to fill out and return a questionnaire)

retrospectively rate the helpfulness of that treatment?" we could do a reader survey. And so on.

>Second, efficacy studies are routinely attacked these days as being artificial in many ways and rightly so. ... but ... efficacy studies should be done first so that we would have something to study when we considered clinical utility<

There is no necessary relationship between the results of an efficacy study and the results of an effectiveness study. Based on an RCT, a treatment could be shown to yield a higher average outcome than a control condition and not help patients at all in clinical settings. Also a treatment could yield an average result in an RCT that is not larger than the average result of the control group and help lots of patients in clinical settings.

>However, many of the potshots continually taken at efficacy studies are based on unproven assumptions about generality and, in some cases, are simply incorrect. To take the latter first, the myth has been widely propagated that comorbidity is not present in efficacy studies. To take the 3 or 4 ongoing multi-site collaborative studies in anxiety disorders as an example, none of these studies require single disorder presentations (which are very rare indeed). All patients in these studies present with extensive comorbidity, including all Axis II disorders.<

I have reviewed hundreds of RCT proposals, and have yet to see one without exclusion criteria. Moreover, these criteria are almost never rationalized. Also, as a colleague pointed out, there is usually no information given about how these criteria will be assessed or about the psychometric properties of these assessment. So, for example, an exclusion criterion might be "suicidal ideation" where we are not told how or how well this is measured.

If, as Barlow does, you have a treatment that has very strong efficacy, you can afford more within cell heterogeneity (among patients). Exclusion criteria are meant to reduce this heterogeneity and give us the best chance of finding a statistically significant effect and a decent effect size.

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Thu, 2 Nov 1995 13:21:19 -0600
To: sscpnet@bailey.psych.nwu.edu
From: k-howard@nwu.edu (Ken Howard)
Subject: Psychotherapy outcome

The outcome problem is a very thorny one. We find no useful relationship between outcome measured by therapist change scores (or ratings) and patient change scores (or ratings) or even between pre-post change scores and termination improvement ratings by the same person.

Sure outcome assessments are significantly correlated, but hardly at a level where they could be viewed as interchangeable or even combinable. So, we have controversy over the Consumer Report retrospective patient evaluations.

Just for fun I ran an analysis comparing reliable improvement (regression-corrected pre-post change scores using patient self-report scales -- Subjective Wellbeing, Symptoms, and Functioning combined into what we call a Mental Health Index [MHI]) with phone follow-up (about three to six months after termination) patient ratings of benefit.

MHI pre-post change was coded as not success (no reliable improvement) vs. success (reliable improvement). And the results were:

Degree of Benefit from Therapy
A great deal of benefit: 0-9 (0 not success, 9 success)
Quite a lot: 8-8
Moderate amount: 2-6
A Little: 5-5
No Benefit: 0-0
Negative Benefit: 1-0

Obviously not much of a relationship between retrospective report and "objective" outcome.

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Wed, 8 Nov 1995 13:16:23 -0600
To: Neil Jacobson <njacob@u.washington.edu>
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: Jacobson's "Psychotherapy outcome"

>Are you saying that there is no connection AT ALL between consumer satisfaction and psychiatric outcome?<

There is hardly any relationship between outcome assessments from different perspectives (therapist; patient) different methods (retrospective; change scores; global ratings) and different measures -- correlations range from zero to low .30's). We may have to resort to raw operationism and use the full "outcome as measured by...."

>If so, what are the implications of Seligman and CR creating a composite?<
>>That is what they did, right? As far as I can tell, they simply lumped consumer satisfaction, psychiatric outcome, and overall functioning together, and the figure in the CR article reports on just that one measure.<<
>Doesn't that mean that, for example, the consumer satisfaction measure could be carrying most of the weight.<

All of these measures were self-report and obtained from the same questionnaire. I would be shocked if they didn't correlate in the high .80's (at least) with one another.

The .04 (reported by Jennifer Lish) correlation between consumer satisfaction and change scores (based on patient reports) is based on a very recent analysis. You might contact her for details
(JENNIFER@COMPASS.COMPASS-IS.COM)

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Wed, 8 Nov 1995 14:44:29 -0600
To: sscpnet@bailey.psych.nwu.edu
From: k-howard@nwu.edu (Ken Howard)
Subject: significance testing

There is a lively debate going on regarding the use of significance tests. I thought you all might be interested in some highlights. S0...

>From: SMEP Network<

>Subject: Objectivity & significance testing<

>To adopt a procedure because it has the virtue of being objective when it has typical error rates of 50%, 60%, 70%, etc. is to make a irrational fetish of objectivity. This is especially so given that a procedure is available that will hold the total error rate to 5%, 1%, or whatever we want it to be: confidence intervals.<

>Flipping a coin to decide whether to accept or reject hypotheses would also be objective. Furthermore, from a statistical power point of view, it would guarantee that the average error rate would not go above 50%--a guarantee the significance test cannot provide. And, of course, it would be more efficient--no study would have to be done. So from the perspective of objectivity as the overriding value, flipping a coin is a better choice than the significance test.<

>What about Stan's implicit assumption that significance testing is objective but that confidence intervals are not? In fact, confidence intervals are just as objective as significance tests. There is nothing less objective about the computation of confidence interval or about the properties of the final confidence interval. What Stan is contending is that the significance test provides a more objective procedure for making a dichotomous decision: whether to accept or reject the hypothesis. He feels that significance tests, whatever their lethal faults, provide a rigid set of rules for making this binary decision in individual studies, whereas confidence intervals do not. Confidence intervals merely provide a point estimate and an error band around that estimate.<

>The problem with this position is that there is no scientific reason to have or want binary decisions to accept or reject hypotheses in individual studies--because the questions of whether to accept or reject an hypothesis does not depend on the results of one study. Instead, decisions to accept or reject are based on results integrated across all studies that have tested that hypothesis. The usual way of doing this is through meta-analysis. (There may be may have tested the hypothesis in different ways or tested different aspects of the hypothesis.)<

>Stan is asking the impossible: he wants a procedure that will provide the answer to whether the hypothesis is correct or incorrect FROM A SINGLE STUDY. We know from statistical power analyses that the significance test cannot do this: the fact that power in typical research is low (e.g., .30 to .50) shows that the individual study does not CONTAIN enough information to allow a conclusion about the truth or falseness of the hypothesis tested. The confidence interval reveals this lack of information more clearly and directly: the confidence intervals are quite wide. The significance test CONCEALS from the researcher the lack of information in the study; only those few researchers who go beyond

computing the significance test itself and compute the statistical power of their study ever actually see how little information there is in their study. This concealment is very convenient for users of significance tests: it allows them to feel good about making accept-reject decisions about hypotheses even when the error rate is 50% or more.<

>Yes, but aren't there at least SOME studies with either large enough Ns, large enough effect sizes, or both to produce very high statistical power? If so, can't such a single study answer a scientific question (i.e., provide a sound basis for accepting or rejecting an hypothesis)? And in the case of such studies, why not use significance tests--because, after all, power is not problem in such a study.<

>Taking the last question first: low statistical power in typical research studies is only ONE of the severe problems associated with significance testing. There are many other reasons for not using significance testing (see Cohen's paper, my paper, Carver's paper, Rozeboom's paper, etc.).<

>Second, there are reasons beyond inadequate power why a single study cannot settle a scientific question. That is, there are additional reasons why replication is necessary in science. There may be, for example, measurement differences between studies (e.g., the dependent variable may be measured differently), treatment operationalization differences, design differences, instructional differences, and hundreds of other differences. Once a body of studies has accumulated, meta-analysis can be used to determine whether any of these differences makes any difference in research outcomes. For example, the mean effect sizes can be compared for the cluster of studies that used operationalization A and the cluster using operationalization B. In this way both methodological and substantive moderators of results can be identified (or shown not to exist). Hence, even if statistical power is very high, a single study still cannot answer a scientific question. A single study cannot examine between-study variability in results--and such variability may hold important information.<

>I am aware that many primary researchers do not like these facts. They strongly want to believe that THEIR individual primary study CAN answer an important scientific question. But wanting something to be true does not make it true.<

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Wed, 8 Nov 1995 14:55:23 -0600
To: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
From: k-howard@nwu.edu (Ken Howard)
Subject: Re: Jacobson's "Psychotherapy outcome"

>If there were a sufficient sample size (as in the CR study) and thus enough statistical power, it should be possible to do some kind of confirmatory factor analysis or structural equations analysis to come up with separate estimates of method (source, e.g. therapist, patient) and trait (outcome) variance. What we are after is an estimate of a conceptual (latent) variable. To resort to operationism would be a regressive move, necessary only in studies without multi-trait, multi-method data or without sufficient power. Right?<

>Don<

In the CR study, they only had patient self-reports.

There are at least six evaluative perspectives: patients (who get the treatment), clients (who request the treatment), therapists (who give the treatment), managers (who decide on the allocation of treatment resources), sponsors (who pay for the treatment), and researchers who evaluate the treatment). Is there any reason to suspect latent agreement (or agreement on a latent variable)?

Within the patient perspective, it turns out that patients' retrospective (global) evaluations of improvement are highly correlated with their status at termination. Their status at termination is highly correlated with their initial status (so the healthier you are at intake, the healthier you will be at termination). Consequently there is rarely a high positive correlation between pre-post change and global improvement ratings. Could this be handled by latent structural analysis? I think not.

Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Wed, 08 Nov 1995 14:56:21 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Jacobson's "Psychotherapy outcome"
To: k-howard@nwu.edu (Ken Howard)

>>Are you saying that there is no connection AT ALL between consumer satisfaction and psychiatric outcome?<<

>There is hardly any relationship between outcome assessments from different perspectives (therapist; patient) different methods (retrospective; change scores; global ratings) and different measures -- correlations range from zero to low .30's). We may have to resort to raw operationism and use the full "outcome as measured by...."<

>>If so, what are the implications of Seligman and CR creating a composite?<<

>>>That is what they did, right? As far as I can tell, they simply lumped consumer satisfaction, psychiatric outcome, and overall functioning together, and the figure in the CR article reports on just that one measure. Doesn't that mean that, for example, the consumer satisfaction measure >>could be carrying most of the weight.<<

>All of these measures were self-report and obtained from the same questionnaire. I would be shocked if they didn't correlate in the high .80's (at least) with one another.<

>The .04 (reported by Jennifer Lish) correlation between consumer satisfaction and change scores (based on patient reports) is based on a very recent analysis. You might contact her for details (JENNIFER@COMPASS.COMPASS-IS.COM) <

>Ken Howard <

If there were a sufficient sample size (as in the CR study) and thus enough statistical power, it should be possible to do some kind of confirmatory factor analysis or structural equations analysis to come up with separate estimates of method (source, e.g. therapist, patient) and trait (outcome) variance. What we are after is an estimate of a conceptual (latent) variable. To resort to operationism would be a regressive move, necessary only in studies without multi-trait, multi-method data or without sufficient power. Right?

Don

Date: Wed, 08 Nov 1995 16:24:24 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Jacobson's "Psychotherapy outcome"
To: k-howard@nwu.edu (Ken Howard)

>>If there were a sufficient sample size (as in the CR study) and thus enough statistical power, it should be possible to do some kind of confirmatory factor analysis or structural equations analysis to come up with separate estimates of method

(source, e.g. therapist, patient) and trait (outcome) variance. What we are after is an estimate of a conceptual (latent) variable. To resort to operationism would be a regressive move, necessary only in studies without multi-trait, multi-method data or without sufficient power. Right?<<

>>Don<<

>In the CR study, they only had patient self-reports.<

>There are at least six evaluative perspectives: patients (who get the treatment), clients (who request the treatment), therapists (who give the treatment), managers (who decide on the allocation of treatment resources), sponsors (who pay for the treatment), and researchers who evaluate the treatment). Is there any reason to suspect latent agreement (or agreement on a latent variable)?<

>Within the patient perspective, it turns out that patients' retrospective (global) evaluations of improvement are highly correlated with their status at termination. Their status at termination is highly correlated with their initial status (so the healthier you are at intake, the healthier you will be at termination). Consequently there is rarely a high positive correlation between pre-post change and global improvement ratings. Could this be handled by latent structural analysis? I think not.<

>Ken Howard
Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859<

Dear Ken:

The articles I was recently reading on the subject of confirmatory factor analysis of multitrait-multimethod data were by David Cole at Notre Dame (phone 219-239-6165). They might not apply specifically to the CR study, but I still think there is a lot to be said for using a latent variable approach instead of old fashioned operationism. Cole, being a statistician (which I admittedly am not) could not doubt speak more eloquently on this topic.

Don

From: "Paul" <CRITS@landru.cpr.upenn.edu>
To: kaechele@sip.medizin.uni-ulm.de (Prof. Dr. Horst Kaechele)
Date: Wed, 8 Nov 1995 16:45:42 EST
Subject: Re: APA validated treatments

Hi Horst -

There is much confusion about the list of empirically validated treatments. In fact, the original list was created by a task force of which I was a member. The list is published as part of a task force report, but APA as a whole has not endorsed this report. Some sections of APA have endorsed some of the specific recommendations of the task force, which mostly had to do with exposing graduate students more to treatments that have an empirical basis. The actual list of treatments in the report is not definitive --in fact, the list is described as "examples of empirically validated treatments," rather than a definitive list. No dynamic therapies made it into the "well established" category, although brief dynamic therapy made it into the "probably efficacious" category. The reason for this is that there was a criteria of at least two controlled (not wait-list) efficacy studies (by different investigators) to make it into the well-established category, and no dynamic treatments qualify. Also, the list examines treatments for specific problems or disorders, and many of the dynamic therapy studies in the literature did not target a specific problem or disorder. The task force is currently revising the list to make it more complete. If you have any other questions about this, I would be glad to answer them.
Paul

Subject: Submissions to PSYCOLOQUY
To: sscpnet@bailey.psych.nwu.edu
From Jeff Berman, Memphis State University

Given the lively discussion that has occurred on SSCPNET for the past few months, it seems that many of the issues regarding clinical psychology and psychotherapy would be of interest to a wider electronic audience. As one of the associate editors of PSYCOLOQUY, the electronic journal sponsored by APA, I would therefore like to encourage members of SSCPNET to consider more formal submissions of their ideas to this forum. For your information, I have attached below a copy of the instructions for submitting articles or commentaries to the journal. Of course, I realize that one risk of making this call for submissions is that Lee Sechrest will now send all his serious work to the journal, leaving us with just his jokes.

INSTRUCTIONS FOR PSYCOLOQUY AUTHORS AND COMMENTATORS

PSYCOLOQUY is a refereed electronic journal (ISSN 1055-0143) sponsored on an experimental basis by the American Psychological Association and currently estimated to reach a readership of 36,000. PSYCOLOQUY publishes brief reports of new ideas and findings on which the author wishes to solicit rapid peer feedback, international and interdisciplinary ("Scholarly Skywriting"), in all areas of psychology and its related fields (biobehavioral, cognitive, neural, social, etc.) All contributions are refereed by members of PSYCOLOQUY's Editorial Board. Target article length should normally not exceed 500 lines [c. 4500 words]. Commentaries and responses should not exceed 200 lines [c. 1800 words]. All target articles, commentaries and responses must have (1) a short abstract (up to 100 words for target articles, shorter for commentaries and responses), (2) an indexable title, (3) the authors' full name(s) and institutional address(es). In addition, for target articles only: (4) 6-8 indexable keywords, (5) a separate statement of the authors' rationale for soliciting commentary (e.g., why would commentary be useful and of interest to the field? what kind of commentary do you expect to elicit?) and (6) a list of potential commentators (with their email addresses). All paragraphs should be numbered in articles, commentaries and responses (see format of already published articles in the PSYCOLOQUY archive; line length should be < 80 characters, no hyphenation). It is strongly recommended that all figures be designed so as to be screen-readable ascii. If this is not possible, the provisional solution is the less desirable hybrid one of submitting them as postscript files (or in some other universally available format) to be printed out locally by readers to supplement the screen-readable text of the article. PSYCOLOQUY also publishes multiple reviews of books in any of the above fields; these should normally be the same length as commentaries, but longer reviews will be considered as well. Book authors should submit a 500-line self-contained *Precis* of their book, in the format of a target article; if accepted, this will be published in PSYCOLOQUY together with a formal Call for Reviews (of the book, not the *Precis*). The author's publisher must agree in advance to furnish review copies to the reviewers selected. Authors of accepted manuscripts assign to PSYCOLOQUY the right to publish and distribute their text electronically and to archive and make it permanently retrievable electronically, but they retain the copyright, and after it has appeared in PSYCOLOQUY authors may republish their text in any way they wish -- electronic or print -- as long as they clearly acknowledge PSYCOLOQUY as its original locus of publication. However, except in very special cases, agreed upon in advance, contributions that have already been published or are being considered for publication elsewhere are not eligible to be considered for publication in PSYCOLOQUY. Please submit all material to psyc@pucc.bitnet or psyc@pucc.princeton.edu Anonymous ftp archive is [pub/harnad/Psycology](ftp://pub/harnad/Psycology) HOST princeton.edu -----

CRITERIA FOR ACCEPTANCE:

To be eligible for publication, a PSYCOLOQUY target article should not only have sufficient conceptual rigor, empirical grounding, and clarity of style, but should also offer a clear rationale for soliciting Commentary. That rationale should be provided in the author's covering letter, together with a list of suggested commentators. A target article can be (i) the report and discussion of empirical research; (ii) an theoretical article that formally models or systematizes a body of research; or (iii) a novel interpretation, synthesis, or critique of existing experimental or theoretical work. Articles dealing with social or philosophical aspects of the behavioral and brain sciences are also eligible.. The service of Open Peer Commentary will be primarily devoted to original unpublished manuscripts. However, a recently published book whose contents meet the standards outlined above may also be eligible for Commentary. In such a Multiple Book Review, a comprehensive, 500-line precis by the author is published in advance of the commentaries and the author's response. In rare special cases, Commentary will also be extended to a position paper or an already published article dealing with particularly influential or controversial research. Submission of an article implies that it has not been published or is not being considered for publication elsewhere. Multiple book reviews and previously published articles appear by invitation only. The Associateship and professional readership of PSYCOLOQUY are encouraged to nominate current topics and authors for Commentary. In all the categories described, the decisive consideration for eligibility will be the desirability of Commentary for the submitted material. Controversially simpliciter is not a sufficient criterion for soliciting Commentary: a paper may be controversial simply because it is wrong or weak. Nor is the mere presence of interdisciplinary aspects sufficient: general cybernetic and "organismic" disquisitions are not appropriate for PSYCOLOQUY. Some appropriate rationales for seeking Open Peer Commentary would be that: (1) the material bears in a significant way on some current controversial issues in behavioral and brain sciences; (2) its findings substantively contradict some well-established aspects of current research and theory; (3) it criticizes the findings, practices, or principles of an accepted or influential line of work; (4) it unifies a substantial amount of disparate research; (5) it has important cross-disciplinary ramifications; (6) it introduces an innovative methodology or formalism for consideration by proponents of the established forms; (7) it meaningfully integrates a body of brain and behavioral data; (8) it places a hitherto dissociated area of research into an evolutionary or ecological perspective; etc. In order to assure communication with potential commentators (and readers) from other PSYCOLOQUY specialty areas, all technical terminology must be clearly defined or simplified, and specialized concepts must be fully described.

NOTE TO COMMENTATORS: The purpose of the Open Peer Commentary service is to provide a concentrated constructive interaction between author and commentators on a topic judged to be of broad significance to the biobehavioral science community. Commentators should provide substantive criticism, interpretation, and elaboration as well as any pertinent

complementary or supplementary material, such as illustrations; all original data will be refereed in order to assure the archival validity of PSYCOLOQUY commentaries. Commentaries and articles should be free of hyperbole and remarks ad hominem.

STYLE AND FORMAT FOR ARTICLES AND COMMENTARIES

TARGET ARTICLES: should not exceed 500 lines (~4500 words); commentaries should not exceed 200 lines (1800 words), including references. Spelling, capitalization, and punctuation should be consistent within each article and commentary and should follow the style recommended in the latest edition of *A Manual of Style*, The University of Chicago Press. It may be helpful to examine a recent issue of PSYCOLOQUY. All submissions must include an indexable title, followed by the authors' names in the form preferred for publication, full institutional addresses and electronic mail addresses, a 100-word abstract, and 6-12 keywords. Tables and diagrams should be made screen-readable wherever possible (if unavoidable, printable postscript files may contain the graphics separately). All paragraphs should be numbered, consecutively. No line should exceed 72 characters, and a blank line should separate paragraphs.

REFERENCES: Bibliographic citations in the text must include the author's last name and the date of publication and may include page references. Complete bibliographic information for each citation should be included in the list of references. Examples of correct style are: Brown(1973); (Brown 1973); Brown 1973; 1978); (Brown 1973; Jones 1976); (Brown & Jones 1978); (Brown et al. 1978). References should be typed on a separate sheet in alphabetical order in the style of the following examples. Do not abbreviate journal titles.

- Kupfermann, I. & Weiss, K. (1978) The command neuron concept.
Behavioral and Brain Sciences 1:3-39.
- Dunn, J. (1976) How far do early differences in mother-child relations affect later developments? In: Growing point in ethology, ed. P. P. G. Bateson & R. A. Hinde, Cambridge University Press.
- Bateson, P. P. G. & Hinde, R. A., eds. (1978) Growing points in ethology, Cambridge University Press.

EDITING: PSYCOLOQUY reserves the right to edit and proof all articles and commentaries accepted for publication. Authors of articles will be given the opportunity to review the copy-edited draft. Commentators will be asked to review copy-editing only when changes have been substantial.

From plaud@badlands.NoDak.edu

Sun Jan 29 20:24:38 1995

Subject: Electronic Journal and WWW (fwd)
To: SSCPnet <sscpnet@bailey.psych.nwu.edu>

Subject: Electronic Journal and WWW

To all list members, we are busily moving forward on our electronic journal programming. For more information, if you have a WWW graphical browser such as Mosaic or NetScape, or the myriad of others, point it to:

<http://rs1.cc.und.nodak.edu/misc/jBAT/joehome.html>

If you are interested in being considered for the editorial board of the journal, please email me with your credentials. The sections of the journal are not yet finalized, and won't be for a while. However, we are attempting to put as much structure as possible to the project. Here is a brief idea of some of the ideas our core group has been discussing:

(1) Review of Materials in other Journals Related to Behavior Analysis and Therapy. One of the primary proposed functions of jBAT will be its attempt at providing a current survey of basic and applied research in the field as represented in other journals. This includes JEAB-type articles, theoretical/philosophical articles, and applied research. For example, in one the latest edition of the Psychological Review, Jack McDowell has a wonderful co-authored article on a linear systems analysis of reinforcement. Review of relevant analyses would be important to keep all readers up to date on important goings on in behavior analysis.

(2) A Coherent Organizational Structure. The way I have conceptualized jBAT, it would be a journal with specific sub-sections; including (but certainly not restricted to) behavioral assessment, theoretical and philosophical issues, EAB, with appropriate sub-divisions which could include operant and respondent conditioning (basic), rule-governed behavior, behavioral allocation and choice, verbal behavior; child bt, adult bt, behavioral med, case studies, abstracts as discussed above, with a special Feature Article inc), rule-governed behavior, behavioral allocation and choice, verbal behavior; child bt, adult bt, behavioral med, case studies, abstracts as discussed above, with a special Feature Article in each issue that highlights a particularly interesting article that attempts to bridge the gap between behavior analysis and therapy. The main concerns of the journal will be BOTH basic experimental and clinical issues, and related behavioral work to this theme would also be acceptable. The overall mission would be a journal dedicated to the design and application of behavioral principles (even "cognitive" principles

appropriately defined and manipulated with behavioral references) to both behavior analysis and clinical psychology/behavior therapy.

(3) Editorial Structure. In accordance with its proposed organizational structure, jBAT will have an editor, and several associate editors, each responsible for a particular content area, such as the abstracts, child bt, etc. Also, the editorial board will be composed of both ABA and AABa particular content area, such as the abstracts, child bt, etc. Also, the editorial board will be composed of both ABA and AABT-like members, representing the best and most behavioral of both the basic and applied behavioral spheres.

The journal will have FULL graphical capabilities on the WWW, but we will also maintain an ascii version for downloading. Also, it will be a fully peer-reviewed journal of behavior analysis and therapy. I do hope that qualified intersted individuals will consider a possible editorial board appointment. At this point Professor Michael Davison is in charge of the experimental/operant etc. sections of the journal, and we are putting together the clinical sections as well. I am also interested in receiving more formal support from ABA, SEAB, Div25 or other behaviorally oriented groups. Please write if you have any questions.

Joe Plaud

+-----
=-=+
|Joseph J. Plaud, Ph.D. University of North Dakota P.O. Box 8380 |
|Department of Psychology Grand Forks, ND 58202-8380 (701) 777-4494 |
|E-mail 1: Plaud@Badlands.NoDak.Edu E-Mail 2: Plaud@Plains.NoDak.Edu
|
WWW: http://rs1.cc.und.nodak.edu/misc/jBAT/joehome.html
=-
List-Owner: Behavior Analysis Net, Div12 Net of the APA, InterPsych
Clinical-Psychologists Forum & Managing Editor, InterPsych Newsletter
+-----
=-=

kaechele @sip.medizin.uni-ulm.de

letters
from
sscpnet
7/95

Rorschach

Abteilung Psychotherapie
Klinikum der Universität Ulm

Date: Wed, 25 Oct 1995 13:56:02 -0500 To: sscpnet@bailey.psych.nwu.edu From: "Nancy D. Vogeltanz" <vogeltan@badlands.nodak.edu> Subject: Re: The guilty Rorschach
--

At 12:56 PM 10/25/95 -0400, Brad Smith wrote:

>I respectfully disagree with Donald Routh's recent assertion that the Rorschach is an "innocent" test. To us an analogy, the Rorschach is to psychological testing as wearing tight blue jeans is to swimming. Neither the Rorschach or wearing jeans when swimming impairs permormance all that much in minimally demanding situations. However, it wastes energy and looks silly.<

>We should be teaching graduate students to use their resources efficiently. This includes teaching about the costs and benefits of various tests, not only in terms of theory and psychometric properties, but also in terms of time and energy. Assuming the R takes at least three hours to administer and score and that most clinics charge around \$120 an hour, is the R worth giving up three hours of direct patient contact or depriving the health care system of \$360? I think not.<

>Brad Smith
ADD Program
WPIC<

As is almost always the case, I strongly agree with Brad Smith's comments regarding the dreaded "R." I seriously question the utility of teaching graduate students how to administer an instrument with poor psychometric propertries AND poor cost-effectiveness. That is not to say that I believe cost-effectiveness should be the bottom line for any assessment or treatment procedure (a well-designed behavioral observation system or token economy can be very time consuming and expensive), but I do believe that time consuming/expensive procedures should have proven effectiveness. Is there anyone out there that can covincingly argue that the Rorschach has a place in clinical assessment other than for the historical significance?

Nancy D. Vogeltanz, Ph.D.
University of North Dakota P.O. Box 8380
Department of Psychology Grand Forks, ND 58202-8380 (701) 777-3790
E-mail: Vogeltan@Badlands.NoDak.Edu

Date: Wed, 25 Oct 1995 14:34:37 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Guilty Rorschach
To: sscpnet@bailey.psych.nwu.edu

Dear Brad:

I agree with what you say as far as it goes. Indeed, I neither use the Rorschach clinically nor do I encourage students to use in in their clinical work.

But what about the use of inkblot tests more generally in research contexts? I gave one example of this in my previous comment (formal inkblot test-based measures of psychotic thinking, e.g. by Philip Holtzman, et al.). Another well justified use of such procedures is in the study of the development of form perception. There are few better demonstrations of Heinz Werner's developmental sequence from global to differentiated to hierarchically integrated responses.

Don

From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Date: Wed, 25 Oct 1995 14:37:02 -0500
Subject: Re: Rorschach, etc. again
To: sscpnet@bailey.psych.nwu.edu

Dear Bruce:

Although I have not read the particular study you mention, that's exactly the kind of thing I had in mind in my comment.

Don

Date: Wed, 25 Oct 1995 14:38:32 -0400
To: sscpnet@bailey.psych.nwu.edu
From: bsmith+@pitt.edu (Brad Smith)
Subject: The guilty Rorschach

I respectfully disagree with Donald Routh's recent assertion that the Rorschach is an "innocent" test. To use an analogy, the Rorschach is to psychological testing as wearing tight blue jeans is to swimming. Neither silly practice impairs permormance all thatmuch in minimally demanding situations. However, both waste energy and look silly.

We should be teaching graduate students to use their resources efficiently. This includes teaching about the costs and benefits of various tests, not only in terms of theory and psychometric properties, but also in terms of time and energy. Assuming the R takes at least three hours to administer and score and that most clinics charge around \$120 an hour, is the R worth giving up three hours of direct patient contact or depriving the health care system of \$360? I think not.

Brad Smith
ADD Program
WPIC

Date: Wed, 25 Oct 1995 15:20:23 -0400 (EDT) From: Marvin Zuckerman <zuckerma@UDel.Edu> To: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Rorschach

Years back I did some research with Gene Levitt and Berny Lubin in which we assessed anxiety, depression, hostility, and dependency with peer ratings, self-ratings, questionnaires, TAT and Rorschach. Rorschach responses were scored by standardized content systems. Both TAT and Rorschach bore no relationship at all to peer ratings or direct self-report measures, or even to each other. Most of what people attempt to assess by this unreliable and invalid method could more effectively be assessed with direct self-ratings or questionnaires.

Marvin Zuckerman

Date: Wed, 25 Oct 1995 15:33:36 -0400 (EDT) From: Marvin Zuckerman <zuckerma@UDel.Edu> To: John F Kihlstrom <john.kihlstrom@yale.edu> Subject: Re: Rorschach

Just read the reviews of the Rorschach and TAT in Buros Mental Measurements Yearbooks over the years and ask yourself "how can a profession which claims to respect the outcome of research above the seat of the pants convictions of clinicians (illusory correlations) continue to use or teach these techniques, particularly since superior and more objective methods have proven of some validity.

Marvin Zuckerman

Date: Wed, 25 Oct 1995 15:35:49 -0400 (EDT)
From: Marvin Zuckerman <zuckerma@UDeI.Edu>
To: Bruce Compas <B_COMPAS@dewey.uvm.edu>
Subject: Re: Rorschach, etc.

Has anyone replicated these findings? One must realize that the Rorschach has thousands or maybe tens of thousands of unpublished negative results and occasionally a positive finding will emerge.

Marvin Zuckerman

On Wed, 25 Oct 1995, Bruce Compas wrote:

Don,
>There is actually an interesting series of studies using TAT responses to predict measures of immune function. I am highly skeptical of all projective methods (and I don't teach them to our students, either), but these studies are very interesting. John Jemmott conducted the research and some of it appeared in the Journal of Behavioral Medicine in the late 1980s. I think it showed that the use of a selected subset of the cards, scored in a reliable and theoretically-based manner, can provide a sample of behavior that is meaningful.<

Bruce Compas
University of Vermont

Date: Wed, 25 Oct 1995 20:39:05 -0700 (PDT)
From: Gerald Davison <gdaviso@alnitak.usc.edu>
To: John F Kihlstrom <john.kihlstrom@yale.edu>
Subject: Re: Rorschach

On Wed, 25 Oct 1995, John F Kihlstrom wrote [on why he is OK with having taught the Rorschach with McClelland some years back at Harvard]

[stuff deleted, my comments in caps for clarity]

>Second, even though Harvard didn't have a clinical training program, it was clear that some of the students were headed for clinical practice, and others were headed for faculty positions in clinical programs, and we felt that they should know something about the Rorschach, just as they should know something about all the other tests that, for better or worse, are commonly employed in clinical settings. But, like Julie Wishner (who taught the Beck system in 4 hours), we taught the Rorschach as a test which had potential, if not promise, but one which had yet to be put on a firm psychometric footing.<

>In other words, we tried to give our students the skill they would need to survive in the environment of clinical practice, at the same time as we tried to convey a sense of healthy, scientific skepticism about the technique itself. We think Julie would have been proud.<

JOHN, NOT CLEAR WHY YOU WOULD NOT ALSO TEACH OTHER NON-VALIDATED ASSESSMENT AND INTERVENTION TECHNIQUES SOLELY BECAUSE THEY ARE IN USE IN APPLIED SETTINGS. I KNOW YOU AGREE WITH ME AND OTHERS THAT ACADEMIC S-P PROGRAMS SHOULD BE LEADING THE WAY, NOT HAVING THEIR CURRICULA DICTATED BY INTERNSHIP AND OTHER PLACEMENT AGENCIES. WHY, THEN, THE RORSCHACH FOR A SEMESTER?

>The third reason, of course, was that since we were teaching the TAT, there was no principled way in which we could omit the Rorschach, as another example of a projective technique.<

THOUGH I AM NOT A FAN OF THE TAT EITHER, I DO RECALL SOME PRETTY AMAZING PREDICTIVE FINDINGS OF MCCLELLAND ET AL, E.G., nACH CODED FROM GRECIAN URNS PREDICTING LATER ECONOMIC ACTIVITY. I SPENT A SEMESTER 100 YEARS AGO IN A JUNIOR SEMINAR WITH MCCLELLAND AND WAS ALWAYS IMPRESSED WITH HIS CLEAR THINKING AND UNCOMPROMISING EMPIRICISM -- PLUS THOSE AND OTHER SIMILAR FINDINGS. SO, THAT IS HOW I WOULD FIND A PRINCIPLED WAY FOR YOU TO TEACH THE TAT, AT LEAST FOR THE SPECIFIC PURPOSES THAT ENJOY SOME DEGREE OF EMPIRICAL VALIDATION. (HAVING SAID THIS, I WOULDN'T TEACH IT IN THE CONTEXT OF CLINICAL ASSESSMENT. BUT MAYBE SOMEONE KNOWS OF SOME DATA THAT WOULD PERSUADE ME OTHERWISE.)

JERRY DAVISON

Date: Thu, 26 Oct 1995 06:52:55 -0400 (EDT)
From: John F Kihlstrom <john.kihlstrom@yale.edu>
To: Gerald Davison <gdaviso@alnitak.usc.edu>
Subject: Re: Rorschach

We didn't teach the Rorschach for a semester. We taught it for 4 hours, just like Julie Wishner did, and it was plenty enough. Frankly, when internships make decisions based on how many Rorschachs applicants have administered, I think we were doing students a disservice not to teach them something about these techniques.

But, Jerry, remember what I said we taught: we taught them to understand what the test was supposed to do, and we also taught them to be skeptical about what it actually did. What more do you want from a science-practice program?

John F. Kihlstrom, Professor
Department of Psychology,
Yale University
P.O. Box 208205,
New Haven, Connecticut 06520-8205
Telephone (203) 432-2596
Facsimile (203) 432-7172

Date: Thu, 26 Oct 1995 07:52:50 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Guilty Rorschach
To: David Rosenhan <rosenhan@psych.stanford.edu>

>Query: Was it Phil Holtzman or Wayne Holtzman with whom you associate the studies of psychotic thinking? I had thought it was Wayne, of the Holtzman Inkblots.<

>David Rosenhan<

Response: Actually, both Holtzmans have contributed to this research literature in significant ways, Wayne Holtzman by his work on the Holtzman inkblot test and Phil Holtzman by research on thought disorder in schizophrenia using mostly Rorschach based measures (for example, I believe he was co-author of such an article in Psychological Assessment not long ago).

Don

Date: Thu, 26 Oct 1995 07:57:29 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Rorschach, etc.
To: "Joseph J. Plaud" <plaud@badlands.NoDak.edu>

>Don, thanks so much for the information. I must say that I was not aware of much of the TAT data you shared. I will investigate further. Thanks again.

>Joe<
> Joseph J. Plaud, Ph.D.
University of North Dakota
P.O. Box 8380
Department of Psychology
Grand Forks, ND 58202-8380
(701) 777-4494
E-mail 1: Plaud@Badlands.NoDak.Edu
E-Mail 2: Plaud@E-Plains.NoDak.Edu
WWW: <http://rs1.cc.und.nodak.edu/misc/jBAT/><

> List-Owner:
Clinical Psychology Net of the American Psychological Association (Div12),
International Clinical Psychologists Net (Clinical-Psychologists or CliniPsy),
Behavior Analysis Net (Behav-An), Psychology-Developmental Disabilities
Net (Psych-DD), & Editor, Journal of Behavior Analysis and Therapy
(jBAT)<

Dear Joe:

Thanks for your interest. Let me know your reactions to the material in the book edited by Smith, if you manage to get hold of a copy.

Don

Date: Thu, 26 Oct 1995 08:02:03 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Rorschach
To: Gerald Davison <gdaviso@alnitak.usc.edu>

On Wed, 25 Oct 1995, John F Kihlstrom wrote [on why he is OK with having taught the Rorschach with McClelland some years back at Harvard]

> [stuff deleted, my comments in caps for clarity]<

>>Second, even though Harvard didn't have a clinical training program, it was clear that some of the students were headed for clinical practice, and others were headed for faculty positions in clinical programs, and we felt that they should know something about the Rorschach, just as they should know something about all the

other tests that, for better or worse, are commonly employed in clinical settings. But, like Julie Wishner (who taught the Beck system in 4 hours), we taught the Rorschach as a test which had potential, if not promise, but one which had yet to be put on a firm psychometric footing.<<

>>In other words, we tried to give our students the skill they would need to survive in the environment of clinical practice, at the same time as we tried to convey a sense of healthy, scientific skepticism about the technique itself. We think Julie would have been proud.<<

>JOHN, NOT CLEAR WHY YOU WOULD NOT ALSO TEACH OTHER NON-VALIDATED ASSESSMENT AND INTERVENTION TECHNIQUES SOLELY BECAUSE THEY ARE IN USE IN APPLIED SETTINGS. I KNOW YOU AGREE WITH ME AND OTHERS THAT ACADEMIC S-P PROGRAMS SHOULD BE LEADING THE WAY, NOT HAVING THEIR CURRICULA DICTATED BY INTERNSHIP AND OTHER PLACEMENT AGENCIES. WHY, THEN, THE RORSCHACH FOR A SEMESTER?<

>>The third reason, of course, was that since we were teaching the TAT, there was no principled way in which we could omit the Rorschach, as another example of a projective technique.<<

>THOUGH I AM NOT A FAN OF THE TAT EITHER, I DO RECALL SOME PRETTY AMAZING PREDICTIVE FINDINGS OF MCCLELLAND ET AL, E.G., EACH CODED FROM GRECIAN URNS PREDICTING LATER ECONOMIC ACTIVITY. I SPENT A SEMESTER 100 YEARS AGO IN A JUNIOR SEMINAR WITH MCCLELLAND AND WAS ALWAYS IMPRESSED WITH HIS CLEAR THINKING AND UNCOMPROMISING EMPIRICISM -- PLUS THOSE AND OTHER SIMILAR FINDINGS. SO, THAT IS HOW I WOULD FIND A PRINCIPLED WAY FOR YOU TO TEACH THE TAT, AT LEAST FOR THE SPECIFIC PURPOSES THAT ENJOY SOME DEGREE OF EMPIRICAL VALIDATION. (HAVING SAID THIS, I WOULDN'T TEACH IT IN THE CONTEXT OF CLINICAL ASSESSMENT. BUT MAYBE SOMEONE KNOWS OF SOME DATA THAT WOULD PERSUADE ME OTHERWISE.)<

>JERRY DAVISON<

Dear Jerry:

I agree with the point I think you are making, here, i.e. that the TAT (or more broadly, thematic methods in the tradition of McClelland and others) could be taught in a principled way in a research course. However, the current research literature does not warrant their use in teaching clinical assessment.

Don

Date: Thu, 26 Oct 1995 08:41:02 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: John F Kihlstrom <john.kihlstrom@yale.edu> Subject: Re: Rorschach
--

The renewed interchange on the Rorschach is dreary, but, I suppose, necessary. Several cognitive (and affective) responses are aroused in me and cause me to comment as follows:

1. We need to keep clear the distinction between the use of the Rorschach as a convenient clinical stimulus and as a standardized psychometric instrument. I agree with Bob Carson that some clever clinicians may be able to spin wonderous webs from the Rorschach responses of clients, but then those clinicians might do just as well if patients were given almost any other stimuli (remember the "Cloud Test," the "Make-a-Picture-Story Test," the "Blacky Test"). For such uses it is the clinician, *not* the Rorschach, that must be *validated.* In clinical training, one assumes, clinicians may need to be taught to be clever (facile) with words and concepts.
2. Other evidence of the validity of any sort of test has to be judged against the standard of "incremental validity," i.e., increment in prediction over what is possible before the test is given. The fact that a test of any sort shows some, almost certainly modest, correlation with some other occurrence of interest, does not mean that the test is necessarily clinically useful. There are opportunity costs associated with the use of every test. For example, even if the Rorschach does provide a window of sorts onto disordered thought, I have been mystified by the idea that a reasonably well trained clinician might be able to detect psychopathology from the Rorschach (after an investment of two hours or so of time) that would not have been detectible by other methods in less time, e.g., a good clinical interview. Or just asking the person whether he or she might be bothered by some unusual ideas. Almost none of our commonly used measures has been tested in what amounts to a hierarchical multiple regression, but when such studies were done back in the late 50s and early 60s, they almost always showed *no* increment in prediction from projective tests, nor, as I recall, from objective tests.
3. The stake that will, when driven through the heart of the Rorschach, kill it, will be a more attractive (not, unfortunately, necessarily better) clinical instrument. John K. is right on that score, and I agree with him that our current array of objective tests will never do the job.
4. The Rorschach could, and should, be discussed in the course on History of Psychology. Otherwise, I think that students should be taught to administer it, feign competence, and treat the whole entthoroughly cynical way. I mean that.

Date: Thu, 26 Oct 1995 08:41:43 -0700 (PDT)
From: "William Thomas O'Donohue" <wto@unr.edu>
To: Marvin Zuckerman <zuckerma@UDel.Edu>
Subject: Re: Rorschach

I think Marvin Zuckerman's posting captures an important part of the quandry re: the R. I think the case that knowledge (science) grows by variation and selective retention has been at least intriguingly made by Campbell, Popper, Skinner and others. "Selective" is the problem for many clinical psychologists. It may be wrong to think that a claim such as "The R. is a psychometrically sound test." has been falsified in any strict logical sense, but looking at the marketplace of assessment methods, it seems like it is a Yugo. It appears to me that an interesting intellectual task would be to attempt to understand the intellectual criteria that are related to something or some idea being placed in the dustbin of clinical psychology. I am afraid that I would bet that 100 years from now a good proportion of clinical psychologists will be selling a lot of Yugos.

Bill O'Donohue

On Wed, 25 Oct 1995, Marvin Zuckerman wrote:

>Just read the reviews of the Rorschach and TAT in Buross Mental Measurements Yearbooks over the years and ask yourself "how can a profession which claims to respect the outcome of research above the seat of the pants convictions of clinicians (illusory correlations) continue to use or teach these techniques, particularly since superior and more objective methods have proven of some validity.<

>Marvin Zuckerman<

From: ssoldz@warren.med.harvard.edu
Date: Thu, 26 Oct 1995 08:49:52 -0400
Subject: Re[2]: Guilty Rorschach
To: rosenhan@psych.stanford.edu,
"Donald K. Routh" <drouth@umiami.ir.miami.edu>

Phil Holtzman is author of the Thought Disorder Index, from the Rorschach, which is used in a number of neuropsychological studies of schizophrenia and correlates with various brain scan phenomena. Marty Shenton and Bill Kremin have done work in this area.

Stephen

Stephen Soldz
Director of Research and Evaluation
Department of Psychiatry
Advocates for Human Potential
Harvard Medical School
Sudbury, MA
531 Beech St.
Director of Research
Roslindale, MA 02131
Boston Institute for Psychotherapy
(617) 469-3576
ssoldz@warren.med.harvard.edu

[Sorry folks, I respond in multiple roles!]

Reply Separator

Subject: Re: Guilty Rorschach Author: "Donald K. Routh" <drouth@umiami.ir.miami.edu> at HMS-Internet Date: 10/26/95 7:52 AM

>Query: Was it Phil Holtzman or Wayne Holtzman with whom you associate the studies of psychotic thinking? I had thought it was Wayne, of the Holtzman Inkblots.<

>David Rosenhan<

Response: Actually, both Holtzmans have contributed to this research literature in significant ways, Wayne Holtzman by his work on the Holtzman inkblot test and Phil Holtzman by research on thought disorder in schizophrenia using mostly Rorschach based measures (for example, I believe he was co-author of such an article in Psychological Assessment not long ago).

Don

From: ssoldz@warren.med.harvard.edu Date: Thu, 26 Oct 1995 08:56:45 -0400 Subject: Re[2]: TAT To: plaud@badlands.NoDak.edu, "Donald K. Routh" <drouth@umiami.ir.miami.edu>

In regards to the TAT, Dan McAdams has established the validity of an intimacy scoring system for it, and Drew Westen has been validating a

measure of object representation. In general, the TAT has been used, with at least modest success, in quite a number of personality studies.

Stephen
Director of Research and Evaluation
Department of Psychiatry
Advocates for Human Potential
Harvard Medical School
Sudbury, MA
531 Beech St.
Director of Research
Roslindale, MA 02131
Boston Institute for Psychotherapy
(617) 469-3576
ssoldz@warren.med.harvard.edu

[Sorry folks, I respond in multiple roles!]

Reply Separator

oben **?????** **??????** **??????** **???????**

Subject: Re: Rorschach, etc. Author: "Donald K. Routh" <drouth@umiami.ir.miami.edu> at HMS-Internet Date: 10/26/95 7:57 AM
--

>Don, thanks so much for the information. I must say that I was not aware of much of the TAT data you shared. I will investigate further. Thanks again.<

Joe

> Joseph J. Plaud, Ph.D.
University of North Dakota
P.O. Box 8380
Department of Psychology
Grand Forks, ND 58202-8380
(701) 777-4494
E-mail 1: Plaud@Badlands.NoDak.Edu
E-Mail 2: Plaud@E-Plains.NoDak.Edu
WWW: <http://rs1.cc.und.nodak.edu/misc/jBAT/><

> List-Owner:
Clinical Psychology Net of the American Psychological Association (Div12),
International Clinical Psychologists Net
(Clinical-Psychologists or CliniPsy), Behavior Analysis Net (Behav-An),
Psychology-Developmental Disabilities Net (Psych-DD),
& Editor, Journal of Behavior Analysis and Therapy (jBAT)<

Dear Joe:

Thanks for your interest. Let me know your reactions to the material in the book edited by Smith, if you manage to get hold of a copy.

Don

Date: Thu, 26 Oct 1995 09:11:26 EDT From: RBDL01A@prodigy.com (DR ARNOLD A LAZARUS) To: sscpnet@bailey.psych.nwu.edu Subject: Projective Tests

In our clinical Psy.D. program at Rutgers, the cognitive-behavioral faculty members were displeased when all sorts of projective tests were taught to our students. There was an interesting exchange on this topic between Don Peterson (when he was our Dean) and Robyn Dawes. Although Terry Wilson and I deplore this practice, we are outvoted by our colleagues who insist that unless our students emerge well prepared (projectively literate!) they will not get into many of the internship programs that demand a knowledge of these measures. For what it may be worth, let me add the fact that although most of our full-time faculty members are of a CBT persuasion, 90% of our students opt for the psychodynamic track.

Arnold

Date: Thu, 26 Oct 1995 09:15:01 -0700 (PDT) From: Gerald Davison <gdaviso@alnitak.usc.edu> To: John F Kihlstrom <john.kihlstrom@yale.edu> Subject: Re: Rorschach (lengthy file)
--

[John, my response follows heavily edited but I hope accurate reiteration of those parts of your message I would like to respond to.]

On Thu, 26 Oct 1995, John F Kihlstrom wrote:

>There is one good practical reason why a science-oriented clinical program would teach something like the Rorschach, and that is that the products of science oriented clinical programs are expected to be able to use the Rorschach on their internships, if not in their post-internship jobs.<
>The point is that a science-oriented internship site expected me to know the Rorschach, the TAT, the HTP, the Bender-Gestalt, the WAIS, and the MMPI.... This is still the case, broadly speaking. ...[we should move] heaven and earth to change the attitudes of internship sites about what a properly trained clinical student should be able and willing to do.<

.... change will occur

>even more rapidly if it is stimulated by the clinical programs themselves.<

Yes, it is the practical consideration that has some of our best clinical programs spend precious time teaching students how to use assessment (and intervention) techniques that have been around a very long time but that lack scientific grounding -- who will accept them as interns?

Tough question. We don't want to make it harder for our students to find internships. But let me recount a little story about the earliest days of the Stony Brook program, which began the same year I went there as a new assistant professor, 1966. I think it applies. We taught very very little in the way of projective testing. Goldfried, who was then completing a scholarly book with Stricker and Weiner on the Rorschach, taught the assessment course -- and he spent little time on the Rorschach. That is telling. We also discussed where our students would intern two or three years later. The decision, led by Len Krasner, was that we would not want our students to intern in places where such background was required, and that we would find and/or help shape internship settings that would match the behavioral, orientation that characterized our program and our students. Somehow we managed this. I don't know what the practices are these days at Stony Brook, but I would assume that little has changed.

To me, John, it's a question of leadership in clinical science and training, something you and many members of SSCP have made significant contributions to. You say that changes will occur if clinical programs take the lead. That is precisely what Stony Brook did in the late 60s and into the 70s and beyond. It was a calculated risk, I suppose, but Krasner et al had determined that it would work out well, and it did. (Never mind that some of what we believed back then was limited and not totally correct. That's the way it is in the science business. The fact is that we were trying to research and teach methods and approaches that offered promise in reshaping the field and improving service. And I think we did.)

Maybe it would help if we stepped back and looked at assessment for what it is, a knowledge gathering, knowledge-creating enterprise. One wants to find out certain things about a patient/group/community for a specific purpose. What are the best ways of learning/constructing what you need to know for that purpose? Put differently, if the Rorschach didn't exist, would we think to drop a blob of ink into a paper, fold it in half, and create an inkblot? Am I so hopelessly biased that I cannot generate a Yes answer to that question?

Final vignette. A year or two after I came here to USC in 1979, I sat with Chuck Nakamura, then DCT at UCLA, in an annual meeting of local Los Angeles VA training directors. A proposal was on the table and about to be implemented that would have required so many practicum hours for VA intern applicants as to distort his program and ours. We fought against it (even though we had not been able to discuss our strategy beforehand), and convinced the VA directors that they would get *better* interns if they did not pressure us into reshaping our programs to resemble those of professional

schools (CSPP's training director was also there, I believe, and he *liked* what the VA had been planning to do). We convinced them that clinically relevant research hours should count toward the clinical hours. It was, as George Albee has said, a small win, but it was a significant one. The result was and is that our clinical programs adhere to APA criteria, and those are good enough for the local VAs.

Sorry to have run on.
Jerry Davison

Date: Thu, 26 Oct 95 09:22:07 EST From: John Hunsley <HUNCH@acadvm1.uottawa.ca> Subject: Time for a Rorschach? To: SSCPNET <sscpnet@bailey.psych.nwu.edu>
--

As we go around yet again (I am really starting to reconsider my disbelief of the concept of repetition compulsion) on the Rorschach, net members may find some interesting data in a study by Ball et al. published last year in *Journal of Personality Assessment* (63, 239-249). The authors surveyed clinical practitioners regarding the time requirements for various commonly used assessment measures, including our dear friend the Rorschach. They report the following information for this instrument: time for administration $M=49$ minutes ($SD=22$), time for scoring $M=46$ minutes ($SD=28$), and time for interpreting $M=50$ minutes ($SD=29$). As recently suggested on this net, given the costs associated with practitioner time for this measure, the question of cost-effectiveness must also be addressed in addition to the standard psychometric concerns for psychological measures.

John Hunsley
University of Ottawa

Date: Thu, 26 Oct 1995 10:07:06 -0400 (EDT) From: "Robert C. Carson" <rcarson@acpub.duke.edu> To: sscpnet@bailey.psych.nwu.edu Subject: Re: Rorschach
--

Friends,

Notwithstanding a prior sacred oath never to again to comment on the perennial Rorschach problem on this net (having been clobbered last time around), I simply can't resist pointing out again that the arguments are, as usual, misdirected. For very many assessment techniques in both medicine and psychology, it just does not make a lot of sense to argue about the "validity" of the instrument independent of the characteristics of the user and his/her expertise with it. For example, I'm quite certain I would be useless or worse with a stethoscope, just as I can prove that I am with the Rorschach. I'm still pretty good with an MMPI profile, having written one of the original interpretation manuals on it, but not nearly as much so as I was when I used that technique practically daily. I'd say exactly the same thing, by the way, about "automated" MMPI interpretation printouts, and I'm sure my good friend Jim Butcher, author of the most successful of them, would agree.

Projective techniques (all of them) have always baffled me personally, but there is, after all, a rich literature on the relationship between the construal of experience and potentially interesting characteristics of the construer. It is far from impossible, therefore, that some folks might be quite good at discerning such relationships, especially where they've spent thousands of hours honing their skills and monitoring their hit rates (the latter being, in my judgment, so important as to approach an ethical obligation among clinical assessors).

Surely it is one of the easiest things in the world to generate a critique of the Rorschach (or the TAT, etc.), and sometimes it's even fun. I often think, however, that much of this criticism simply misses the point, an important one. By the way, I am not in saying this advocating extended teaching of particular assessment techniques in our already overcrowded graduate programs. I do believe we need to teach our kids how to think straight; if they learn along the way that there are no easy answers in our field, so much the better.

Bob Carson

Date: Thu, 26 Oct 1995 10:20:04 -0700 (PDT)
From: John McQuaid <mcquaid@itsa.ucsf.EDU>
To: John F Kihlstrom <john.kihlstrom@yale.edu>
Subject: Re: Rorschach

I am not convinced, in this Rorschach debate, that the Rorschach is such a pervasively required instrument at internship sites. In my internship process, I had no difficulty finding internships and being accepted, despite having never given a Rorschach. Since my personal experience differs from the perception of folks on this net, I wonder if anyone has data on internship requireents. I would expect that there is a reasonable number of internships that are not just counting # of Rorschachs. I would also suspect that students who are interested in this net would be more apt to apply to those internships that are less likely to require training in Rorschach.

If I am correct, and you don't HAVE to have learned this particular test to get a reasonable internship, then I think it frees students up to look at the internship process in a more functional way; basically, where do I match well. The best internship outcomes (and the greatest likelihood of acceptance) should come when a student finds an internship that matches their interests and goals. It bothers me to see folks training to "just get an internship" rather than thinking about what their overall goals are.

Given these thoughts, I think it is unwise to encourage training in and utilization of an instrument, like the Rorschach, with such major questions regarding validity. Continued use perpetuates the belief that you have to know it to get by (which is not true from what I have seen), and perpetuates the use of a questionable instrument when other, more valid techniques are available.

Date: Thu, 26 Oct 1995 10:30:49 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: "Donald K. Routh" <drouth@umiami.ir.miami.edu>,
Subject: Re: Rorschach

All

Re the Rorschach. It's a social dilemma situation. As lone as (50 + X)% of clinical training programs teach anything about Rorschach administration and interpretation, the APA true-believer dominated internship[sites are going to expect trainees who know how to "use" the Rorschach. The dominating strategy is then to tach it in order to keep one's students from being at a disadvantage and/or being one up on others in the internship site.

This dilemma was precisely the one we faced at Oregon when I was department head there. We were in a unique situation because our program for a number of reasons (in part, location) was so highly selective that no matter what we did or didn't do to the students, almost all eventually did wellin their internships--despite some loud and angry complaints at the

beginning of the year from their supervisors there that they didn't know the least bit about clinical psychology, e.g. the use of projective techniques or the spinning of psychoanalytic hypotheses in therapy contexts.

So we were able to get away with choosing the cooperative, dominated response

But that can't be expected as a general choice in this situation.

You people are talking about the utility of the R as if that followed simply from the consequences for the individual student and client. There's a broader problem there, and in the lack of concerted (meaning "in concert") action, nothing is going to improve.

Except for any emotional benefit of "abreaction" via email.

Robyn

P.S. a "healthy" arrogance that one does not need to worry about getting the (external) "suckers payoff" helps in a social dilemma.

Date: Thu, 26 Oct 1995 10:58:29 -0400 (EDT) From: John F Kihlstrom <john.kihlstrom@yale.edu> To: Gerald Davison <gdaviso@alnitak.usc.edu> Subject: Re: Rorschach

There is one good practical reason why a science-oriented clinical program would teach something like the Rorschach, and that is that the products of science oriented clinical programs are expected to be able to use the Rorschach on their internships, if not in their post-internship jobs.

I am a product of a science-oriented clinical training program, albeit one that gave us the bare minimum of practical training in assessment and therapy. I took my internship at Temple, the Department of Psychiatry, very research oriented, where the doyen of the department was Charles Shagass, himself an Iowa MA in psychology, and someone who held clinical psychology, and psychological research, in the highest regard. My supervisors were Barry Shmavonian and Lyle Miller, two highly regarded clinical researchers. THE VERY FIRST THING Shamvonian did, on the first day of my internship, was to thrust a set of Rorschach cards in my hands and tell me to test an inpatient. Had Julie Wishner not taught me how to administer and score the Rorschach, I would not have been able to do this. WHEN THAT PATIENT WAS PRESENTED AT GRAND ROUNDS Shagass immediately called for a report of the psychological testing, and wanted a full account of the Rorschach. Had Julie Wishner not taught me how to score and interpret the Rorschach, I would not have been able to do this.

Whether I did so holding my nose, which in fact I did, is beside the point. The point is that a science- oriented internship site expected me to know the Rorschach, the TAT, the HTP, the Bender-Gestalt, the WAIS, and the MMPI. This is still the case, broadly speaking. And given such an environment, we

do our students a gross disservice if we do not (a) train them to administer these tests; (b) train them to be extremely skeptical about these tests until their reliability, validity, and utility is proven; (c) train them to develop better tests, and use them when and if they become available; and (d) move heaven and earth to change the attitudes of internship sites about what a properly trained clinical student should be able and willing to do.

Reform of clinical training may include abandoning the projective techniques, though frankly the OBJECTIVE techniques aren't all that great either. But reform cannot occur unless and until the market is changed. Managed care will help change that market, and that is one of the reasons I believe that managed care will be good for clinical psychology. But change will occur even more rapidly if it is stimulated by the clinical programs themselves.

The continuing insistence of internship sites that students be trained in dubious techniques is the best argument I know for the proposition that every clinical training program should control its own internship. Unless this happens, clinical training programs will never be able to control their own curricula, and our students will continue to be trained in dubious techniques as a matter of economic necessity.

John F. Kihlstrom, Professor
Department of Psychology,
Yale University
P.O. Box 208205,
New Haven, Connecticut 06520-8205
Telephone (203) 432-2596
Facsimile (203) 432-7172

From: BOBAPPIC@aol.com Date: Thu, 26 Oct 1995 11:14:16 -0400 To: sscpnet@bailey.psych.nwu.edu Subject: Re: Projective Tests
--

Once again, internships are cited as the reason why other- minded graduate programs still require their students to learn the Rorschach. Once again, I point out that there are over 500 internship programs in the APPIC directory, many of whom do not require background in the Rorschach of their applicants.

I would like to see grad programs determine what *they* believe to be the best in graduate training, and help their students apply to internships that are compatible with their training programs. Those grad programs that view heavy projective experiences as critical to training good psychologists will find plenty of internships happy to accept people with that graduate school background; those programs that eschew the Rorschach will find plenty of programs (including my own) that are more than happy to consider applicants without that background. Of course, this tack means that a student of yours

who wants an internship whose program or goals are different from your grad program's goals will have a lesser chance of admission to their preferred program. That's a dilemma that each grad program faculty must weigh.

Bob Klepac
bobappic@aol.com

Date: Thu, 26 Oct 1995 12:04:39 -0700 (PDT) From: Jake Jacobs <jakej@alnitak.usc.edu> To: sscp <sscpnet@bailey.psych.nwu.edu>

We are looking for something called the Expressional Fluency Test (with a reference to Christensen, P. R. & Guilford, J. P. 1958) also called the Christensen-Guilford fluency test. It was (at sometime in the past) available from the Sheridan Supply Company in Beverly Hills, CA. They are no longer in business...and our library has not been able to find the reference. Any clues, copies, reprints, or direction from members of this group would be greatly appreciated.

W. Jake Jacobs

Date: Thu, 26 Oct 1995 12:57:33 -0500 To: "William Thomas O'Donohue" <wto@unr.edu> From: "Joseph J. Plaud" <plaud@badlands.nodak.edu> Subject: Re: Rorschach

At 08:41 AM 10/26/95 -0700, William Thomas O'Donohue wrote:

>I am afraid that I would bet that 100 years from now a good proportion of clinical psychologists will be selling a lot of Yugos.<

Quite true, Bill, but hopefully *not* on SSCPnet!

With regard to the "R" it is important to remember that it is being utilized by lots of psychologists in varieties of clinical situations, and we are really being led to believe that external criteria (such as internships and the "real world" demands it) justify its use. Wow! Imagine a quote like this coming from the worlds of biology, chemistry, physics of medicine:

"The Rorschach has not yet gained the status of scientific respectability enjoyed by many other personality tests, and perhaps it never will"

(Reference: Gregory, R. J. (1996). Psychological Testing. Boston: Allyn & Bacon, p. 516).

This is typical of the type of reports that Marvin gave us in a recent post. By all established criteria, the "R" is in its own orbit, *but* we still use it, still support its use in graduate training and internships, in clinics and hospitals, still entertain its usage here in the Society for a Science of Clinical Psychology. Should this be the way we do business? I think you know my vote...

Joe Plaud
Back in the Saddle

Date: Thu, 26 Oct 1995 13:31:03 -0400 (EDT) From: John F Kihlstrom <john.kihlstrom@yale.edu> To: SSCP Network <sscpnet@bailey.psych.nwu.edu> Subject: My Last Word on the Rorschach
--

Evolution is slow, but it probably doesn't have to be as slow as it has been with regard to clinical assessment. Both clinical training programs and internship sites have to pick up the pace. But until the mammals have clearly triumphed over the saurians, can someone please answer a couple of rhetorical questions for me?

(1) Why are accredited internships allowed to require that applicants be trained in techniques of dubious reliability and validity? Eliminating those requirements should be a condition of accreditation.

(2) Until such requirements are removed, why would any clinical training program offer a curriculum that effectively constrains its students' choices of future internship sites?

So long as a critical mass of internship sites require their applicants to know how to administer, score, and interpret the Rorschach, I don't see any justification for failing to acquaint our students with this material. Spending just a little time on the Rorschach, and explicitly evaluating the technique against established psychometric standards, would certainly be in students' short-term interests, but it would also be in their long-term interests. After all, scientist practitioners are supposed to carry scientific values -- healthy skepticism, attention to evidence, etc. -- into their clinical practices. Learning how the Rorschach stacks up might well reinforce student's abstract learning of such concepts as reliability, validity, and utility; and this would be a lesson that might last a lifetime.

John F. Kihlstrom, Professor

Department of Psychology,
Yale University
P.O. Box 208205,
New Haven, Connecticut 06520-8205
Telephone (203) 432-2596
Facsimile (203) 432-7172

Date: Thu, 26 Oct 1995 13:33:48 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: "Robert C. Carson" <rcarson@acpub.duke.edu> Subject: Re: The blot on our escutcheon
--

Lee Sechrest, Univ. of Arizona

The Rorschach is an embarrassment; of that there can be no doubt. So what to do?

One strategy might be benign neglect. That may, in fact, be something of what is involved in John Kihlstrom's solution. Teach enough of it to fill a void with gas but no more. Eventually it will die out.

Another strategy might be (malignant) embrace. A. Bandura remarked somewhere or other that theories never actually are actively abandoned. They simply, gradually become so limited by boundary conditions that they become uninteresting. Maybe we should actively *teach* the Rorschach in the light of what (even if little) evidence it is for its reliability, validity, and utility. Encourage students to use it within those limitations but for no other purposes. Interest would decline.

A third possibility, and one that we should undertake anyway, would be to develop better understanding of just what decisions it is that clinical psychologists have to make and the kind of evidence that would be useful in making those decisions. Then consider whether the Rorschach provides any such evidence. I do not think we know very much at all about what decisions it is that clinicians have to make for which they rush to the test cupboard. There is a Society for Medical Decision Making that studies in a systematic way the decisions that physicians have to make and the processes that underlie those decisions. There is no such society for clinical psychology decision making.

My guess is that the Rorschach is used mostly proforma for the purposes of "unbounded prediction" as Leon Levy called it. That is, the Rorschach is most often given not because of any particular decision that has to be made but as an all purpose instrument to predict pretty nearly anything at all that might occur to anyone involved in the case. That is pretty stupid when you think of

it. What other field has such an instrument? Or pretends to? (Astrology, palm reading, for two)

Otherwise, the Rorschach is probably used most in forensic or quasi-forensic settings, where it is most dangerous and for which it is least suited. We ought to emphasize to (scare the pants off of) students that the use of the Rorschach and other such instruments in forensic cases could expose them to all sorts of legal dangers, we ought to do more to provide direct information to the legal profession about the limitations of the Rorschach so that its use would never go unchallenged, and we ought to publish more review papers to make the case.

Jim Wood at UT El Paso has done some work on the Exner system that deserves to be widely known. Maybe this network could be a useful medium by which to disseminate information. Get people to write the apa monitor, protest to the accreditation committee, insist that the Science Directorate become more proactive on such matters.

Date: Thu, 26 Oct 1995 14:06:34 -0500 To: sscpnet@bailey.psych.nwu.edu (SSCPnet) From: "Nancy D. Vogeltanz" <vogeltan@badlands.nodak.edu> Subject: Re: Rorschach etc..

To Jerry Davison: Yes, yes, yes. What would this field do without you?

To Bill O'Donohue: I'm not sure I completely understood that crafty Yugo metaphor (you philosophers are so very deep), but I think it means that psychologists don't seem to behave according to the principles of natural selection. Now I'm really getting frightened.

To the Esteemed A. Lazarus: If 90% of your faculty is cognitive-behavioral and 90% of your students go into the psychodynamic track, it seems that your token Freudian is really getting a work out while the rest of you sit idly by wishing for someone to influence. It is hard for me to believe that the 90% cog-behav. faculty doesn't come up with some clever PR work (maybe you could use the Yugo metaphor) and convince these students that experimentally-based clinical science is better.

Nancy D. Vogeltanz, Ph.D.
University of North Dakota P.O. Box 8380
Department of Psychology Grand Forks, ND 58202-8380 (701) 777-3790
E-mail: Vogeltan@Badlands.NoDak.Edu

Date: Thu, 26 Oct 1995 15:14:34 -0400 (EDT) From: "Robert C. Carson" <rcarson@acpub.duke.edu> To: SSCP@acpub.duke.edu Subject: Re: Rorschach

Lee,

Your cogent remarks stimulated a further thought. I guess what I really believe about this stuff is that it's virtually **always** the clinician that must be "validated," whether he/she uses interviews, (my own sine qua non), inkblots, Blacky pictures, MMPI printouts (notoriously self-contradictory), Tarot cards, or tea leaves. It's an interesting question whether or not there would be some sort of "g" factor operative across across instruments, but as far as I know that's never been addressed, partly I suspect because of our preoccupations with the alleged merits (or their lack) of the instruments--the old saw about a hammer and its uses occurs to me here, but I'll suppress it. Almost certainly, differing techniques vary on the general quality of the data they yield as a platform for inferences, irrespective of overall examiner perspicacity. Nevertheless, I'd bet that the instrument-examiner interaction term would always be non-negligible in well-designed, comprehensive assessments of the assessment enterprise--again mostly because some folks who work at it teach themselves to be quite good with particular techniques, which I find not exactly surprising.

Bob Carson

Date: Thu, 26 Oct 1995 16:02:35 -0700 (PDT) From: Steven Taylor <taylor@unixg.ubc.ca> To: Samuel M Turner <turnersm@musc.edu> Subject: Rorschach and internships

I agree with Dr. Turner that it would be most interesting to see data on the types of assessment skills that are required and encouraged by internship programs.

Our internship program (which is APA and CPA accredited) does not require interns to be familiar with projectives. Nor do we encourage their use. We emphasize behavioural assessment and structured interviews, with self-report measures as useful adjuncts. However, I believe we do have a Rorschach set buried somewhere in our test library, along with the Szondi cards.

Steve Taylor

On Thu, 26 Oct 1995, Samuel M Turner wrote:

>We keep hearing that internships are responsible for programs needing to teach projective techniques. I'm not convinced that internships are responsible. Does someone have data on this? Most of the science oriented internships I know do not require such skill, and I think I know a lot of them. But then there usually is a lot of information floating around about what internship programs require, most of which is erroneous. I do suspect that some of the old traditional departments (we all know who they are) whether science oriented or not, continue to value projectives. I do believe, however, that anyone looking for an internship that does not require projectives can find it.<

>Samuel M. Turner<

Date: Thu, 26 Oct 1995 16:16:17 -0500 (CDT) From: clark lee anna <laclark@blue.weeg.uiowa.edu> To: "Robert C. Carson" <rcarson@acpub.duke.edu> Subject: Re: Rorschach
--

On Thu, 26 Oct 1995, Robert C. Carson wrote:

>Your cogent remarks stimulated a further thought. I guess what I really believe about this stuff is that it's virtually *always* the clinician that must be "validated," whether he/she uses interviews, (my own sine qua non), inkblots, Blacky pictures, MMPI printouts (notoriously self-contradictory), Tarot cards, or tea leaves.<

What ever happened to Meehl's documentation of the superiority of actuarial (vs. clinical) prediction?

LAC

Lee Anna Clark, Ph.D.

Department of Psychology
University of Iowa
E 11 Seashore Hall
Iowa City, IA 52242-1407
Tel: (319) 335-3391
Fax: (319) 335-0191

Date: Thu, 26 Oct 1995 17:06:15 -0400 (EDT) From: Samuel M Turner <turnersm@musc.edu> To: Gerald Davison <gdaviso@alnitak.usc.edu> Subject: Re: Rorschach (lengthy file)

We keep hearing that internships are responsible for programs needing to teach projective techniques. I'm not convinced that internships are responsible. Does someone have data on this? Most of the science oriented internships I know do not require such skill, and I think I know a lot of them. But then there usually is a lot of information floating around about what internship programs require, most of which is erroneous. I do suspect that some of the old traditional departments (we all know who they are) whether science oriented or not, continue to value projectives. I do believe, however, that anyone looking for an internship that does not require projectives can find it.

Samuel M. Turner

Date: Thu, 26 Oct 1995 18:12:17 -0500 (CDT) From: "Joseph J. Plaud" <plaud@badlands.NoDak.edu> Subject: Re: Rorschach To: clark lee anna <laclark@blue.weeg.uiowa.edu>

On Thu, 26 Oct 1995, clark lee anna wrote:

>On Thu, 26 Oct 1995, Robert C. Carson wrote:<

>>Your cogent remarks stimulated a further thought. I guess what I really believe about this stuff is that it's virtually *always* the clinician that must be "validated," whether he/she uses interviews, (my own sine qua non), inkblots, Blacky pictures, MMPI printouts (notoriously self-contradictory), Tarot cards, or tea leaves.<<

>What ever happened to Meehl's documentation of the superiority of actuarial (vs. clinical) prediction?<

>LAC<

It has been banished to out-of-print status (where *most* of the good books in psychology happen to be anyway). I use Meehl's Psychodiagnosis text for my graduate clinical assessment sequence, and last year the University of Minnesota Press stopped publishing it. I had to make special arrangements to have our University bookstore xerox it (paying the copyright fee, of course) for my graduate students. Contained within the xeroxed pages are the most seminal of Meehl's writings on this most important topic that should be used in the R debate...excellent point!

Joe Plaud
University of North Dakota

Date: Thu, 26 Oct 1995 18:13:21 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: sscpnet@bailey.psych.nwu.edu, BOBAPPIC@aol.com Subject: Re: Projective Tests
--

Bob,

Maybe things have improved re the internships since 1985 (the last year I was head of a psychology department).

hope so.

Base rate reasoning, however, does not yield optimism.

Robyn

Date: Thu, 26 Oct 1995 19:02:28 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: "Robert C. Carson" <rcarson@acpub.duke.edu>, clark lee anna <laclark@blue.weeg.uiowa.edu> Subject: Re: Rorschach

"Whatever happened to?" See Dawes, Faust, and Meehl *Science*, 1989, 243, 1668-1674. See Grove and Meehl (in press) in *Psychology, Public Policy, and Law*.

What has happened is replication after replication after replication in qualitatively diverse fields.

Together with an increasingly absurd discrepancy between this research literature and how "psychologists" actually make a living.

Robyn Dawes

Date: Thu, 26 Oct 1995 20:47:04 -0400 (EDT) From: "Robert C. Carson" <rcarson@acpub.duke.edu> To: clark lee anna <laclark@blue.weeg.uiowa.edu> Subject: Re: Rorschach
--

Lee Anna,

Meehl is still right, as usual. The problem here is that it is enormously difficult and expensive to generate actuarial data on many of the questions that confront clinicians (and I mean here at both ends, predictors and outcomes), not to mention substantial noise due to local boundary conditions. Where such applicable data exist, I went on record at least 25 years ago to assert that the "actuarial" is always to be preferred to the "clinical." I also believe that Dawes et al. are right in asserting that local clinicians could significantly enhance their eventual effectiveness by beginning the task of generating an adequate actuarial data base for the decisions they are normally asked to render. Meanwhile, however, decisions have got to be made. I just don't happen to think the "validity" (such as it is) of various assessment instruments is as crucial a variable in this context as the people who use them.

Bob Carson

Date: Thu, 26 Oct 1995 21:01:31 -0400 (EDT) From: "Robert C. Carson" <rcarson@acpub.duke.edu> To: Lee B Sechrest <sechrest@U.ARIZONA.EDU> Subject: Re: The blot on our escutcheon
--

On Thu, 26 Oct 1995, Lee B Sechrest wrote:

>Date: Thu, 26 Oct 1995 13:33:48 -0700 (MST)
>From: Lee B Sechrest <sechrest@u.Arizona.EDU>
>To: "Robert C. Carson" <rcarson@acpub.duke.edu>
>Cc: SSCP@acpub.duke.edu, SSCP Network <sscpnet@bailey.psych.nwu.edu>
>Subject: Re: The blot on our escutcheon<

>Lee Sechrest, Univ. of Arizona<

>The Rorschach is an embarrassment; of that there can be no doubt. So what to do?<

Lee,

I have doubts. I am far less embarrassed by the existence of 10 inkblots and the half-dozen "systems" purporting to tell how to use them than I am by the stupidity and mysticism that has grown to surround them--or by the *requirements* that intern applicants certify that they have "given" a dozen or more of these; the latter really does strike me as an extravagant misunderstanding of what the hell our field is all about.

I sort of like your alternative modes of coping though.

Bob

Date: Thu, 26 Oct 1995 21:10:38 -0400 (EDT) From: "Robert C. Carson" <rcarson@acpub.duke.edu> To: Robyn Dawes <rd1b+@andrew.cmu.edu> Subject: Re: Rorschach
--

Robyn,

I'm in essential agreement on this point (see my earlier reply to Lee Anna). As I learned last time around in this general debate, it's very difficult to avoid stereotypic attributions in such a heated atmosphere. But then, I'm sure you've had your own experience with such matters.

Bob Carson

Date: Fri, 27 Oct 1995 06:32:27 -0700 (PDT) From: Jake Jacobs <jakej@alnitak.usc.edu> To: Steven Taylor <taylor@unixg.ubc.ca> Subject: Re: Rorschach and internships

As I read the continuing exchange about the R, I find I am increasingly worried. Most programs continue to teach the R. Meanwhile, we are seriously debating whether projectives should or should not be taught...we do so in the face of this fact: Many programs do not teach fundamentals such as behavior assessment and behavior analysis. There are at least three top-rated programs within my range of knowledge where students can achieve their Ph.D. in clinical psychology without any exposure to these basic, well-established, and highly useful tools. Although I do not mean for the tone of this note to be ungracious, it seems to me that the leaders of our field have lost their way. Unfortunately, they are us.*

W. Jake Jacobs

*with apologies to Pogo

On Thu, 26 Oct 1995, Steven Taylor wrote:

>I agree with Dr. Turner that it would be most interesting to see data on the types of assessment skills that are required and encouraged by internship programs.<

> Our internship program (which is APA and CPA accredited) does not require interns to be familiar with projectives. Nor do we encourage their use. We emphasize behavioural assessment and structured interviews, with self-report measures as useful adjuncts. However, I believe we do have a Rorschach set buried somewhere in our test library, along with the Szondi cards.<

>Steve Taylor<

From: ssoldz@warren.med.harvard.edu Date: Fri, 27 Oct 1995 08:20:26 -0400 Subject: Re[2]: The blot on our escutcheon To: rcarson@acpub.duke.edu, Lee B Sechrest <sechrest@U.ARIZONA.EDU>

Lee,

I think that the best strategy of those you suggest is to teach the R in the context of decision making. As you point out, there is little relation between the decisions clinicians make and the tools they use. A typical referral for projectives when I did my internship 9 years ago was : "Is this person psychotic?" or "Is pt. schizophrenic or depressed"? One amusing aspect of the latter question was that all the indicators we were taught, even if valid, had been developed in DSM-II days and we were living in DSM-III times! Thus, any valid indicator from the literature would have to have terrible hit rates for us. But I could never get even a discussion going regarding this glaring contradiction. As to the former question, if the Doc and staff on an inpatient/day hospital didn't know if someone was psychotic, the testing wasn't going to help. In one case, my supervisor (a prominent neuropsychologist) interpreted the R to show gross thought disorder. The Doc thought this didn't match her experience. So, the testing was ignored (as it should have been. Even if it was "true" in some sense, it would have had no treatment implications).

Another fundamental problem is that testers who rely on their clinical experience usually have no way of learning from that experience. Either a testing confirms what was already known re: a pt., in which case there is no incremental validity, or else it says something unknown (as in the pt. above). In the latter case, how can the tester know if (s)he is correct? In the vast majority of clinical settings, (s)he will never get systematic feedback. So, one will probably only hear anecdotal confirming instances: "You know, you said that pt. was suicidal, and. you know, 9 months later they did attempt suicide.

Amazing!" But never: "You know, you said that pt. was suicidal, and. you know, 9 months later, they are quite happily married."

As to the forensic issues, they would require systematic and continued public presentation of the issues. I once had a lawyer friend who was representing a woman who was about to have her child removed. The State DSS had a testing report saying that she was danerous, based on the R. I told my friend of all the negative evidence regarding the validity of the R and gave her some references. She made the mistake of asking the psychologist about these issues: he was totally ignorant. But when it got to court and he was on the stand, and he started asking about the validity of the R, he'd done some reading and started sounding like an expert. My friend, (this not being the OJ case, with millions behind her, but your average pro bono case), did not know enough to challenge his references and the psychologist looked quite competent, loosing the case for my friend. The moral is that lawyers would have to have a lot of info to challenge testers, as they can always develop a few sources to make them look good.

As to Exner, even if the research was valid, it would be irrelevant to most clinicians. The reason is that Exner's norms are based on his administration of the R, which has minimal inquiry (regarding location and reasons for percepts). But surveys consistently show that most clinicians, including those who claim to use Exner, use must more inquiry, presumably because they want to also do content interpretation. So the Exner norms will be totally invalid for the majority of those relying on them.

Cheers,
Stephen
Stephen Soldz
Director of Research and Evaluation
Department of Psychiatry
Advocates for Human Potential
Harvard Medical School
Sudbury, MA
531 Beech St.
Director of Research
Roslindale, MA 02131
Boston Institute for Psychotherapy
(617) 469-3576
ssoldz@warren.med.harvard.edu

Date: Fri, 27 Oct 1995 08:35:18 -0700 (PDT) From: Steven Taylor <taylor@unixg.ubc.ca> To: Jake Jacobs <jakej@alnitak.usc.edu> Subject: Re: Rorschach and internships

Jake,

You raise an important point. How many internship programs require prospective interns to be trained in behavioural assessment (including behavioural analysis)? My guess is 'not many.' We don't, and for some interns we provide their first experience in BA. In previous postings on this topic, people have suggested that the R be taught in graduate school because interns will be required to use on internship. I have never heard of anyone using the same argument for teaching BA.

Steve

On Fri, 27 Oct 1995, Jake Jacobs wrote:

>As I read the continuing exchange about the R, I find I am increasingly worried. Most programs continue to teach the R. Meanwhile, we are seriously debating whether projectives should or should not be taught...we do so in the face of this fact: Many programs do not teach fundamentals such as behavior assessment and behavior analysis. There are at least three top-rated programs within my range of knowledge where students can achieve their Ph.D. in clinical psychology without any exposure to these basic, well-established, and highly useful tools. Although I do not mean for the tone of this note to be ungracious, it seems to me that the leaders of our field have lost their way. Unfortunately, they are us.*<

>W. Jake Jacobs
*with apologies to Pogo<

>On Thu, 26 Oct 1995, Steven Taylor wrote:<

>>I agree with Dr. Turner that it would be most interesting to see data on the types of assessment skills that are required and encouraged by internship programs.<<

>>Our internship program (which is APA and CPA accredited) does not require interns to be familiar with projectives. Nor do we encourage their use. We emphasize behavioural assessment and structured interviews, with self-report measures as useful adjuncts. However, I believe we do have a Rorschach set buried somewhere in our test library, along with the Szondi cards.<<

>>Steve Taylor<<

Date: Fri, 27 Oct 1995 08:38:14 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Blatt-Ford Book
To: sscpnet@bailey.psych.nwu.edu

To: Sidney J. Blatt
From: Donald K. Routh

The book summarized in your e-mail note today seems to be an example of serious research still going on quietly these days using Rorschach, TAT, and other such procedures, and mostly ignored by the discussants on this net in the last couple of days.

One question I have is why these data are published in a book rather than in ordinary peer reviewed journal articles.
Don

Date: Fri, 27 Oct 1995 10:26:44 -0400 (EDT)
From: "Alan S. Bellack" <abellack@umabnet.ab.umd.edu>
To: Gerald Davison <gdaviso@alnitak.usc.edu>
Subject: Re: Rorschach (lengthy file)

Jerry,

As always, you are right on target about training programs taking the lead. As someone who has been a Director of Training at an APA approved Clinical program (Univ of Pittsburgh) and Director of Psychology at two medical schools with APA approved internships (MCP and Univ of Maryland) I have an additional point about training content. My sense is that most internships that demand specific technical skills (e.g., # of Rorschachs completed in graduate school) want/need interns to serve as cheap labor. They need interns to pick up heavy caseloads right away. If they are truly interested in "teaching", they can develop desired skills throughout the year. The most important requirements for intern applicants should be: a) that students have had enough sound clinical experience to be trusted to sit with patients without a supervisor present and not do anything stupid or unethical (a surprisingly difficult requirement to meet given the sparse clinical training provided by many Clinical programs), and b) that they desire the nature and content of the training provided. Graduate students should be trained for their careers, not for 1-year of forced labor.

Alan

Date: Fri, 27 Oct 1995 10:34:29 -0400 (EDT)
From: Sidney Blatt <blatts@minerva.cis.yale.edu>
To: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Blatt-Ford Book

We did publish two reports in journals -- one in 1988 and the other this year, but the data are extensive and complex and we need a much broader canvas to present the total picture. Also we got a lot of grief from editors and reviewers for our refusal to use DSM diagnoses, which I think for the most part are meaningless and trivial; and for our not using the Exner system, which I have had considerable difficulty with for years, primarily because many of his ratio and summary are atheoretical and based on claims of empirical support much of which was never peer reviewed. And often Exner's ratios and summary scores made little sense to me based on my theoretical, clinical, and research experience with the Rorschach.

Sid

On Fri, 27 Oct 1995, Donald K. Routh wrote:

>To: Sidney J. Blatt<
>From: Donald K. Routh<

>The book summarized in your e-mail note today seems to be an example of serious research still going on quietly these days using Rorschach, TAT, and other such procedures, and mostly ignored by the discussants on this net in the last couple of days.<

>One question I have is why these data are published in a book rather than in ordinary peer reviewed journal articles.<

>Don<

Date: Fri, 27 Oct 1995 10:56:32 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Blatt-Ford Book
To: Sidney Blatt <blatts@minerva.cis.yale.edu>

Earlier today (10-27-95) Sidney J. Blatt stated:

>We did publish two reports ...>see above<

Reply to Sidney Blatt's comment:

Being a journal editor myself, I certainly know what a hassle one has to go through to publish articles, but I think the field benefits from this process. This is especially true when an author operates from an unpopular theoretical perspective or uses unpopular procedures. For this reason, I was pleased to

see the appearance of a journal such as Psychological Assessment that would presumably be more open to the kind of research we are discussing but would apply rigorous standards, too.

Don

Date: Fri, 27 Oct 1995 10:59:35 -0700 (PDT) From: Gerald Davison <gdaviso@alnitak.usc.edu> To: "Alan S. Bellack" <abellack@umabnet.ab.umd.edu> Subject: Re: Rorschach (lengthy file)

On Fri, 27 Oct 1995, Alan S. Bellack wrote:

>Jerry,
>...[deleted] I have an additional point about training content. My sense is that most internships that demand specific technical skills (e.g., # of Rorschachs completed in graduate school) want/need interns to serve as cheap labor. They need interns to pick up heavy caseloads right away. If they are truly interested in "teaching", they can develop desired skills throughout the year....<

Alan,

Yes, indeed. In an otherwise good internship at Va Palo Alto, I was plunked my first day into the role of sole psychologist on a partially closed ward. Sink or swim is not always the best didactic strategy. The reason, of course, was cheap labor. Six months later I was pretty much on top of things, but not at the outset.

Jerry

Date: Fri, 27 Oct 1995 11:04:32 -0700 (PDT) From: Gerald Davison <gdaviso@alnitak.usc.edu> To: Steven Taylor <taylor@unixg.ubc.ca> Subject: Re: Rorschach and internships
--

On Fri, 27 Oct 1995, Steven Taylor wrote:

How many internship programs require
>prospective interns to be trained in behavioural assessment (including behavioural analysis)? My guess is 'not many.' We don't, and for some interns we provide their first experience in BA. In previous postings on this topic, people have suggested that the R be taught in graduate school because interns will be required to use on internship. I have never heard of anyone using the same argument for teaching BA.<

>Steve<

Excellent points by Steve Taylor and USC colleague Jake Jacobs. The reason is simple: behavioral analysis is not "a test." It therefore does not fit into a "test battery." The approach also hasn't been around as long as the R and other things, hence it is suspect. We all know that techniques and approaches that have been in use a long time are ipso facto valid and useful.

Jerry Davison

Date: Fri, 27 Oct 1995 11:25:53 -0400 (EDT) From: Sidney Blatt <blatts@minerva.cis.yale.edu> To: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Blatt-Ford Book

Don: I welcome a ranking journal in assessment that is not doctrinaire. I must admit that I felt a strange sense of compassion in the recent communications on the net from the behavioral types who felt misunderstood and unappreciated by colleagues. From the other end of the spectrum, I have felt the same sense of alienation being classified as psychodynamic and interested in fuzzy procedures like the Rorschach. It has been difficult to get funding, and editorial reviews seem sometimes driven by less than rational considerations. Editorial reviews can be valuable source of learning and of shaping the final product if they are done with an openness to currently unpopular views and if the work is judged in its own merit. The Riggs-Yale project, which was based on incredibly rich clinical data, gathered in a remarkably systematic and thorough way, was denied support by NIMH. Fortunately we were able to get MacArthur Foundation support for the project.

Thanks for your interest in the work.

Sid

Date: Fri, 27 Oct 1995 17:19:51 -0400 From: stricker@sable.adelphi.edu To: sscpnet@bailey.psych.nwu.edu Subject: The Rorschach

I had vowed to stay out of the Rorschach discussion until Jerry Davison was good enough to remember the Rorschach book that I did years ago with Marv Goldfried and Irv Weiner. I still think of it as the best work on the issue, because it approached the Rorschach as a psychometric instrument and evaluated it by those criteria. The findings were mixed - not as bleak as some portray, and nowhere as promising as the votaries assume. Because it was so good (I humbly assert), it now is out of print. Anyway, the very last paragraph of the book stated:

"In summary, then, the Rorschach may be approached either as a perceptual-cognitive task, as a structured interview, or as a stimulus to fantasy productions, and Rorschach indices all relate primarily to one or another of these approaches. Indices based on the first two approaches emphasize the subject's manner of structuring his percepts and verbalizing his responses, and these indices are fairly directly representative of behavior; indices based on the third approach concern response content and constitute symbolic representation of behavior. It is the directly representative indices that offer the most promise for cross-validation and broad applicability, whereas symbolic indices appear particularly prone to errors of measurement and the persistence of illusory correlation despite empirical contradiction."

Thus, used at its best, the Rorschach may reduce to a structured interview, raising Lee's question of incremental validity (but it does not take a skilled interviewer anywhere near 3 hours to administer, score and interpret), and satisfying Bob Carson's desire for the interview as the ultimate diagnostic technique. And, data about the utility of interviews ain't so good either.

George Stricker

Internet address (STRICKER@SABLE.ADELPHI.EDU)
Telephone (516) 877-4803
Fax (516) 877-4805

Significance

letters from sscpnet 8/95

Date: Mon, 13 Nov 1995 10:46:29 -0600 To: sscpnet@bailey.psych.nwu.edu From: k-howard@nwu.edu (Ken Howard) Subject: More on significance testing

From another net:

Not that I agree with such a sharp distinction between individual studies and meta-analyses. A meta-analysis of all studies of feral children is still going to be based on less than a couple dozen observations of any given effect (presuming we could find a particular effect that had been measured in each case) and will provide a correspondingly imprecise estimate of that effect. The crucial variable (if we assume homogeneity of studies with respect to scores on important moderating variables) is simply total N -- as your paper hints at in treating the "single larger validity study with 1,428 subjects" as if its correlation of .22 were a population value.

Of course, studies are not usually homogenous, so consistency across different studies conducted under very different conditions provides considerably greater confidence than a single study with the same total N that the effect is a robust one.

I'm also pleased that you disavow one-tailed tests (along with all significance tests), though I would sure appreciate a note of caution to your readers that one-tailed tests are an especially abhorrent variant, lest readers as careless as I take what appears to be the pejorative tone of your comments on the lower power of two-tailed tests (which is true only if your a priori hypothesis is correct; overall power depends on the likelihood that you're right, i.e. on the relative weight given the slightly higher power when right and the zero power when wrong about the population sign) as indicating the reverse preference between these two "evils".

While I'm spouting off anyway, let me suggest that your statement that significance level gives NO information about the probability of replication is too strong. You're certainly right that significance at the alpha level doesn't imply a replication probability of $1 - \alpha$.

However, with N and sigma fixed (or N fixed and effect size expressed in standard-deviation units) the larger the population effect size, the lower the average p-value across a set of studies. Similarly, the larger the effect size, the higher the probability that two successive studies will yield statistical significance at a given alpha level in the same direction. Thus there will be a negative correlation between p-value and probability that an effect will replicate in the sense just described. In a paper that's in press in Psychophysiology (authored by Tony Greenwald, Rich Gonzalez, Don Guthrie, and me, though the result

I'm about to focus on was mostly Don's work) a reasonable set of assumptions leads to the conclusion that a p-value of .005 (note the extra zero) is associated with about an 80% chance that an exact replication (same N, sampling from the same population, etc.) will yield statistical significance at the .05 level and in the same direction as the initial study. (The initial issue presented to the committee Tony oversaw was under what conditions an

author reporting a result inconsistent with past results or with generally accepted theory should be required to replicate the finding before publication.)

Finally, to your list of false beliefs about the benefits of significance testing, I wish to append what I feel is a correct belief: While a highly significant (e.g., $p < .001$) result does not necessarily imply that the effect tested is large or important, it DOES indicate that you are considerably less likely to be wrong in your conclusion about the sign of the population effect than if your significance test had yielded a p-value of, say, .049. No conclusion based on variable data (whether single-study or meta-analytic) is error-free, but a researcher is justified in feeling less nervous about asserting that the population effect matches in sign that of the sample result if he has a highly significant result, rather than a barely significant one.

Oh, well. One more. A number of participants in this discussion have suggested that our null hypothesis is best considered an hypothesis that the population effect size doesn't depart by more than a trivial (unimportant) amount from the null-hypothesized value. I teach my students (as I'm sure lots of us do) that the one exception to the general rule that one never accepts H_0 (just fail to reject it) is that you may be able to accept the MODIFIED null hypothesis of no IMPORTANT departure from the hypothesized value (usually zero). The conditions for doing so are, first, that you specify and justify on practical or theoretical grounds just what a minimally important effect size is (i.e., what the boundary between unimportant and important effects is) and, second, that the CI (95%? 99%?) around your point estimate not include any values larger than this minimally important effect size. This is not an easy thing to do.

Dana Quade (Harris & Quade, J. Ed. Stat., 1992) showed that the sample size needed to have a 1-alpha chance that your $(1-\alpha)*100\%$ CI include no important values is at least 4 times as great as the N you would need to disprove the null hypothesis (with 50% power) if the population effect were just above the minimally- important threshold. And a large proportion of consultees are, in my experience, very reluctant to commit themselves to a particular effect size as minimally important. I believe that one of the recent notes cited a paper that recommends a highly similar procedure and suggests, essentially, that when the modified null hypothesis of no important difference between two means has been accepted we say that the EQUIVALENCE of the two population means has been established -- a much more stringent condition than simply finding a nonsignificant difference between them.

Bottom line: I gather we're in agreement that we should push hard to eliminate misconceptions about significance testing (especially the treatment of nonsignificance as proof of H_0), to encourage reporting of confidence intervals, and to provide some mechanism for generating uncensored data bases of studies testing given effects. Where we disagree is in whether abolishing (OK, recommending the abandonment of) significance tests and prohibiting (discouraging) ever drawing conclusions from single studies are the right approaches to these goals.

Ken Howard

Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Mon, 13 Nov 1995 11:02:06 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Giving up on significance testing
--

Lee Sechrest, Univ. of Arizona and Jim Wood, Univ. of Texas, El Paso

Confidence limits offer one alternative to conventional significance testing. Confidence limits can be used to accomplish significance testing but will provide additional information.

A more radical departure is the use of likelihood ratios. For information about likelihood ratios, we would recommend that you start with the following article, quite readable and informative.

Goodman, S.N., & Royall, R. (1988). Evidence and scientific research. *American Journal of Public Health*, 78, 1568-1574.

Likelihood ratio testing produces a statistic that reflects the relative likelihoods of two (or more) hypotheses given the observed data. Hence, it is Bayesian in form (the inverse probability as described by Gigerenzer). The likelihood ratio does not estimate the probability of any given hypothesis; rather it estimates the evidential value of observed data in relation to one or more hypotheses, the probative value of the evidence.

Put another way, a given bit of evidence may support any number of different hypotheses. Likelihood ratio testing is a way of evaluating the relative degree of support for alternative hypotheses that is afforded by particular evidence. As an example, suppose that "red spots" occur in 90% of measles cases. The odds of red spots are therefore 90:10, or simply "9" for measles. Suppose further that "red spots" occur in 60% of chicken pox cases. The odds of red spots are therefore 60:40, or "6" for chicken pox. The "likelihood ratio" of red spots would be $90:60 = 1.5$ favoring the measles hypothesis over the chicken pox hypothesis. By conventional standards a likelihood ratio of 1.5 would be considered at best weak support, and one can sense that, given that a child has some red spots, the evidence in favor of a diagnosis of measles rather than chicken pox would be slight. The actual diagnosis, however, would depend on other factors, most notably the base rate for the two diseases at the time of the observation of red spots.

Another example: A child is brought to an emergency room with a grid-like burn on his hand. According to a parent, the child burned himself accidentally on a barbecue. Suppose the emergency room physician has seen many such burns. She estimates that grid-like marks are found in about 5% of accidental hand burns, but in about 20% of hand burns that are the result of abuse. The likelihood ratio for the grid-like burn is therefore 4:1 -- indicating (according to conventions suggested by Goodman and Royall) not strong evidence of abuse. It is important to note that the 4:1 likelihood ratio does not indicate the probability that the child has been abused. To calculate that probability, we would need to know the "prior odds" of abuse among burn victims. What the likelihood ratio does tell us, however, is the "probative value" of the evidence. The use of odds likelihood ratios requires that two or more alternative hypotheses be exactly specified. In our usual approach to significance testing, we specify only one hypothesis exactly: the null. We specify that one alternative is that we will find *no* difference between groups; *any* alternative is equivalent to any other.

If we did likelihood ratios for psychotherapy outcome studies, we might end up with hypotheses like the following:

1. A difference between groups of $d = .2$ or less would not be interesting; therefore, adopt the hypothesis that the effect of psychotherapy is no greater than $.2$.
2. On the other hand, the therapy to be tested appears strong, and it is fairly expensive. Therefore a difference of $d = .5$ is not unreasonable and would probably have to be that large in order for the therapy to have any utility. Then we would test the therapy. Let us say we get a result of $d = .41$. We then calculate the likelihoods for the two alternatives. For example, a value of $.41$ might have a likelihood of $.08$ if the true effect of therapy were $.2$, and that same value might have a likelihood of $.24$, given an expected effect of $.5$ for therapy. The likelihood ratio would, therefore, be $.24:.08$ or 3:1. Thus the observed finding of $d = .41$ favors the hypothesis of $.5$ over $.2$ by a margin of 3 to 1. The evidence constituted by the observed $d = .41$ is better evidence for an effect of $.5$ than it is for an effect of $.2$.

For any given phenomenon, there is, obviously an infinite number of possible hypotheses, e.g., the effect of therapy is 2.00, 2.01, 2.12...4.68...5.06, etc. The choice of specific values to test would be arbitrary, but not capricious.

No single standard exists by which to judge the strength of likelihood differences, although in medicine likelihood ratios of less than 2 are generally regarded as too small, and ratios of as high as 4 or 5 are regarded as only fairly weak bases for choosing between hypotheses. (For mathematical reasons, likelihood ratios are usually expressed in terms of log likelihoods, a detail that is not unimportant but that is not directly germane to the present discussion.)

It will not escape anyone at this point that the use of likelihood ratios would change *markedly* our way of testing hypotheses. We would be forced to make point estimates or predictions, something Meehl has been urging on us

for years and that is undoubtedly overdue. That would be demanding, to say the least. But it would also, in our view, advance our science greatly. Surely it is time to be moving beyond the level of formulating hypotheses such as that the effect of therapy will be greater than nothing or that the accuracy of a diagnostic test will be greater than zero. We would come to rely much more heavily on extant literature in order to formulate our hypotheses and would also be more likely to use our results in order to refine our estimates.

Think about it. And read Goodman and Royall.

Date: Tue, 14 Nov 1995 13:43:01 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: signifying

Lee Sechrest, Univ. of Arizona

THIS IS IN RESPONSE TO A RECENT POSTING BY KEN HOWARD, ALTHOUGH I GATHER THAT THE TEXT WAS NOT ACTUALLY AUTHORED BY KEN. PROBABLY HE APPROVES OF IT, THOUGH. MY RESPONSES ARE IN UPPER-CASE (FORGIVE ME, BUT I LOST THE ORIGINAL DENOTATION OF THE TEXT)

Not that I agree with such a sharp distinction between individual studies and meta-analyses. A meta-analysis of all studies of feral children is still going to be based on less than a couple dozen observations of any given effect (presuming we could find a particular effect that had been measured in each case) and will provide a correspondingly imprecise estimate of that effect.

AGREED, A SMALL NUMBER OF CASES *MAY* YIELD IMPRECISE ESTIMATES. THAT DEPENDS ON THE HOMOGENEITY OF THE EFFECT (REMEMBER EARLIER DISCUSSIONS OF THE HIPPO PHENOMENON). IN ANY CASE, HOWEVER, A META-ANALYSIS EVEN ON A SMALL NUMBER OF CASES IS LIKELY TO YIELD A BETTER ESTIMATE THAN IS OBTAINABLE FROM ANY EVEN SMALLER NUMBER OF CASES, AS FROM A SINGLE STUDY.

The crucial variable (if we assume homogeneity of studies with respect to scores on important moderating variables) is simply total N -- as your paper hints at in treating the "single larger validity study with 1,428 subjects" as if its correlation of .22 were a population value.

NO, EVEN MORE IMPORTANT THAN EQUALITY ON MODERATING VARIABLES WILL BE THE PRECISION WITH WHICH THE STUDY IS CARRIED OUT. STUDIES SHOULD BE WEIGHTED BY SAMPLE SIZE

ONLY TO THE EXTENT THAT THEY ARE HOMOGENEOUS WITH RESPECT TO "QUALITY."

Of course, studies are not usually homogenous, so consistency across different studies conducted under very different conditions provides considerably greater confidence than a single study with the same total N that the effect is a robust one.

RIGHT. CAMPBELL REFERS TO "HETEROGENEITY OF IRRELEVANCIES," AS A DESCRIPTION OF THIS PHENOMENON.

I'm also pleased that you disavow one-tailed tests (along with all significance tests), though I would sure appreciate a note of caution to your readers that one-tailed tests are an especially abhorrent variant, lest readers as careless as I take

what appears to be the pejorative tone of your comments on the lower power of two-tailed tests (which is true only if your a priori hypothesis is correct; overall power depends on the likelihood that you're right, i.e. on the relative weight given the slightly higher power when right and the zero power when wrong about the population sign) as indicating the reverse preference between these two "evils".

ANY DISCUSSION OF THE RELATIVE MERITS OF OR JUSTIFICATION FOR ONE-TAILED VS. TWO-TAILED TESTS ASSUMES (A LA NEYMAN-PEARSON) THAT SOME MAGIC INHERES IN THE LEVEL OF ALPHA THAT IS CHOSEN. THAT IS, IF ONE'S SIGNIFICANCE TEST EXCEEDS $P = .05$, THEN ONE MAY CELEBRATE; IF IT IS ONLY $.10$, THEN ONE MUST, WHATEVER ONE'S COMMON SENSE AND LOGIC MIGHT SAY, GET A BEER TO CRY INTO. THE FISHERIAN VIEW OF SIGNIFICANCE, AND MAYBE THE HYBRID (SEE GIGERENZER, REFERRED TO PREVIOUSLY IN THIS STRING), WOULD JUST AS SOON SAY "SIGNIFICANT AT THE $.10$ LEVEL" AS "SIGNIFICANT AT THE $.05$ LEVEL BY A ONE-TAILED TEST." OTHERS WONDER WHAT THE PROBLEM IS IN THE FIRST PLACE.

While I'm spouting off anyway, let me suggest that your statement that significance level gives NO information about the probability of replication is too strong. You're certainly right that significance at the alpha level doesn't imply a replication probability of $1 - \alpha$. However, with N and sigma fixed (or N fixed and effect size expressed in standard-deviation units) the larger the population effect size, the lower the average p-value across a set of studies.

BE AWARE, THOUGH, THAT THE ASSUMPTION THAT SIGMA CAN BE "FIXED" IS MAJOR. SOME YEARS AGO BILL YEATON AND I LOOKED AT A NUMBER OF INSTANCES OF "FAILURE TO

REPLICATE," AND MORE OFTEN THAN NOT THE FAILURE WAS NOT IN REPLICATING DIFFERENCES BETWEEN MEANS BUT IN NOT REPLICATING INITIALLY SMALL ERROR VARIANCES. ONE OF THE REASONS THAT A STUDY MAY HAVE A VERY LOW P VALUE IS THAT IT ALSO HAS AN ANOMALOUSLY SMALL ERROR TERM.

Similarly, the larger the effect size, the higher the probability that two successive studies will yield statistical significance at a given alpha level in the same direction. Thus there will be a negative correlation between p-value and probability that an effect will replicate in the sense just described. In a paper that's in press in Psychophysiology (authored by Tony Greenwald, Rich Gonzalez, Don Guthrie, and me, though the result I'm about to focus on was mostly Don's work) a reasonable set of assumptions leads to the conclusion that a p-value of .005 (note the extra zero) is associated with about an 80% chance that an exact replication (same N, sampling from the same population, etc.) will yield statistical significance at the .05 level and in the same direction as the initial study. (The initial issue presented to the committee Tony oversaw was under what conditions an author reporting a result inconsistent with past results or with generally accepted theory should be required to replicate the finding before publication.)

WOW! SO THEN WHAT IS THE CHANCE THAT A P-VALUE OF .05 OR SO WILL BE SUPPORTED BY A FOLLOW-ON STUDY THAT MAY NOT HAVE THE SAME N, THE SAME SPECIFIC CONDITIONS, ETC.? THAT THE FOLLOW-ON STUDY WILL EVEN RESULT IN A DIFFERENCE WITH THE SAME SIGN!?!

Finally, to your list of false beliefs about the benefits of significance testing, I wish to append what I feel is a correct belief: While a highly significant (e.g., $p < .001$) result does not necessarily imply that the effect tested is large or important, it DOES indicate that you are considerably less likely to be wrong in your conclusion about the sign of the population effect than if your significance test had yielded a p-value of, say, .049. No conclusion based on variable data (whether single-study or meta-analytic) is error-free, but a researcher is justified in feeling less nervous about asserting that the population effect matches in sign that of the sample result if he has a highly significant result, rather than a barely significant one.

NOTE THE IMPLICIT ACCEPTANCE OF THE HYBRID MODEL OF SIGNIFICANCE TESTING THAT CANNOT QUITE DECIDE WHETHER CERTAIN VALUES OF ALPHA SHOULD BE SET IN ADVANCE AND CONSIDERED CRITICAL (NEYMAN-PEARSON) OR WHETHER P-VALUE SHOULD BE TREATED AS A VARIABLE.

Oh, well. One more. A number of participants in this discussion have suggested that our null hypothesis is best considered an hypothesis that the population effect size doesn't depart by more than a trivial (unimportant)

amount from the null-hypothesized value. I teach my students (as I'm sure lots of us do) that the one exception to the general rule that one never accepts H_0 (just fail to reject it) is that you may be able to accept the MODIFIED null hypothesis of no IMPORTANT departure from the hypothesized value (usually zero). The conditions for doing so are, first, that you specify and justify on practical or theoretical grounds just what a minimally important effect size is (i.e., what the boundary between unimportant and important effects is) and, second, that the CI (95%? 99%?) around your point estimate not include any values larger than this minimally important effect size. This is not an easy thing to do. Dana Quade (Harris & Quade, J. Ed. Stat., 1992) showed that the sample size needed to have a $1-\alpha$ chance that your $(1-\alpha)*100\%$ CI include no important values is at least 4 times as great as the N you would need to disprove the null hypothesis (with 50% power) if the population effect were just above the minimally-important threshold. And a large proportion of consultees are, in my experience, very reluctant to commit themselves to a particular effect size as minimally important.

I THINK THAT THE LATTER PROBLEM CARRIES WITH IT A FAIRLY PRACTICAL SOLUTION TO GETTING CONSULTEES TO COMMIT THEMSELVES TO A PARTICULAR EFFECT SIZE AS MINIMALLY IMPORTANT. TELL THEM THE SAMPLE SIZE THAT WILL BE REQUIRED (THEREFORE THE COST ENTAILED) TO DETECT EFFECTS AT DIFFERENT LEVELS. THEY WILL MOST LIKELY LOSE INTEREST IN SMALL EFFECTS. BUT IT IS ONE OF THE TASKS OF THE CONSULTANT TO HELP CONSULTEES (UGLY WORD!) THINK THROUGH THEIR EFFECT SIZE ISSUES SO THAT THEY UNDERSTAND THEM AND CAN MAKE INFORMED DECISIONS.

I believe that one of the recent notes cited a paper that recommends a highly similar procedure and suggests, essentially, that when the modified null hypothesis of no important difference between two means has been accepted we say that the EQUIVALENCE of the two population means has been established -- a much more stringent condition than simply finding a nonsignificant difference between them.

HOW ABOUT SAYING THAT "THAT THE TWO MEANS ARE ROUGHLY EQUIVALENT SEEMS ACCEPTABLE?" THE TERM "ESTABLISHED" IMPLIES RATHER MORE CONFIDENCE THAN I THINK WE OUGHT TO FEEL. I WOULD CERTAINLY AGREE, HOWEVER, THAT WHEN WE WISH TO "ACCEPT" THE NULL HYPOTHESIS, WE OUGHT TO HAVE MORE STRINGENT STANDARDS THAN SIMPLY THAT WE HAVE NOT BEEN ABLE TO ESTABLISH SIGNIFICANT DIFFERENCES.

Bottom line: I gather we're in agreement that we should push hard to eliminate misconceptions about significance testing (especially the treatment of nonsignificance as proof of H_0), to encourage reporting of confidence

intervals, and to provide some mechanism for generating uncensored data bases of studies testing given effects. Where we disagree is in whether abolishing (OK, recommending the abandonment of) significance tests and prohibiting (discouraging) ever drawing conclusions from single studies are the right approaches to these goals.

I AM GOING TO RESPOND TO THIS LAST PARAGRAPH IN A SEPARATE POSTING SINCE IT RAISES THE INTERESTING AND CHALLENGING QUESTION OF JUST WHAT WE CAN CONCLUDE FROM THE RESULTS OF A SINGLE STUDY.

Date: Tue, 14 Nov 1995 17:12:53 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Signifying To: sscpnet@bailey.psych.nwu.edu
--

Dear Lee:

The dialogue you are continuing with Ken Howard seems to focus a lot on the issue of the sample size, and always in the sense of the number of separate individuals who have been included in a single study or a group of studies under meta-analysis.

Does the logic of your position extend to the study of a large number of observations for a single individual? It seems to me that the study of a single individual, for example under various experimental conditions, with many observations per condition, has the possibility of greater precision than do group designs. Individuals (other than identical twins) differ in an unknown number of respects from each other in ways possibly relevant to a particular study. Even identical twins differ in possibly relevant ways in their histories of experience.

Don

Date: Wed, 15 Nov 1995 08:46:13 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Signifying To: DonaldK737@aol.com
--

>single subject multiple crossover designs seem useful to show there is someone that an intervention helps but there is no generalizeable effect size or guide to sample selection. i like it only for early phase two where there is little chance of carryover.

don klein <

Dear Don:

I agree. The down side of single subject studies is the lack of information about generalizability. Nevertheless, I sometimes think we should take the same attitude as astronomers or geologists, who after all have only the one universe or the one world to study. They can still find generalizations that apply in the particular domain.

Don

From: DonaldK737@aol.com Date: Wed, 15 Nov 1995 10:33:41 -0500 To: drouth@umiami.ir.miami.edu Subject: Re: Signifying
--

dear don,

you said to my note

> single subject multiple crossover designs seem useful to show there is someone that an intervention helps but there is no generalizeable effect size or guide to sample selection. i like it only for early phase two where there is little chance of carryover.

don klein <

Dear Don:

I agree. The down side of single subject studies is the lack of information about generalizability. Nevertheless, I sometimes think we should take the same attitude as astronomers or geologists, who after all have only the one universe or the one world to study. They can still find generalizations that apply in the particular domain.

I believe that when nonexperimental generalizations have stood the test of time ,they have always(eventually) been shown to be deducible from experimentally testable generalizations. for instance the lamarckian generalization (inheritance of acquired characteristics) about evolutionary patterns in its day was quite reasonable but doesnt fit the experimental genetic facts whereas darwinian variation and selection does (even though darwin allowed for the inheritance of acquired characteristics.).

In neither case is a direct experimental approach applicable but the experimentally testable context decides.

One can make the same case for ptolemy and copernicus. in the context of testable newtonian physics copernicus wins.

I have no problem with single case observations in the context of discovery, but i dont think they are an answer to the inductive problem.

cordially

don klein

Date: Wed, 15 Nov 1995 10:13:02 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Foundation Money To: sscpnet@bailey.psych.nwu.edu
--

This is in response to what Lee Sechrest just remarked about the small size of the grants (\$10K) offered by the American Suicide Foundation in relation to the difficulty of studying a problem such as youth suicide.

We should not forget that it is often the role of Foundations to provide "seed money" so that reseachers can do pilot studies. With pilot data in hand, the researchers can then come up with more convincing proposals to send in to the people with the deeper pockets, namely the federal government. Looked at in that light, the amounts of money Foundations can provide seem more reasonable.

Don

Date: Wed, 15 Nov 1995 11:04:59 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Signifying
To: DonaldK737@aol.com

dear don,
you said to my note ...
 >see above<

Dear Don K.:

There are really two issues here, best addressed separately. One concerns the use of single subjects in research. Single subject studies can be experimental as shown by Skinner and his followers and by psychopharmacologists (mostly working with animals) who have adapted Skinnerian methods to their own purposes. The historical example I had given earlier on the sscpnet was Ebbinghaus's five year long study with himself as the only subject--this discovered many new laws of learning and memory, and these in general were replicated by others. In some respects his findings were distorted by his methods. For example a lot of proactive inhibition is generated by learning so many lists of meaningless syllables.

The other issue is the use of experimental vs. nonexperimental methods. My own training was strictly experimental, and for a long time I did not feel right about it unless I was doing some kind of experiment. However, I spent six years sitting on the NIH Behavioral Medicine Study section, where among other things I learned to have a great deal of respect of epidemiology and its frequent use of large scale nonexperimental studies. As a journal editor, also, I have come to recognize that a large proportion of the good research in the field of child and adolescent psychopathology is nonexperimental in nature and will no doubt continue to be so. Our most influential role model for the use of epidemiological methods in the study of child and adolescent psychopathology is Michael Rutter.

Don R.

Date: Wed, 15 Nov 1995 09:46:11 -0800 (PST)
From: sylvia@netcom.com (Sylvia Caras)
Subject: Periodic Post: MADNESS List

MADNESS is an electronic action and information discussion list for people who experience moods swings, fright, voices, and visions (People Who). MADNESS creates an electronic forum and distribution device for exchanging ways to change political systems that touch People Who, and for distributing any information and resources that might be useful. A basic

premise of science and research is also a value of MADNESS: to share your findings with others.

MADNESS generates at least 40 messages a day.

For more information send LISTSERV@sjuvvm.stjohns.edu the command INDEX MADNESS.

To join, e mail LISTSERV@sjuvvm.stjohns.edu the message SUBSCRIBE MADNESS yourfirstname yourlastname

key words: mind mad crazy moods voices visions fear consumer survivor ex-patient rights advocacy

From: DonaldK737@aol.com Date: Wed, 15 Nov 1995 10:33:41 -0500 To: drouth@umiami.ir.miami.edu Subject: Re: Signifying
--

dear don,

you said to my note

>single subject multiple crossover designs seem useful to show there is someone that an intervention helps but there is no generalizeable effect size or guide to sample selection. i like it only for early phase two where there is little chance of carryover.

don klein

Dear Don:

I agree. The down side of single subject studies is the lack of information about generalizability. Nevertheless, I sometimes think we should take the same attitude as astronomers or geologists, who after all have only the one universe or the one world to study. They can still find generalizations that apply in the particular domain.

I believe that when nonexperimental generalizations have stood the test of time, they have always(eventually) been shown to be deducible from experimentally testable generalizations. for instance the lamarckian generalization (inheritance of acquired characteristics) about evolutionary patterns in its day was quite reasonable but doesnt fit the experimental genetic facts whereas darwinian variation and selection does (even though darwin allowed for the inheritance of acquired characteristics.).

In neither case is a direct experimental approach applicable but the experimentally testable context decides.

One can make the same case for ptolemy and copernicus. in the context of testable newtonian physics copernicus wins.

I have no problem with single case observations in the context of discovery, but i dont think they are an answer to the inductive problem.

cordially

don klein

Date: Wed, 15 Nov 1995 10:13:02 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Foundation Money To: sscpnet@bailey.psych.nwu.edu
--

This is in response to what Lee Sechrest just remarked about the small size of the grants (\$10K) offered by the American Suicide Foundation in relation to the difficulty of studying a problem such as youth suicide. We should not forget that it is often the role of Foundations to provide "seed money" so that reseachers can do pilot studies. With pilot data in hand, the researchers can then come up with more convincing proposals to send in to the people with the deeper pockets, namely the federal government. Looked at in that light, the amounts of money Foundations can provide seem more reasonable.

Don

Date: Wed, 15 Nov 1995 11:04:59 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Signifying To: DonaldK737@aol.com
--

dear don,

you said to my note
>single subject multiple crossover designs seem useful to show there is someone that an intervention helps but there is no generalizeable effect size or guide to sample selection. i like it only for early phase two where there is little chance of carryover.
don klein <

Dear Don:

>I agree. The down side of single subject studies is the lack of information about generalizability. Nevertheless, I sometimes think we should take the same attitude as astronomers or geologists, who after all have only the one universe or the one world to study. They can still find generalizations that apply in the particular domain. i believe that when nonexperimental generalizations have stood the test of time ,they have always(eventually) been shown to be deducible from experimentally testable generalizations. for instance the lamarckian generalization (inheritance of acquired characteristics) about evolutionary patterns in its day was quite reasonable but doesnt fit the experimental genetic facts whereas darwinian variation and selection does (even though darwin allowed for the inheritance of acquired characteristics.). in neither case is a direct experimental approach

applicable but the experimentally testable context decides. one can make the same case for ptolemy and copernicus. in the context of testable newtonian physics copernicus wins. i have no problem with single case observations in the context of discovery, but i dont think they are an answer to the inductive problem.
cordially <
don klein

Dear Don K.:

There are really two issues here, best addressed separately. One concerns the use of single subjects in research. Single subject studies can be experimental as shown by Skinner and his followers and by psychopharmacologists (mostly working with animals) who have adapted Skinnerian methods to their own purposes. The historical example I had given earlier on the sscpnet was Ebbinghaus's five year long study with himself as the only subject--this discovered many new laws of learning and memory, and these in general were replicated by others. In some respects his findings were distorted by his methods. For example a lot of proactive inhibition is generated by learning so many lists of meaningless syllables.

The other issue is the use of experimental vs. nonexperimental methods. My own training was strictly experimental, and for a long time I did not feel right about it unless I was doing some kind of experiment. However, I spent six years sitting on the NIH Behavioral Medicine Study section, where among other things I learned to have a great deal of respect of epidemiology and its frequent use of large scale nonexperimental studies. As a journal editor, also, I have come to recognize that a large proportion of the good research in the field of child and adolescent psychopathology is nonexperimental in nature and will no doubt continue to be so. Our most influential role model for the use of epidemiological methods in the study of child and adolescent psychopathology is Michael Rutter.

Don R.

Oben schon erwähnt ???

From: DonaldK737@aol.com Date: Wed, 15 Nov 1995 10:33:41 -0500 To: drouth@umiami.ir.miami.edu Subject: Re: Signifying
--

dear don,
you said to my note
see above
I believe that when nonexperimental generalizations have stood the test of time, they have always(eventually) been shown to be deducible from experimentally testable generalizations. for instance the lamarckian generalization (inheritance of acquired characteristics) about evolutionary patterns in its day was quite reasonable but doesnt fit the experimental genetic facts whereas darwinian variation and

selection does (even though darwin allowed for the inheritance of acquired characteristics.).

In neither case is a direct experimental approach applicable but the experimentally testable context decides.

One can make the same case for ptolemy and copernicus. in the context of testable newtonian physics copernicus wins.

I have no problem with single case observations in the context of discovery, but i dont think they are an answer to the inductive problem.

cordially
don klein

Date: Wed, 15 Nov 1995 10:13:02 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Foundation Money To: sscpnet@bailey.psych.nwu.edu
--

This is in response to what Lee Sechrest just remarked about the small size of the grants (\$10K) offered by the American Suicide Foundation in relation to the difficulty of studying a problem such as youth suicide.

We should not forget that it is often the role of Foundations to provide "seed money" so that reseachers can do pilot studies. With pilot data in hand, the researchers can then come up with more convincing proposals to send in to the people with the deeper pockets, namely the federal government. Looked at in that light, the amounts of money Foundations can provide seem more reasonable.

Don

Oben schon erwähnt ???

Date: Wed, 15 Nov 1995 11:04:59 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Signifying To: DonaldK737@aol.com
--

dear don,
you said to my note
see above
Dear Don:

I agree....see above

- > i believe that when nonexperimental generalizations have stood the test
- > of time ,they have always(eventually) been shown to be deducible
- > from experimentally testable generalizations.
- > for instance the lamarckian generalization (inheritance of
- > acquired characteristics) about evolutionary patterns in its day was
- > quite reasonable but doesnt fit the experimental genetic facts

> whereas darwinian variation and selection does (even though
 > darwin allowed for the inheritance of acquired characteristics.).
 >
 > in neither case is a direct experimental approach applicable but
 > the experimentally testable context decides.
 >
 > one can make the same case for ptolemy and copernicus. in the context
 > of testable newtonian physics copernicus wins.
 >
 > i have no problem with single case observations in the context
 > of discovery, but i dont think they are an answer to the inductive
 > problem.
 > cordially
 > don klein
 >
 Dear Don K.:

There are really two issues here, best addressed separately. One concerns the use of single subjects in research. Single subject studies can be experimental as shown by Skinner and his followers and by psychopharmacologists (mostly working with animals) who have adapted Skinnerian methods to their own purposes. The historical example I had given earlier on the sscpnet was Ebbinghaus's five year long study with himself as the only subject--this discovered many new laws of learning and memory, and these in general were replicated by others. In some respects his findings were distorted by his methods. For example a lot of proactive inhibition is generated by learning so many lists of meaningless syllables.

The other issue is the use of experimental vs. nonexperimental methods. My own training was strictly experimental, and for a long time I did not feel right about it unless I was doing some kind of experiment. However, I spent six years sitting on the NIH Behavioral Medicine Study section, where among other things I learned to have a great deal of respect of epidemiology and its frequent use of large scale nonexperimental studies. As a journal editor, also, I have come to recognize that a large proportion of the good research in the field of child and adolescent psychopathology is nonexperimental in nature and will no doubt continue to be so. Our most influential role model for the use of epidemiological methods in the study of child and adolescent psychopathology is Michael Rutter.

Don R.

From: "Paul D. Rokke" <rokke@plains.nodak.edu>
To: sscpnet@bailey.psych.nwu.edu
Date: Thu, 16 Nov 1995 11:53:45 +0000
Subject: (Fwd) Statistical Significance

-----Original message-----

Dear Netters:

I thought this might be of interest to you.

I have been on rmnet now for over a year. Although you have not heard from me often, I have been carefully reading many of the exchanges on the net. I have observed that many of the inquiries concern how to conduct statistical significance tests; e.g., how to test one's findings to see if they are statistically significant, what test to use, and how to compute it, etc. In the course of these observations, I have come to believe that many netters hold false beliefs about the value and benefits of significance tests in analyzing research data.

In fact, there are no legitimate benefits at all, and statistical significance tests should never be used. But I digress. The purpose here is to inform you that the APA Board of Scientific Affairs is looking into the question of whether use of significance testing should be discontinued. See below.

Board of Scientific Affairs Action on Significance Testing

By Frank Schmidt

In an article last year in the American Psychologist, Jacob Cohen (1994) urged that psychologists completely discontinue the use of statistical significance testing in analyzing research data and instead employ point estimates of population parameters and confidence intervals. In his Division 5 Presidential Address at the 1994 APA convention, Schmidt (in press) reached the same conclusion.

Albert Bartz, a psychologist who has authored several statistical texts, brought this issue to the attention of the APA Board of Scientific Affairs in March of 1995. He proposed that the Board appoint a Task Force to make recommendations as to how to implement the phasing out of statistical significance testing in course texts, journal articles, etc. The Board was provided with copies of the Cohen article, the Schmidt paper, and other materials.

At its November 3-5 meeting the Board took up Dr. Bartz' proposal. Board member Duncan Luce took the lead in laying out this issue for the Board. The Board approved in principle the notion of a Task Force to study this question and make recommendations. The Board also felt that the question was larger than APA; they felt that APS, Division 5, the Society for Mathematical Psychology, and other organizations should be given the opportunity to be involved. They also felt they should at least check out the potential involvement of other disciplines, such as statistics (through the American Statistical Association). The Board plans to bring this question up at a meeting of the Federation of Behavioral and Social Sciences (which includes Anthropology, Sociology, Economics, and other social sciences).

The Board appointed a committee of its members to study this question and make recommendations to be acted on at its March 1996 meeting, specifically:

1. What the plan for the Task Force should be.
2. What the budget for the Task Force should be.
3. Who should be on the Task Force.

The subcommittee will talk to a variety of people outside and inside APA before making its recommendations. The chair of the subcommittee is Duncan Luce.

Suzanne Wandersman, APA staffer to the Board, reported that the Board appeared to be very favorable to the idea of doing away with statistical significance testing. She thinks this effort will go forward and that there will be quite a bit of activity on it in 1996.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.

Schmidt, F.L. (in press). Statistical significance testing and cumulative knowledge psychology: Implications for the training of researchers. *Psychological Methods*.

Paul D. Rokke

Department of Psychology
North Dakota State University
Fargo, ND 58105-5075

Voice: 701/231-8626

Fax: 701/231-8426

Internet: rokke@plains.NoDak.edu

Date: Fri, 17 Nov 1995 09:00:03 -0700 (MST) From: James H Cavender <jhc@GAS.UUG.Arizona.EDU> To: John F Kihlstrom <john.kihlstrom@yale.edu> Subject: Re: Statistical vs. Clinical Significance

On Fri, 17 Nov 1995, John F Kihlstrom wrote:

I don't mean to start up another thread, but I remember a discussion, perhaps initiated by NJ, about the difference between statistically and clinically significant effects of therapy. I am now very interested in this issue, and I would be grateful for any references to published material on this issue. If people could send me some citations (privately, so as not to clog up the list) I'd be grateful.

Also, I remember some listmembers offering opinions as to what effect size might be construed as clinically as opposed to (merely) statistically significant. I'd be interested in those thoughts, as well.

To whom it may concern,

I echo Prof. Kihlstrom's request. As a fledgling researcher in clinical psychology I am very interested in the differences between statistical

and clinical significance, either applied or theoretical, on therapy. I would appreciate any informative citations on this issue sent my way.

James (Jim) Harrison Cavender
Student
University of Arizona
Department of Psychology
Tucson, AZ 85721
Phone: (520) 318-1769
(520) 621-7447
Fax: (520) 318-1769
e-mail: JamesC@aruba.ccit.arizona.edu
JHC@gas.uug.arizona.edu

Date: Mon, 20 Nov 1995 10:43:59 -0500 (EST) From: Samuel M Turner <turnersm@musc.edu> To: "Paul D. Rokke" <rokke@plains.nodak.edu> Subject: Re: (Fwd) Statistical Significance

Just a point of clarification on Dr. Rokke's comments regarding the BSA Task Force on Statistical Significance. As a member of BSA, it is my recollection that the task force would study the issue in the broadest of contexts. The task will not be to decide if we should continue to use significance testing but rather to elucidate the issues for the field. I believe I am correct in asserting that such a task force would not have the power to mandate anything in any event.

Samuel M. Turner

On Thu, 16 Nov 1995, Paul D. Rokke wrote:
see above

From: DonaldK737@aol.com Date: Thu, 16 Nov 1995 23:58:52 -0500 To: sechrest@U.ARIZONA.EDU Subject: Re: further signifying
--

dear all

lee says "Any single study has boundaries set by the specific conditions represented in it. Those conditions may include population and sample definitions, investigator and laboratory identities (some "findings" are stubbornly resistant to replication by other people or in other sites), specific treatment parameters, measures used, methods of analysis employed, and so on and so on. In a paper that deserves to be much better known, Tom Cook

has provided a comprehensive discussion of the conditions under which we may or may not be in a position to generalize about "cause," which, after all, is really what we are after.

Cook, T.D. (1990). The generalization of causal connections: multiple theories in search of clear practice. In L. Sechrest, E. Perrin, and J. Bunker (Eds.) Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Washington: D.C.: DHHS, Agency for Health Care Policy and Research, 9-31.

Also available as:

Cook, T.D. (1993). A quasi-sampling theory of the generalization of causal relationships. In L. Sechrest and A.G. Scott (Eds.) Understanding causes and generalizing about them. New Directions in Program Evaluation. San Francisco: Jossey-Bass, No. 57.

I come down generally on the side of Don Klein, who argues that replicability is the key to developing our knowledge in a field, at least if by replicability we mean constructive extension of our research and theory. "

lee represents my views quite exactly.

I would amplify that since correct random sampling of the population of interest never occurs we are stuck with the best we can do , a crass approximation by unsystematic replication. if that often falls within a theoretical framework (effect sizes are even crasser obfuscating approximations) then we can progressively narrow in to what popper calls verisimilitude.

This makes independent replications the key to advance, and devalues the single study (including a determination of the speed of light) as only useful pointers towards further work.

That our field leaps on the occasional "significant" finding as definitive betrays our nostalgia for certitude, which should have gone out with divine revelation.

cordially

don klein

Date: Fri, 17 Nov 1995 05:42:38 -0500 (EST) From: John F Kihlstrom <john.kihlstrom@yale.edu> To: SSCP Network <sscpnet@bailey.psych.nwu.edu> Subject: Statistical vs. Clinical Significance
--

I don't mean to start up another thread, but I remember a discussion, perhaps initiated by NJ, about the difference between statistically and clinically significant effects of therapy. I am now very interested in this issue, and I would be grateful for any references to published material on this issue. If people could send me some citations (privately, so as not to clog up the list) I'd be grateful.

Also, I remember some listmembers offering opinions as to what effect size might be construed as clinically as opposed to (merely) statistically significant. I'd be interested in those thoughts, as well.

John F. Kihlstrom, Professor
Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205
Telephone (203) 432-2596 Facsimile (203) 432-7172

Date: Fri, 17 Nov 1995 08:40:26 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Small N To: donaldek737@aol.com
--

Dear Don K.:

Oddly enough, I almost used Piaget as another example of an investigator who used very sample sizes to good effect. Perhaps some of his results did not replicate, but when people in the U.S. began to take his work seriously, an amazing number of them did replicate. It is hard to think of a student of cognitive development who made greater contributions. Piaget also made good use of what he called the clinical method. Of course, many others (notably Freud) abused this method, giving it a bad name.

While I am mentioning examples, another one I thought of was the field of psychophysics. S.S. Stevens is an example of a rigorous experimentalist in this area. He advanced the view that we should honor Fechner but repeal his law, substituting a power law. As I understand it, Stevens liked to run extensive studies of individuals one at a time, on the argument that psychophysical functions may vary, depending upon the peculiarities of a particular individual's sensory apparatus (and life history?).

Don R.

Date: Mon, 20 Nov 1995 08:14:03 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: Small N To: DonaldK737@aol.com

"Dear Don K.:
>Oddly enough, I almost ...< see above

dear don

i guess we disagree about piagets verisimilitude. You remember his remark when told that newborns imitatively protruded their tongue(in complete contradiction to his developmental schemas.).

" Very rude,indeed",quoth he.

If stevens studies individual A 10000 times and find that his log/log weber-fechner weight estimate ratio is $1.2 \pm .05$. Is stevens justified in proclaiming a power law as the modal model for individuals c,d,e.... or that A's power parameter estimate is a transindividual constant??

But since im not really interested in A per se, but want to make generalizations, hopefully on the basis of an inferred model, stevens is stimulating but not satisfying.

So the question boils down to whether intensive longitudinal small sample observation is a fruitful strategy for hypothesis formation. My read is sometimes yes but often no.

So lets distinguish hypothesis formation, which can spring from any set of observations or delusions from hypothesis testing. My small sample observation that imipramine blocked spontaneous panics was immediately countered by marks' report that his small sample observations were contradictory. It took experimentation to resolve the issue (except for marks).

This is hardly a brilliant, novel insight but its still central.

Freud did not give the clinical method a bad name. his 1900 colleagues (including the mythomaniac fliess) pointed out these deficiencies immediately. However clinical wishful thinking gave his arbitrary interpretive procedures an unwarranted good name.

cordially

don

Dear Don K.:

If Stevens studies individual A intensively and finds that a certain quantitative equation fits the psychophysical data from that person, that is an important first step. Whether this is discovery or confirmation depends on what Stevens thought beforehand. Now, obviously, the next thing is to run another subject, B. If the same equation fits, a bit of generalizability of the equation is established. If C, D, and E and so on also produce data that fit the equation, Stevens is on his way to a principle of some generality.

Now, suppose that the data for B (or any individual in the series) are a poor fit with the equation. Now the possibility arises that there are qualitative differences between individuals in psychophysical functions. For example, as is well known, there are several varieties of color blindness that would be associated with different reactions to visual displays. In this case one must search for ways to identify independently the groups that require different equations.

In this scenario, what might be the most misleading would be to get one data point from A, another from B, another from C, and so on, and compose the psychophysical function from all of these mixed in together. The result would perhaps be an equation that did not fit any particular individual, a sort of a bastard law that only fit a heterogeneous group.

At the most general level, what I am arguing is that if individuals are different with respect to certain phenomena, a proper study of the phenomena must take this into account.

Don

Date: Tue, 21 Nov 1995 10:53:32 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Re: "Clinical Significance," etc. To: John F Kihlstrom <john.kihlstrom@yale.edu>

Don,

I don't know how good it is, but I do think that the statistical vs. clinical significance problem is pretty interesting.

By the way, on your dissertation, I would have couched the finding in terms of longstanding debates about the cognitive capacities of infants -- how far back can you push their learning -- and also the equally longstanding debate about the role of reinforcement in language acquisition.

I bet your aunt would have understood.

John

John F. Kihlstrom, Professor
Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205 Telephone
(203) 432-2596 Facsimile (203) 432-7172

Dear John:

Yes, I was exaggerating a little bit there. The real point (in relation to the distinction among clinical, statistical, and scientific significance) was simply that the infant conditioning research did not appear to provide the infants with any benefits. That would have been clinically significant in Jacobson's terms as I understand them.

To provide another example that will be closer to home for SSCP Net participants, we might consider David Shakow's famous series of studies on the reaction time of schizophrenics. The research appeared to say something about attentional deficits in schizophrenia, but it was never argued that the subjects benefited in any way from participating in the enterprise. The whole category of experimental psychopathology, as well as much nonexperimental research on psychopathology, has no immediate clinical significance, and yet one hopes that its findings will be important to the clinician and to the consumer in the long run (or else why would we try to get taxpayers to support the activity?).

Don

Date: Tue, 21 Nov 1995 10:03:51 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: "Clinical Significance," etc.
To: John F Kihlstrom <john.kihlstrom@yale.edu>

It was I.

John F. Kihlstrom, Professor
Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205 Telephone
(203) 432-2596 Facsimile (203) 432-7172

Thanks. It is good to know that.

Don

Date: Tue, 21 Nov 1995 09:18:56 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: "Clinical Significance," etc.
To: sscpnet@bailey.psych.nwu.edu

During the last few days, someone on the SSCP Net requested some references on the topic of "clinical vs. statistical significance" and related matters. I know that Neil Jacobson is well known for his writings in this area and assume that it has been discussed various times before on this Net, perhaps to the extent that some had gotten tired of it. As a relatively new participant, I am not tired of it yet. Perhaps there are others out there who are not, either.

My own take on this issue is that clinical significance is a sort of proxy consumer judgment that a particular treatment was effective enough to be worth the trouble, cost, or risk it involved. The example that came to mind when I was thinking about this yesterday and today was one of the earliest ones, historically. In France J. Rodriguez Pereira taught some deaf children to speak and made his reputation thereby. People in Paris could judge the success of his work by their encounter with these successfully treated individuals, who were deaf but not mute.

Statistical significance in contrast refers to a research finding that appears to be replicable but is of no particular interest. For example, a manuscript I reviewed recently compared children of short stature who were receiving growth hormone therapy to the normative population of a popular child behavior checklist. There were some small differences on certain scales,

large enough to be associated with low "p" values in the analysis, but no theoretical rationale was presented for them, and as a reviewer I was unsure they were not due to some artifact or confound, e.g., some demographic difference between the short children and the normative group.

It seems to me that there should be a third category of differences that are of some scientific importance but not of any particular clinical significance. For example, my own doctoral dissertation many years ago demonstrated that one could condition young infants to make more vocalizations of one type or another (conditioning of response differentiation), a finding of some interest to students of conditioning or of infancy. It was not easy for me to explain to my Aunt Leila Katherine (or other non-psychologists) why my dissertation was of any importance. Yet the research was published in a respectable journal, elicited reprint requests from people like Lew Lipsitt, etc. If I had taught a deaf child to speak, my aunt would not have needed an elaborate explanation of why this was important.

Well, this comment is already long enough. Let's see if it elicits any response.

Don

Date: Tue, 21 Nov 1995 11:33:23 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: John F Kihlstrom <john.kihlstrom@yale.edu> Subject: Re: Statistical vs. Clinical Significance
--

Lee Sechrest, Univ. of Arizona

At the risk of clogging up the list, as John warned not to do, I am sending this out on the net because I think it is generally important and interesting.

On Fri, 17 Nov 1995, John F Kihlstrom wrote:

I don't mean to start up another thread, but I remember a discussion, perhaps initiated by NJ, about the difference between statistically and clinically significant effects of therapy. I am now very interested in this issue, and I would be grateful for any references to published material on this issue. If people could send me some citations (privately, so as not to clog up the list) I'd be grateful.

Also, I remember some listmembers offering opinions as to what effect size might be construed as clinically as opposed to (merely) statistically significant. I'd be interested in those thoughts, as well.

John,

This is in response to your inquiry about clinically important effect sizes. I assume that you will have heard, or will hear from, N. Jacobson or someone who is familiar with his ideas and calculations. Jacobson's idea

about assessing clinically "significant" change by demonstrating that people are moved into a "normal" range with respect to some outcome measure or other, although Ciarlo may not ever have coupled the idea of moving people to within two standard deviations of "normal" with the requirement that it also be demonstrated that they show "significant" change. I do not have a reference to Ciarlo handy, but you should not have much trouble retrieving it, and he is still working away in Colorado. What follows are some other ideas.

Some years back, a few behaviorists wrote about, and some even did something about, the idea of assessing therapeutic outcomes in terms of "normalizing" behavior, an idea similar to Jacobson's, but preceding it. The behaviorists called their idea "social validity," and you can learn all about it from your colleague Al Kazdin. Basically, though, the idea involved in social validity was that people (patients, clients, etc.) who are noticeably different from "normal" persons before treatment should not be so after treatment. Or at least the difference should be markedly reduced. The idea of social validity never became widespread, but I was impressed by it and remember one study in particular that showed that socially deviant adolescent females who were clearly distinguishable from "normal" adolescent females (if there is such a thing) before treatment for communications problems were not distinguishable afterwards (Fawcett and Miller, 1975).

At about that time, Bill Yeaton and I were considering the idea of the "just noticeable difference" in functioning, and we did write about it (Sechrest and Yeaton, 1981), and I messed around with the idea (i.e., I never actually produced any real data). Dan Ozer did learn about our idea, however, and he did collect some data. For a variety of reasons, Dan's work remained as a file drawer problem for quite some time until it was finally published in 1993 (Ozer, 1993).

I regard Dan's paper as a landmark. Without trying to summarize it here, I will just say that his paper suggests that across several measures of "personality" people must differ by about two standard deviations in order for the difference to be noticeable by raters who have been observing them in interpersonal settings for a period of several (about 8) hours. Put in other terms, two people would have to be about as different as those at the 32nd and 68th percentile on some personality measure in order for it to be obvious to people who did not know them until having the opportunity to observe them for a day.

More recently, and somewhat inspired to get back to the problem by Dan's paper, I and one of my students, Pat McKnight, have been working on the general problem that I call "calibrating" measures, by which I mean learning how to translate the empty metrics of typical psychological measures (e.g., 8 points on a depression scale) into "behavioral implications." Among behavioral implications of interest are "just noticeable differences." Think of it this way: how different in introversion-extraversion would two job candidates have to be in order for 75% of interviewers to make the correct judgment based on a one-half hour interview? The idea of behavioral implications is broader than just noticeable differences, but what we are

interested in is what differences in behavior can be expected (be implied by) on the basis of given differences on psychological measures, i.e., tests.

Without going into the details of our work too much, Pat McKnight and I concluded very much as did Dan Ozer, that a difference of about two standard deviations on a psychological measure is required in order to have much in the way of behavioral implications. At least behavioral implications at all readily

noticed. Differences of about a half-standard deviation on intelligence tests may be noticeable over a long period of time in an individual's life. For example, about a half of one standard deviation is associated with the difference between average high school and average college graduates. And about one standard deviation is associated with the difference between the average college graduate and the average Ph.D. (shockingly small!)

Most people are surprised at the large difference on psychological measures that seems to be required for important behavioral implications, but the limited validity of psychological measures and regression toward the mean are sufficient to explain the phenomena. If (generously) the validity of our measures is about .50, then that means that regression toward the mean is about half-way. Consequently, two standard deviations of observed difference on some measure translates into about one standard deviation on the underlying behavior. Thus, two individuals who are two standard deviations apart (32nd and 68th percentiles) on some personality measure will be only about one standard deviation apart on an observable behavior (e.g., about from the 41st to the 59th percentile). That is not a lot of difference between people as we usually think of it. Obviously, we have not extended such thinking to the extremes of differences, where percentile differences would not be large in relation to standard deviation units. But we are not usually greatly interested in differences at the extremes. We would not be greatly impressed at a therapeutic intervention that moved a person from four standard deviations out on potential for violence to only three standard deviations out, i.e., from maybe the 99.999th percentile to the 99.99th percentile.

Pat and I have a version of our calibration paper (given a couple of times at meetings) written up and would be willing (not necessarily glad) to send you a copy if you are interested to see it.

Fawcett, S.B., Miller, L.K. (1975). Training public speaking behavior: an experimental analysis and social validation. *Journal of Applied Behavior Analysis*, 8, 125-135.

Ozer, D.J. (1993). Classical psychophysics and the assessment of agreement and accuracy in judgments of personality. *Journal of Personality*, 61, 739-767.

Sechrest, L., and Yeaton, W.H. (1981). Assessing the effectiveness of social programs: Methodological and conceptual issues. In S. Ball (Ed.), *New Directions in Evaluation Research*. San Francisco: Jossey-Bass, No. 9, 41-56.

As a post-script, Larry Stricker of ETS and I have talked about the j.n.d. approach, and it is my understanding that he is trying it out on the English proficiency test used with foreign students. He wants to determine how much of a difference in the test score is required in order to produce a j.n.d. in classroom performance (ratings, I think).

Date: Tue, 21 Nov 1995 11:40:42 -0800 (PST) From: Neil Jacobson <njacob@u.washington.edu> To: Lee B Sechrest <sechrest@u.Arizona.EDU> Subject: Re: Statistical vs. Clinical Significance

Dear Lee, John, and colleagues,

Our original ideas about clinical significance were published in Behavior Therapy in an article by Jacobson, Follette, and Revenstorf in 1984. A couple of critical comments were made by Larry Christensen at Texas A & M and Bruce Wampold at University of Utah, published in 1996, and we published our response in Behavior Therapy in 1986. Then I edited a special issue of Behavioral Assessment in 1988 devoted to the subject in 1988, including an article with Dirk Revenstorf. In that issue, many interesting commentaries appeared. Finally, the most recent version of our methods was published in an article by Jacobson & Truax, JCCP, 1991.

We have gone back and reanalyzed data from previous outcome studies, using our clinical significance statistics. To do so requires raw data, which investigators have been kind enough to send us. This led to some new publications, most of which led to pessimistic conclusions about the efficacy of psychotherapy, when examined under the microscope of clinical significance. Essentially, for every disorder which we have examined, and regardless of the modality of therapy, a minority of couples "normalize" during the course of therapy. Most clients do not get what they came for, in my view. However, they do improve a bit, enough for statistically significant advantages relative to a control group, and in some cases relative to a placebo. But, typically, an empirically validated treatment means nothing more than that a treatment works better than nothing, but is not that impressive.

A couple of recent exceptions: the treatment of panic, developed by people like David Clark, Tim Beck, and Dave Barlow; and CBT for bulimia, the Fairburn version.

Finally, Andy Christensen and I published the Psych Science article on the overwhelming evidence that neither professional training nor amount of clinical experience influence outcome. I put all of this material together for a magazine, the Family Therapy Networker, written for clinicians.

I am happy to send anyone who is interested the "Clinical significance packet". Just ask for it by sending an e-mail to my assistant, Joan Giacomini: joang@u.washington.edu

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Tue, 21 Nov 1995, Lee B Sechrest wrote:

Lee Sechrest, Univ. of Arizona

At the risk of clogging up the list, as John warned not to do, I am sending this out on the net because I think it is generally important and interesting.

On Fri, 17 Nov 1995, John F Kihlstrom wrote:

>I don't mean to start up another thread, but ...< see above

>John,

This is in response to your inquiry ...< see above

Date: Tue, 21 Nov 1995 18:06:28 -0800 (PST) From: Bill Follette <follette@unr.edu> To: Neil Jacobson <njacob@u.washington.edu> Subject: Re: Statistical vs. Clinical Significance
--

Neil-

There are a couple of other references on the topic to point out:

Speer published a piece in JCCP on clin sig which is referenced in the Truax piece.

Also, Mike Lambert and colleagues have a piece coming out in the SPR journal Psychotherapy Research where I and Glenn Callaghan have responded and (I think) Ken Howard.

Happy holidays,
Bill Follette

Date: Sat, 25 Nov 1995 22:29:09 -0800 (PST) From: Gerald Rosen <groser@u.washington.edu> To: Neil Jacobson <njacob@u.washington.edu>
--

Subject: Re: Statistical vs. Clinical Significance

This is a bit tardy but I had vague memories of earlier references on this issue and went back to a Psych Bull review of self-help therapies Russ Glasgow and I did in 1978. Sure enough in that paper we cite an article by J.R. Lick, "Statistical vs. clinical significance in research on the outcome of psychotherapy." International Journal of Mental Health, 1973, 2, 26-37 and I wonder if that is the first article on the topic of interest a few weeks ago.
-Jerry

Gerald Rosen, Ph.D.
#1910 Cabrini Medical Tower
901 Boren Avenue
Seattle, WA98104

Tel: (206) 343-9474
Fax: (206) 682-5859
e-mail: grosen@u.washington.edu

Date: Sun, 26 Nov 1995 10:42:19 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: Gerald Rosen <grosen@u.washington.edu>
Subject: Re: Statistical vs. Clinical Significance

I am glad you reminded me of John Lick's article. I had forgotten all about it. He was a Northwestern doctoral student (Dick Bootzin PhD).
Lee Sechrest

Date: Sun, 26 Nov 1995 11:46:29 -0700 (MST)
From: Lee B Sechrest <sechrest@u.Arizona.EDU>
To: sscpnet <sscpnet@bailey.psych.nwu.edu>
Subject: More on less significance testing

Lee Sechrest, Univ. of Arizona

Two more references that may be of use, particularly if you get involved in arguments with backward journal editors or referees.

Loftus, G.R. (1991). On the tyranny of hypothesis testing in the social sciences. Review of G. Gigerenzer, et al. "The empire of chance: how probability changed science and everyday life. Cambridge, England: Cambridge Univ. Press, 1989. Contemporary Psychology, 36, 102-105.

G. Loftus, as most of you know, is a statistician and author of a statistics text, so he writes with some authority. The Gigerenzer book is well worth reading in its own right.

Loftus, G. (1993). Data analysis: a picture is worth a thousand p values. *Memory and Cognition*, 21, p. 3.

The reference is to a fragment of the Editorial Comment with which Loftus began his editorship of *Memory and Cognition*. Let me quote briefly from it:

1. By default, data should be conveyed as figure depicting sample means with associated standard errors and/or, where appropriate, standard deviations.
2. More often than not, inspection of such a figure will immediately obviate the necessity of any hypothesis-testing procedures. In such situations, presentation of the usual hypothesis-testing information (F values, p values, etc.) will be discouraged.

There you have it.

Date: Mon, 27 Nov 1995 12:57:46 -0700 (MST) From: Lee B Sechrest <sechrest@u.Arizona.EDU> To: sscpnet <sscpnet@bailey.psych.nwu.edu> Subject: Still more on non significance-testing

Lee Sechrest, Univ. of Arizona

Here is another comment from a completely unrelated field, traffic safety, written by an engineer, and a dozen years old. I quote from the abstract: Sensible management of traffic safety is predicated on having reasonable expectations about the effect of various safety countermeasures. It is the role of evaluative research to derive such intelligence from empirical data. In spite of decades of research and experience, the safety effect of many countermeasures remains unknown. This sorry state of affairs is largely due to the objective difficulty of conducting conclusive experiments. Recognition of this objective difficulty should lead to the realization that in transport safety, knowledge is accumulated gradually from small, noisy and diverse experiments. The statistical tools used to extract knowledge from data should reflect this aspect of reality. One must therefore question the usefulness of classical tests of significance as a device for scientific progress in this field. It is argued that the unquestioning and all-pervasive use of significance testing in evaluative research on transport safety amounts to a self-inflicted learning disability. In contrast, it is shown that the classical Point Estimation, Likelihood Support and Bayesian methods can all make good use of experimental evidence which comes in small doses. In particular the likelihood function is an efficient device for the accumulation of objective information and a necessary ingredient for Bayesian decision analysis. Woe! Aid and comfort to the enemy. I was only going to use the last part of the abstract having to do with significance testing since I know how Ken Howard will now bash me--and some of you, too--with the first part. But I

was overcome with a strange impulse, honesty apparently, and felt compelled to provide the entire abstract.

Hauer, E. (1983). Reflections on methods of statistical inference in research on the effect of safety countermeasures. *Accident Analysis and Prevention*, 15, 275-285.

Date: Tue, 28 Nov 1995 07:40:29 -0800 (PST) From: Gerald Rosen <groten@u.washington.edu> To: sscpnet@bailey.psych.nwu.edu Subject: Re: Statistical vs. Clinical Significance (fwd)

For some reason this message got to the individuals but not the maillist. I'm sending it out again for general info.

-Jerry

Date: Sat, 25 Nov 1995 22:29:09 -0800 (PST)
From: Gerald Rosen <groser@u.washington.edu>
To: Neil Jacobson <njacob@u.washington.edu>
Subject: Re: Statistical vs. Clinical Significance

This is a bit tardy but I had vague memories of earlier references on this issue and went back to a Psych Bull review of self-help therapies Russ Glasgow and I did in 1978. Sure enough in that paper we cite an article by J.R. Lick, "Statistical vs. clinical significance in research on the outcome of psychotherapy." International Journal of Mental Health, 1973, 2, 26-37 and I wonder if that is the first article on the topic of interest a few weeks ago.

-Jerry

Gerald Rosen, Ph.D.
#1910 Cabrini Medical Tower
901 Boren Avenue
Seattle, WA 98104

Tel: (206) 343-9474
Fax: (206) 682-5859
e-mail: groser@u.washington.edu

Date: Tue, 19 Dec 1995 09:15:01 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Clinical Significance Again
To: sscpnet@bailey.psych.nwu.edu

A number of days ago a new discussion of clinical significance got started on the SSCP net. At one point Neil Jacobson came on the net with an offer to send people a packet of materials on the topic. All of us who were engaged in the discussion sort of stopped in our tracks. At my request, Jacobson's lab was kind enough to send me the packet. Last night I leafed through the reprints there (I had read some of the articles before, but not all). First of all I was impressed all over again with the solid contribution Jacobson and his colleagues had made to the issue of how to evaluate the results of therapy research.

Naturally, I transposed Jacobson's concepts to an area of treatment I know more about. Consider, therefore, the clinical significance of efforts to deal with the serious problem of mental retardation. Three examples of such intervention came to mind: prevention of PKU by a low phenylalanine diet, early intervention with demographically high risk children via enriched developmental daycare (e.g. Ramey's Abecedarian project), and prevention of the undue deterioration of mental functioning in children with Down syndrome by keeping them at home rather than institutionalizing them.

Compared to treatment research on depression or panic disorder, these examples are all a bit peculiar. First of all, mental development in children is a dynamic process, so we typically use standard scores relative to an age-appropriate normative group (e.g. IQ). There are relatively good norms for the general population, which facilitates the use of some of Jacobson's

concepts. What we don't usually have is normative data for the clinical population of persons with MR taken as a whole. We know that the vast majority of them fall within the range of mild MR, so for purposes of discussion I will assume a mean IQ of about 60 and a SD of 15. Second, the interventions do not usually raise IQ. Instead, they prevent it from going down (or going down so much).

The PKU case would fit Jacobson's concept of clinical significance (and certainly, my own). We have a group of youngsters who in the absence of treatment might end up microcephalic with IQs of around 35, and after treatment they approximate 100 (or at worst borderline, say 85). Now that is therapy!

The Abecedarian case does not fit Jacobson's concept, but I would argue that it is nonetheless clinically significant in my sense (a proxy consumer judgment of whether the treatment was worth the effort involved). The youngsters in the control group ended up with IQs of about 85, and those in the intervention group with IQs of about 100.

The DS case also does not fit Jacobson's concept but does fit mine. At best, an adult with DS is going to be at least mildly mentally retarded. Such children start out looking very normal in infancy and go down hill from then on (some have thus argued that it should be considered a dementing condition, even aside from the likelihood that adult DS individuals are at high risk for Alzheimer's disease). In the old days, doctors told parents routinely to institutionalize their DS babies, and the outcomes of institutionalized persons with DS were in the range of moderate to severe MR (rather than mild).

At least, now we can perhaps continue the discussion of clinical significance better informed, courtesy of Jacobson's efficient assistant.

Don

Date: Tue, 19 Dec 1995 16:08:58 -0500 Message-Id: <0d72f5a0@warren.med.harvard.edu> Subject: Re: Clinical Significance Again To: sscpnet@bailey.psych.nwu.edu,

Don Routh just restarted the discussion of clinical significance. By coincidence, I got my packet of stuff from Neal's lab just as I was reading Marty Seligman's American Psychologist piece on the CR study. While I concur that the work on "clinical significance" is very important, especially in sharpening the discussion, I still feel that it is frequently too strict a criterion, and feel that the name is a rhetorical trick that confuses the waters. Don's examples from retardation are excellent ones where return to "normal" may not be in the cards, but enormous progress may be. Clinically significant in my book. In the CR study, many people felt that they had improved "quite a bit". Isn't this closer to the lay concept of "clinically significant" than "return to the normal distribution"? Even improving "a fair amount" may be clinically

significant. Shouldn't the strict interpreters at least provide empirical evidence of what therapy patients want, rather than simple interpretations based on their "clinical experience"?

Cheers,
Stephen

Stephen Soldz
Director of Research and Evaluation
Advocates for Human Potential
Sudbury, MA
531 Beech St.
Director of Research
Boston Institute for Psychotherapy
ssoldz@warren.med.harvard.edu
[Sorry folks, I respond in multiple roles!]

Department of Psychiatry
Harvard Medical School

Roslindale, MA02131
(617) 469-3576

Date: Tue, 19 Dec 1995 16:03:24 -0800 (PST) From: Neil Jacobson <njacob@u.washington.edu> To: ssoldz@warren.med.harvard.edu Subject: Re: Clinical Significance Again

Dear all,

I never meant to equate clinical significance with "return to normal functioning". I simply used "return to normal functioning" as an example of a meaningful criterion that is not measured by either a p value or an effect size. In fact, there are numerous alternatives in the literature, many of which make sense, considering the population. I say as much in the "Networker" article. However, consumers who enter therapy don't care much about statistically significant improvement. I think that the general expectation that clients have when they enter therapy is that they leave therapy without the problem they came in with. That means "returning to normal functioning". There are exceptions, like some forms of mental retardation, schizophrenia, or autism. But "returning to normal functioning" is not an uncommon expectation, especially among the worried well. In that regard, it is interesting that in the Consumer Reports study, only 54% thought therapy helped alot. Imagine how much lower that percentage would have been if dysfunctional people who hadn't benefited from therapy had responded. The survey obviously pulls for those who "felt good" about the experience.

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105

Phone: (206) 543-9871
Fax: (206) 685-3293

Date: Tue, 19 Dec 1995 18:58:04 -0800 (PST) From: Bill Follette <follette@unr.edu> To: ssoldz@warren.med.harvard.edu Subject: Re: Clinical Significance Again
--

Once again, I'll try to prop up Jacobson's career by coming to his defense. There is no requirement one responds more like someone who doesn't have a clinical problem than one who does after an intervention. It just seems like that is one important piece of information one would want to know when evaluating both what to expect, whether treatment was worth the cost, and how to compare alternatives. Unless all one is selling is nonspecific expectations of getting some benefit from treatment, there is no harm informing consumers of this information. If one had terminal cancer with no possibility of returning to a disease free state, people may well wish to make choices between lesser options

that they value other than complete return to health.

I have heard all this before and in an upcoming article have suggested (reluctantly) that data be presented as we suggested in Jacobson, Follette, and Revendorf (1984), but the addition of location on a third distribution which represents the data from the current best available alternative treatment that has been subjected to empirical testing. That way people can judge their results with the best of what's been found to work and the gold standard (return to like I didn't have the problem). That third best available treatment (BAT) distribution will change, but it solves your and Ken Howard's and others objections that wanting to return to the normal distribution is too strict. I must say, there ought to be no problem telling people what we are and are able to do when they seek to spend money for services. In your original post asked if people really expected to get back to the way they were before the problem occurred (my paraphrase) and that we should collect data on that and not assume it. Go ahead. But what if people said they didn't expect that much out of treatment. Would you stop research? If one treatment produced better results than people seemed satisfied with, would you not recommend it over one that was "just satisfying" *ceteris paribus*. If you fall and break your wrist, don't you ask yourself if not the doctor "Am I gonna be able to play the piano again (implicitly as well as before)?"

People who see the value of clinical significance aren't insisting life be a bed of roses, only that we keep track of where we are and where we're going. If I had cancer and I was choosing between 5 yrs or 10 yrs of survival, I'd choose 10. I'd want a full life expectancy. I'd take the best I could get, but I'd want full life expectancy. Wouldn't you? Plus I have always loathed the days when doctors used to withhold that information from patients because the patient couldn't handle it, or more likely the doctor couldn't admit they couldn't help

(yet). Nothing in reporting clinical significance data undermines hope and confidence in a particular instance. I hope my comments didn't miss the point of your and earlier posts (some of which I missed).

Happy holidays to all,
Bill Follette

Date: Wed, 20 Dec 1995 03:18:24 -0500 Message-Id: <951220031823_95164824@mail02.mail.aol.com> To: sscpnet@bailey.psych.nwu.edu Subject: Re: Clinical Significance Again
--

In a message dated 95-12-19 19:06:02 EST, njacob@u.washington.edu (Neil Jacobson) writes:

>In that regard, it is interesting that in the Consumer Reports study, only 54% thought therapy helped alot. Imagine how much lower that percentage would have been if dysfunctional people who hadn't benefited from therapy had responded. The survey obviously pulls for those who "felt good" about the experience.<

Yes that may be true, but a cynic might say that the ones who benefitted greatly might take it all for granted and the ones who hadn't might be more likely to respond in order to complain.

John Roraback
Southpark Psychology
Moline, IL

From: seligman@CATTELL20.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: Clinical Significance and CR To: njacob@u.washington.edu (Neil Jacobson) Date: Wed, 20 Dec 1995 08:55:11 -0500 (EST)

Neil Jacobson wrote:

>In that regard, it is interesting that in the Consumer ...< see above

I have to respond from memory since I am away from the original data: Let's get the facts straight first: On a 5 point scale: made things a lot better, better things somewhat better, no change, made things somewhat worse, made things a lot worse. 57% of respondents said therapy "made things a lot better" and another 34% said therapy made things "somewhat better." So 91%--this was almost as true of people who started in the "very poor"--I could barely

manage" category, thought therapy helped. I am not exactly sure whether return to normal or to within 2sd's of normal should be criteria of clinical significance, but I suggest that 91% of the CR respondents may have been helped to a "clinically significant" level.

How many CR respondents did not respond because they were "dysfunctional" is unknown. "Imagining", as Neil wants us to do, won't help decide this. We need a prospective CR-type study that has no dropouts to find out. My guess is "quite a number" from the AA category and "not many" from the other disorders.

Neil alleges that the CR study OBVIOUSLY pulls for people who felt good about therapy. "Obviously" is a large overstatement. That is Neil's speculation only. There was no internal evidence that this was so, including evidence from a validation study. But more important, Neil should tell us why CR's studies of insurance companies, airlines, luggage, and used cars do not "pull for people who liked the product", but somehow the therapy study does.

Again, the argument is ad hoc and only a prospective study with no drop outs will answer it.

What does need an answer--urgently-- is that Neil has claimed that the effects of psychotherapy are very small and clinically of marginal or no significance. If memory serves, his data base is short term efficacy studies with a manual and fixed number of sessions. Why are Neil's estimates of clinical significance so small, and CR's so large (even 57% clinical significance is, I assume, much bigger than the estimate Neil has been touting). Let's do the right study to find out.

marty seligman

Date: Mon, 18 Dec 1995 10:29:28 -0500 To: sscpnet@bailey.psych.nwu.edu From: mes3@po.cwru.edu (Milton E. Strauss) Subject: low base rates and reliability
--

On December 17, 1995 David A. Smith wrote:

>While we're on the topic of interobserver agreement, can anyone help me with this issue? Imagine 4 judges of 6 targets (a la Shrout and Fleiss sample data matrix). All four judges assign exactly the same number to each of the targets (e.g., 24 "0s" are assigned -- the matrix is all 0s -- for instance, all 6 targets are asymptomatic). <

Dave Smith
Ohio State

>This seems to be an extreme variant of the "low base rate" problem (or "high base rate" problem, logically the same) described by Meehl and Rosen years ago. Why observe at all, when you can be 100% correct by simply calling all cases asymptomatic?<

Don

Dons comment brought to mind that Dominic Cicchetti has written several papers on the issue of reliability when base rates are low (some with specific emphasis on kappa). One is in Journal of Clin and Exp Neuropsychology, 1991, 12, 328-338 [aren't computerized reprint files handy!].here are a couple of others, but I lent them to a student before entering them [...but only if one is systematic!]. The limiting case that David noted is an extreme, but low base rate disorders are common and the solution is not simple as Cicchetti describes.

Milton

Milton E. Strauss, Ph.D.
Department of Psychology
Case Western Reserve University
Cleveland, OH 44106-7123
Phone: (216) 368-2695
Fax: (216) 368-3537
E-mail: mes3@po.cwru.edu

Date: Mon, 18 Dec 1995 09:30:01 -0500 (EST) From: Barbara Baumann <bauman+@pitt.edu> Subject: Re: ICC = Shrout & Fleiss 1979 To: dasmith+@osu.edu
--

On the subject of ICC calculations, macros with formulas for all 6 Shrout and Fleiss ICCs are available from SPSS. I have been using these recently with good success. Someone simply wrote a program to create a matrix specific to each of Shrout and Fleiss' assumptions. They are useful when there is a concern or hypothesis regarding magnitude of ratings in addition to ranking. I would be happy to provide these macros to anyone interested.

Barbara Baumann
Clinical/developmental psychology grad student
University of Pittsburgh

On Fri, 15 Dec 1995 dasmith+@osu.edu wrote:

>To make sure you are using the correct computations and values for Shrout & Fleiss ICCs, you might want to run their data through your stats package first. Here is a SPSS program to illustrate. It shows two separate ways to get the same anova table (viz. "reliability" and "manova"). You have to collect the appropriate MSs from the output and hand calculate the specific ICC you are after.
As Howard Berenbaum notes, there are at least 6 ICCs, so it is always necessary to be clear which you use (e.g., ICC(2,1) Shrout & Fleiss, 1979). That is why stats packages do not just give you "the ICC" -- there are many. In S&F, they range from .17 to .91 on the same data!<

Dave Smith
Ohio State

*****Shrout & Fleiss (1979). ICC. PsyBull, 86, 420-428

*****SPSS/PC verification program

data list free / target judge1 judge2 judge3 judge4.

begin data.

1 09 2 5 8

2 06 1 3 2

3 08 4 6 8

4 07 1 2 6

5 10 5 6 9

6 06 2 4 7

end data.

reliability variables judge1 to judge4
/statistics all.

manova judge1 to judge4
/wsfactors judge(4)
/wsdesign.

ANALYSIS OF VARIANCE

SOURCE OF VARIATION/ SUM OF SQ /DF/ MEAN SQUARE/
F/PROB.

BETWEEN PEOPLE	56.2083	5	11.2417	
WITHIN PEOPLE	112.7500	18	6.2639	
BETWEEN MEASURES	97.4583	3	32.4861	31.866
000				
RESIDUAL	15.2917	15	1.0194	
NONADDITIVITY	2661	1	2661	248
626				
BALANCE	15.0256	14	1.0733	
TOTAL	168.9583	26	7.3460	

Date: Mon, 18 Dec 1995 08:58:12 -0500
 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
 Subject: Re: Reliability Analyses
 To: dasmith+@osu.edu (David A. Smith, Ph.D.)

On December 17, 1995 David A. Smith wrote:

While we're on the topic of interobserver agreement, can anyone help me with this issue? Imagine 4 judges of 6 targets (a la Shrout and Fleiss sample data matrix). All four judges assign exactly the same number to each of the targets (e.g., 24 "0s" are assigned -- the matrix is all 0s -- for instance, all 6 targets are asymptomatic). In a sense, there is perfect agreement; if percent agreement were computed it would be 100%. However, all ICCs will be 0.0. Perhaps one could say that there is perfect agreement and zero reliability. Clearly, with no variance there can be no covariance, and with no covariance no reliability. However, I want to know how much variance is "enough" to warrant using reliability statistics (ICC) over agreement statistics (% agreement, kappa) for the purpose of establishing the "quality" of the data. Might weighted kappa be adapted as a multi-purpose solution? Any takers? I'll be happy to summarize private replies for the list to accomodate you shy lurkers. Thanks. Happy Holidays!
 Dave Smith
 Ohio State

This seems to be an extreme variant of the "low base rate" problem (or "high base rate" problem, logically the same) described by Meehl and Rosen years ago. Why observe at all, when you can be 100% correct by simply calling all cases asymptomatic?

Don

Date: Wed, 20 Dec 1995 06:59:33 -0800 (PST)
From: Neil Jacobson <njacob@u.washington.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL20.psych.upenn.edu>
Subject: Re: Clinical Significance and CR

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Wed, 20 Dec 1995, Martin E. P. Seligman wrote:

Neil Jacobson wrote
cut

In that regard, it is interesting that in the Consumer Reports study, only 54% thought therapy helped alot. Imagine how much lower that percentage would have been if dysfunctional people who hadn't benefited from therapy had responded. The survey obviously pulls forthose who "felt good" about the experience. I have to respond from memory since I am away from the original data: Let's get the facts straight first: On a 5 point scale: made things a lot better, better things somewhat better, no change, made things somewhat worse, made things a lot worse. 57% of respondents said therapy "made things a lot better"

It is 54%. I don't find this very impressive. First, the sample itself is a worried well sample, as you admit in the AP article. Second, who wouldn't say that therapy has helped "somewhat"? Especially among a worried well sample? Third, 46% said that therapy did not help alot, even among a worried well sample. Fourth, is it not logical that the response bias would pull for those who did well, if for no other reason than the fact that people who were not helped, and therefore are still dysfunctional, are less inclined to complete surveys of any kind? Fifth, most of CR is devoted to prospective field trials, not surveys. One cannot equate a "road test", which is prospective, and uses measures which are objective, to a retrospective survey, where only 4,000 therapy seekers responded out of 186,000 people who were surveyed. Sixth, having seen the questionnaire used in the survey, it is clear that the one question pertaining to psychiatric symptoms, directly adjacent to the question on more generic consumer satisfaction, shares so much method variance that it can be safely interpreted as simply another way of assessing consumer satisfaction. Seventh, I would guess that 54% of worried well folks do feel good about their experiences in therapy. What does this have to tell us about the efficacy of psychotherapy for moderate to serious problems? Not much. What do these 54% feel good about. I think Jerome Frank said it best.

and another 34% said therapy made things
"somewhat better." So 91%--this was almost as true of people who started in the
"very poor"--I could barely manage" category, thought therapy helped. I am not

exactly sure whether return to normal or to within 2sd's of normal should be criteria of clinical significance, but I suggest that 91% of the CR respondents may have been helped to a "clinically significant" level.
How many CR respondents did not respond because they were "dysfunctional" is unknown.

If people stopped assuming that the CR study had anything to say about the efficacy of psychotherapy, I would stop speculating about the nonresponders.

"Imagining", as neil wants us to do, won't help decide this. We need a prospective CR-type study that has no dropouts to find out. My guess is "quite a number" from the AA category and "not many" from the other disorders.

The AA data are quite damning. They are so obviously distorted that they call into question all of the data. They show how easily it is for a retrospective survey such as this to create a biased sieve which selects for true believers.

Neil alleges that the CR study OBVIOUSLY pulls for people who felt good about therapy. "Obviously" is a large overstatement. That is Neil's speculation only. There was no internal evidence that this was so, including evidence from a validation study. But more important, Neil should tell us why CR's studies of insurance companies, airlines, luggage, and used cars do not "pull for people who liked the product", but somehow the therapy study does.

This is a sleight of hand comparison, as I have pointed out above. The other studies you refer to are field trials, prospective, with objective measures of outcome. CR does not rely primarily on surveys.

Again, the argument is ad hoc and only a prospective study with no drop outs will answer it. What does need an answer--urgently-- is that Neil has claimed that the effects of psychotherapy are very small and clinically of marginal or no significance.

I do not believe that psychotherapy is of little value. My guess is that it has considerable value, especially for particular types of problems (like panic disorder). I also stated in my "Networker" article that the process of being in therapy is an important end in itself for many people, especially the worried well, and that purchase of friendship should not be underestimated. The CR study actually suggests to me that this "process" effect is weaker than I thought it would be.

If memory serves, his data base is short term efficacy studies with a manual and fixed number of sessions. Why are Neil's estimates of clinical significance so small, and CR's so large (even 57% clinical significance is, I assume, much bigger than the estimate Neil has been touting). Let's do the right study to find out.

marty seligman

??? ????? ???? oben ?

From: seligman@CATTELL20.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: Clinical Significance and CR continued To: njacob@u.washington.edu (Neil Jacobson) Date: Wed, 20 Dec 1995 11:46:53 -0500 (EST)

Neil says

cut

It is 54%. I don't find this very impressive. First, the sample itself is a worried well sample, as you admit in the AP article.

Marty replies:

I nowhere say that the CR sample is a "worried well" sample. I don't even know what this derogatory term means. It is not part of science or my vocabulary. As I recall about 40% of the sample reported that they could "barely manage" when they went for help. Is that "worried well?" Again a prospective survey with diagnostic procedures would decide whether what Neil asserts as true, rather than hypothetical, is true or false

Neil says:

Fifth, most of CR is devoted to prospective field trials, not surveys. One cannot equate a "road test", which is prospective, and uses measures which are objective, to a retrospective survey, where only 4,000 therapy seekers responded out of 186,000 people who were surveyed.

Marty replies:

Again, CR's rating of used car reliability, of insurance brokers, of doctors, and lawyers (forthcoming) is based on just the same survey method with similar return rates. In none of these, if memory serves, do anything like 91% of clients say the service made things the equivalent of "a lot better" or "somewhat better." Neil needs to show us why as some sort of sampling artifact, psychotherapy does so well and every other service tested by CR with the same methods does worse. The argument, as Neil makes it, is simply ad hoc.

Neil says:

The AA data are quite damning. They are so obviously distorted that they call into question all of the data. They show how easily it is for a retrospective survey such as this to create a biased sieve which selects for true believers.

Marty replies:

Again what is "obvious" to Neil is not "obvious" to me, but rather an hypothesis in need of test. An alternative hypothesis, which I am not inclined to, is that AA does very well in real life with CR type clients.

Neil says:

This is a sleight of hand comparison, as I have pointed out above. The other studies you refer to are field trials, prospective, with objective measures of outcome. CR does not rely primarily on surveys.

Marty replies:

Neil is wrong here. For the professions cited above, CR relies primarily on surveys and those services don't come off as well as therapy.

marty seligman

At any rate, have a good vacation Neil.

Date: Wed, 20 Dec 1995 09:07:27 -0800 (PST) From: Neil Jacobson <njacob@u.washington.edu> To: "Martin E. P. Seligman" <seligman@CATTELL20.psych.upenn.edu> Subject: Re: Clinical Significance and CR continued

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Wed, 20 Dec 1995, Martin E. P. Seligman wrote:

Neil says

It is 54%. I don't find this very impressive....

Marty replies:

I nowhere say that the CR sample is a...
>see above<

Both in the CR article and in the forthcoming AP article (Seligman, in press), although the term "worried well" is not used, it is explicitly stated that this was not primarily a sample of people with serious mental disorders, but rather a typical sample of adults seeking outpatient therapy, and there is a clear implication that we are not talking about the "difficult to treat". By the way, I don't find the term "worried well" derogatory. It describes most people I

know, at least those who have been in therapy. Some of my best friends are worried well.

Neil says:

Fifth, most of CR is devoted...

Marty replies:

Again, CR's rating of used car reliability...

>see above<

Take a look at any issue of CR. Count the number of articles devoted to consumer satisfaction surveys, vs. prospective tests with performance measures. I actually don't know what the outcome of such a study would be. My guess is that the "meat" of CR is in the prospective performance tests, not in the surveys. That is why I used to subscribe to it. But Marty is right, it is an empirical question. But this 91% figure combines those who said therapy helped "alot" and "somewhat". I don't know anyone who wouldn't say that they have been in therapy sometime and it didn't help at all. But I know alot of people who have been in therapy with the same problems now that they had then. None of the measures on the CR questionnaire tap into measures of their "problems". The items alleged to do so are simply additional self-report items, which undoubtedly are highly correlated, and loading on the same factor, as the consumer satisfaction ratings. The shared method variance guarantees this. The kind of research Marty calls for has been done, by Ken Howard, among others. He finds that there is NO correlation between self-report ratings of "helpfulness" and psychiatric symptoms or functioning in the world. And these data are prospective, using psychometrically sound measures. The CR study is retrograde compared to already existing methodologies for studying psychotherapy naturalistically.

Neil says:

The AA data are quite damning....

Marty replies:

Again what is "obvious" to...

Neil says:

This is a sleight of ...

Marty replies:

Neil is wrong here. For the professions cited above, CR relies primarily on surveys and those services don't come off as well as therapy.

marty seligman

At any rate, have a good vacation Neil.

>see above<

From: seligman@CATTELL20.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: Clinical Significance and CR continued

To: BTHYER@UGA.CC.UGA.EDU (Bruce Thyer)
Date: Wed, 20 Dec 1995 13:51:11 -0500 (EST)

People have no trouble reporting that they hate their cars, lawyers, doctors, airlines, and vacations in CR. Many of these involve lots of expense and lots of time. The striking fact is that psychotherapy seems to fare much better than all these services and products--in spite of roughly equal money and time.

When the CR data become publically available, we will surely analyze for the cost-time X outcome results. At any rate, I don't think cognitive dissonance explains the pattern of data very well.

marty

Date: Wed, 20 Dec 1995 14:55:39 -0400 (EDT)
From: "Al Lang" <allang@darwin.psy.fsu.edu>
To: seligman@CATTELL20.psych.upenn.edu, sscpnet@bailey.psych.nwu.edu
Subject: Re: Clinical Significance and CR continued

Marty:

I'm reluctant to get into this CR/psychotherapy effectiveness thing again as I've already spent too much time on this net lately, but let me reiterate a derivation of my query on the topic originally raised a month or so ago: How much confidence would you have in the effectiveness of herbal remedies if CR had evaluated them by asking the subset of their readers who had used them to any significant degree to report on their experiences by responding to questions analogous to those they asked about psychotherapy?

Al

Alan R. Lang, Ph.D.	phone:(904)644-6065
Psychology Department	fax: (904)644-0790
Florida State University	
Tallahassee, FL 32306-1051	allang@psy.fsu.edu

From: ssoldz@warren.med.harvard.edu
Date: Wed, 20 Dec 1995 15:31:18 -0500
Subject: Re[2]: Clinical Significance Again
To: Bill Follette <follette@unr.edu>

Bill,

I agree that "clinical significance" is a useful way to present data on what effects are likely to occur from TX. I certainly agree that some alternative to statistical significance would be an improvement to treatment planning, and, when we have good data, public policy. My biggest concern is linguistic and ideological. The term "clinical significance" is very loaded and easily subject to misinterpretation. It also is easily to say that therapy does not attain "clinical significance" in a large number of cases and is, therefore, of limited value. I think that some of the writings by Neil suggest this; for example, his family Therapy Networker piece on therapy being oversold. (BTW, I think that its overall a great piece and intend to distribute it to many students and colleagues in order to stimulate discussion on this overselling problem. It may not seem like it from my contributions here, but I take a lot of heat from clinical colleagues on precisely this issue.)

To some degree, I think that our differences come from having different modal patients in mind. Most of my research and practice involve chronic conditions in people who may never have been "normal" to begin with, but want to do better than they are. Personality disorders and substance abusers, for example. In substance abuse, most treatment programs don't expect that a single treatment episode will lead even to permanent abstinence, much less, "normality" in the broader psychosocial realm. If this is accepted (don't oversell treatment), then the question is, what are the expected effects, both of single treatment episodes, and of a succession of treatment episodes. Even of rigid cost-benefit grounds, substance abuse treatment might never reach "clinical significance" and yet be extremely valuable (and valued by) both clients and society. A reduction in substance use could (and often is) associated with a reduction in associated crime, medical problems (cirrosis, AIDS, etc.) and costs, improved parenting (which may affect the next generation), clients spending more time working and less on public assistance, etc. All this is possible without the treatment ever attaining "clinical significance", yet it surely seems clinically significant to me (and to the Center for Substance Abuse Treatment and NIDA, who fund such programs without expecting permanent abstinence as a necessary criterion of success).

So how about using a less loaded term than "clinical significance"?

As to the idea of comparing a treatment to the best available alternative, its an interesting one, which I'm in priniple sympathetic too. I'm sure we'd disagree on what data is relevant -- target problems, vs. associated psychosocial functioning. Also, as to the quality of the existent data. However, I'm very interested in ideas like this to communicate data to patients. My clinical colleagues like to assert the sanctity of the therapist-patient relationship for treatment decision making. Leaving aside the complex issues of "public" funding, I usually point out that patients usually don't have the data to make informed decisions, as they only know what this therapist does, not what the alternatives are, or how these alternatives would affect them (a form of David Rubin's missing data).

I hope this long post at least clarifies some of our differences.

Cheers,
Stephen

Stephen Soldz

Director of Research and Evaluation
Advocates for Human Potential
Sudbury, MA
531 Beech St.

Department of Psychiatry
Harvard Medical School

Director of Research
Boston Institute for Psychotherapy
ssoldz@warren.med.harvard.edu

Roslindale, MA 02131
(617) 469-3576

[Sorry folks, I respond in multiple roles!]

Date: Wed, 20 Dec 1995 17:34:20 -0500 From: "Donald K. Routh" <drouth@umiami.ir.miami.edu> Subject: Clinical Significance Again To: sscpnet@bailey.psych.nwu.edu

Maybe we could adopt some eponym such as Jacobson Type A, B, or C significance for the valuable concepts Neil has introduced and return the term "clinical significance" to its traditional colloquial use, which it seems several of us still consider to be worthwhile.

Don

Date: Wed, 20 Dec 1995 15:02:37 -0800 (PST)
From: Neil Jacobson <njacob@u.washington.edu>
To: ssoldz@warren.med.harvard.edu
Subject: Re: Re[2]: Clinical Significance Again

Like I said yesterday and in the "Networker" article, I do not equate clinical significance with "returning to normal functioning". Never did. Never will. I agree that it would be a mistake.

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Wed, 20 Dec 1995 ssoldz@warren.med.harvard.edu wrote:

> Bill,
> I agree that "clinical significance" ...<
>see above<

Date: Wed, 20 Dec 1995 15:05:20 -0800 (PST)
From: Neil Jacobson <njacob@u.washington.edu>
To: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Clinical Significance Again

Don,

Perhaps instead we should just use the term "Jacobson" instead of cure. "I saw my shrink today, and he Jacobsoned me". I like it. I have always been jealous of people who develop assessment instruments and get to use their names.

Happy Holidays!

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Wed, 20 Dec 1995, Donald K. Routh wrote:

> Maybe we could adopt some eponym such as Jacobson Type A, B, or C
significance for the valuable concepts Neil has introduced and return the term
"clinical significance" to its traditional colloquial use, which it seems several of us
still consider to be worthwhile.<
Don

From: seligman@CATTELL20.psych.upenn.edu (Martin E. P. Seligman) Subject: Re: Clinical Significance and CR continued To: allang@darwin.psy.fsu.edu Date: Wed, 20 Dec 1995 19:46:58 -0500 (EST)

Of course: If herbal remedies elicited the same pattern of responses to the same set of questions, I would take them almost as seriously as I do psychotherapy. As a Bayesian, I would still take such data more seriously for an endeavor with a higher a priori credibility--TX--than ginseng. The last time CR wrote as glowing an article on a product as they did about TX, I bought it--the Lexus. I have methodological reservations about both the TX article, and about the Lexus article, n=1, but none of these reservations come close to explaining the pattern of data away.

marty

Marty:

I'm reluctant to get into this CR/psychotherapy effectiveness thing again as I've already spent too much time on this net lately, but let me reiterate a derivation of my query on the topic originally raised a month or so ago:
How much confidence would you have in the effectiveness of herbal remedies if CR had evaluated them by asking the subset of their readers who had used them to any significant degree to report on their experiences by responding to questions analogous to those they asked about psychotherapy?

Al

Alan R. Lang, Ph.D.
Psychology Department
Florida State University
Tallahassee, FL 32306-1051

phone: (904)644-6065
fax: (904)644-0790
allang@psy.fsu.edu

Date: Thu, 21 Dec 1995 10:12:50 -0400 (EDT)
From: "Al Lang" <allang@darwin.psy.fsu.edu>
To: seligman@CATTELL20.psych.upenn.edu, sscpnet@bailey.psych.nwu.edu
Subject: Re: Clinical Significance and CR continued

Marty:

I think you better hope CR never does such a survey on herbal medicine or you may find yourself drinking a lot of ginseng. I can just see the true believers stuffing the ballot box, and still wonder if that is not, in part, what happened with the psychotherapy study. As for the Lexus study, I'd wager that the quality of Lexuses is a lot more uniform than that of psychotherapists and they did do some objective tests on the car. Enough of this...best wishes for the holidays.

Al

In message Wed, 20 Dec 1995 19:46:58 -0500 (EST),
seligman@psych.upenn.edu (Martin E. P. Seligman) writes:

>Of course: If herbal remedies elicited ...
>see above<

Date: Thu, 21 Dec 1995 10:24:27 -0800 (PST)
From: Bill Follette <follette@unr.edu>
To: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: Re: Clinical Significance Again

How about if we don't.

Bill Follette

On Wed, 20 Dec 1995, Donald K. Routh wrote:

> Maybe we could adopt some eponym such as Jacobson Type A, B, or C
significance for the valuable concepts Neil has introduced and return the term
"clinical significance" to its traditional colloquial use, which it seems several of us
still consider to be worthwhile.<

Don

Date: Thu, 21 Dec 1995 10:23:24 -0800 (PST)
From: Bill Follette <follette@unr.edu>
To: ssoldz@warren.med.harvard.edu
Subject: Re: Re[2]: Clinical Significance Again

Stephen-

Thanks for the comments. Lexically, the term "clinical significance" was offered as a contrast to "statistical significance." We were asking that people make such a distinction when reporting data in scientific journals. Implicitly, whether they use those same terms when talking to clients is not important, but there is a qualitative difference between treatment outcomes that should somehow be communicated if such differences exist. We wouldn't presume to say you must present treatment by saying "Hey, treatment won't make any clinically significant difference in your life, but what the heck." I'm being silly with the example, but I do think researchers, clinicians, and clients all need to be informed about what we can and cannot do. It does not detract from that fact that less than "clinically significant changes" can still "matter." I feel a little proprietary about the c.s. term since it has been used for over a decade. Lambert et al have a piece coming out that agrees with your position. That's the piece I responded to suggesting the Best Available Alternative rather than give up the C.S. term which has obvious social validity.

I enjoy the discussion.

WCF

Date: Thu, 21 Dec 1995 08:18:45 -0500
From: "Donald K. Routh" <drouth@umiami.ir.miami.edu>
Subject: I have been Jacobsoned
To: Neil Jacobson <njacob@u.washington.edu>

Dear Neil:

Yes, that is the kind of thing I had in mind. You could be right up there with Binet and Wechsler. Your acceptance has Jacobsoned me of my guilt over disputing your concepts of clinical significance. Of course, my message also had a sly purpose, sort of like the article S.S. Stevens once wrote "to honor Fechner and repeal his Law." Happy Holidays to you, too.

Don

On December 20, 1995, Neil Jacobson replied:

Don,

> Perhaps instead we should just use the term "Jacobson" instead of cure. "I
saw my shrink today, and he Jacobsoned me". I like it. I have always been jealous
of people who develop assessment instruments and get to use their names. <
Happy Holidays!

Neil S. Jacobson
University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

On Wed, 20 Dec 1995, Donald K. Routh wrote:

> Maybe we could adopt some eponym such as Jacobson Type A, B, or C
significance for the valuable concepts Neil has introduced and return the term
"clinical significance" to its traditional colloquial use, which it seems several of us
still consider to be worthwhile. <

Don

Letters
from
SSCP net

No 9

ed
horst u kaechele

ulm- london august 96
for private use only

Date: Mon, 15 Jul 1996 11:55:47 -0400
From: PhilF101@aol.com
To: seligman@CATTELL.PSYCH.UPENN.EDU
cc: njacob@U.WASHINGTON.EDU, **ssepnet@bailey.psych.nwu.edu**
Subject: Seligman/Consumer Report (CR) Surprises

Marty and Neil,

My comments on the Consumer Reports(CR) Study are:

*Long term therapy worked much better than short term therapy.

This finding surely came as a surprise to me and is definitely not consistent with my own clinical work and clinical research. In fact as I have stated before my clinical research with outpatients indicates it typically takes from 12 to 16 weeks to move from the clinically distressed range using a variety of measures including stress symptom measures, the Friedman Well-Being and Belief Scales, the PANAX affect scales etc. These measures were tracked session by session.

I very, very rarely see a client for 2 years (long term therapy in the CR study) and when I do I generally consider myself relatively ineffective. I do though have a very good relationship with a client by that time. Although there may be a dose response curve operating in my clinical population in therapy most of the gains clearly come in the early sessions. Also my average client was 33 years old with an individual, income of \$31, 000 and joint income of \$54,000. The CR clients were older (about 44 I believe) and I assume wealthier though I haven't seen statistics on this.

*No modality of therapy or drug worked better than any other for any problem.

Reading the Lambert reviews of the outcome literature seems to indicate that behavioral or cognitive-behavioral therapy is better for certain problems.

*Marriage counselors and family doctors did significantly worse than other mental health specialists

This surprised me somewhat. Were the marriage counselors well-trained in marital and family therapy and as well-trained as the other therapists I would ask.? Another related question. When I ask clients who they saw previously for therapy and what their professional degree was a large percentage of them have no idea what the professional background or degree was of their previous therapist. How come the CR clients were so knowledgeable of this information?

*Insurance limitations were associated with worse outcomes.

This finding is not immediately obvious though I am pleased at the result. (irrelevant to the finding however)

*Group therapy worked as well as individual therapy

I am somewhat surprised this held up across all problems

*The sheer size of the benefits was considerably larger than I expected.

Could you clarify what you mean here?

I think it will take a more powerful effectiveness design (total population, blind diagnosis, better reporting of modality, longitudinal) to test whether these findings are reliable, but a good case can be made that the findings told us quite a bit that surprised us.

Yes but how reliable and repeatable are they. As I have indicated I am skeptical of some of them from my professional experience. Nevertheless it was a great effort and a great stimulus for further research in these areas.
Best Regards,

Phil

P.S. I have discussed some of my reactions to the CR study with you recently in person at the PSCP event in Philadelphia but if you have any reactions to my comments I would appreciate hearing them.

Philip H. Friedman, Ph.D

Exec.: Director: Foundation for Well-Being
P.O. Box 627
Plymouth Meeting, Pa. 19462
Assistant Professor: Hahnemann University
and Medical College of Pennsylvania
Author: Friedman Well-Being, Belief and Quality of Life Scales
PhilF101@aol.com
610-828-4674
"The World is As You See It"
"Forgiveness is the Key to Happiness"

Date: Fri, 19 Jul 1996 15:13:50 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: clinical@cattell.psych.upenn.edu, Subject: Qualifications of Therapists and Outcomes
--

Neil Jacobson claimed that the CR study "confirmed what we already knew" that increasing levels of experience, skill, and education do not make for better therapy. The CR study did nothing of the sort, nor do "we already know it" from elsewhere. This is a mischievous claim in today's marketplace. Profit-driven health care schemes will send patients to the cheapest provider they can find-- particularly if scholarship can be invoked to justify the idea that less qualified providers (less experience and less education) make for just as good outcomes.

The CR dataset was conflicting on this issue. Social workers (presumably mainly MSW's, although degree was not queried) did as well as doctoral level providers. Marriage counselors (presumably mainly MFC's, MFT's, and other nondoctoral providers--although degree was not queried), on the other hand, did significantly worse. This was not an artifact of the fact that marriage counselors see couples in troubled relationships. CR compared doctoral level providers treating marital and sexual problems to marriage counselors treating the same problems, and the doctoral level providers did significantly better. Here are some relevant data: The marriage counselors (n=197) seeing respondents who checked "marital or sexual problems" had significantly poorer global outcome than other therapists (n=753) with these patients ($p<.0017$, controlling for initial severity).

The argument for equivalence derives from Christensen and Jacobson (1993) and their argument is seriously flawed. For they review studies, some badly flawed, in which manuals are used, mild and uncomplicated clinical problems are the diagnosis (but the diagnosis is made by doctoral level providers), and duration of therapy is brief and fixed.

Precisely the situations in which clinical judgment, experience, and education matter very little. A seven- year-old may be able to fly a one-engine plane in clear weather, but this does not mean she can handle a 747 in a thunderstorm.

The disinterested conclusion should be:

- 1) If a case is simple, if a manual must be followed, if the diagnosis is made by a more experienced and better trained clinician, and if treatment must be very brief, less experienced and less well-educated providers may do as well as doctoral level specialists.
 - 2) It seems likely that in real therapy, where cases are complicated and more severe, no manuals are used, diagnosis as well as therapy must be done, and clinical judgment is important, that more education and more experience of providers will improve outcomes.
 - 3) Effectiveness studies of level of education and experience of providers and the cost-benefit analysis are urgently needed. Instead MCO's use this state of affairs to justify using less experienced and less well-trained providers even in complicated and severe cases.
- Again, I believe, patients are being deprived of adequately skilled treatment on a massive scale. Until this issue is resolved by the appropriate effectiveness study, I recommend that MCO's err on the conservative side and provide experienced and highly educated providers in all but the simplest and least severe cases.

Date: Sun, 21 Jul 1996 11:36:13 -0700 (PDT)
From: Neil Jacobson <njacob@u.washington.edu>
To: section3 <:sscpnet@bailey.psych.nwu.edu>
Subject: anyone want a pre-print debunking the CR study?

Marty Seligman has commented that my remarks regarding the CR study affirming what we already know were "mischievous". As an example, he chose my observation that we already knew that social workers do as well as psychiatrists and psychologists. We do know that. His counter-argument may not be mischievous (I don't know his intent), but it is a sleight of hand, and an interesting example of how data can be used as rhetorical devices, in support of one's a priori position. He picks once again on those poor marriage counselors to show that the CR study really does provide evidence that training makes a difference.

Here is why I find this such an interesting counter-argument. The assumption is that marriage counselors are either less trained or less experienced than the other professions. Yet, clients often call anyone who does therapy for couple problems a marriage counselor, regardless of profession. It is a term often used pejoratively, when therapy is not working. How do we know what the profession of the marriage counselor was? Not many people are certified as marriage counselors, and so most of these counselors were either psychiatrists, psychologists, social workers, or certified counselors (a more generic sort of certification). The fact that they did more poorly is irrelevant to my argument. The fact that social workers did as well as Ph.D's and M.D.'s is directly relevant to my argument, and proves my point.

I also find it interesting that, while the lawyers of CR refuse to release the data, little tidbits get leaked when such data become necessary to refute a criticism of the study. I suspect we will continue to see such selective leakage in the October issue of the American Psychologists, where critiques of the CR study are to be published.

My major point was that the CR study supports the lack of a relationship between level of training and outcome. Seligman assumes that I am basing my conclusions that no such relationship exists on the Christensen and Jacobson (1993) article in Psych Science. He finds major flaws in this article, because, unlike the CR study, this review includes studies with methodological problems. I wonder how many of the studies he has read. Has he read the Strupp and Hadley (1979) study, for example, where psychodynamic therapists

with 25 years of experience on the average failed to outperform college professors in non mental health areas, even though they were both untrained and unsupervised? The studies in that review vary considerably in their methodological quality, but the bottom line is this: in the meta-analysis reported by Smith, Glass, and Miller, the correlation between training and outcome was .00!

However, I do believe that, despite these null findings, Larry Beutler is correct when he points out that the acquisition of specific skills does enhance outcome in certain treatments for certain problems.

Finally, I find it troubling that people keep confusing substantive questions with methodological ones. Effectiveness studies ask different questions than do efficacy studies, but that doesn't mean that the answers require different methodologies. While clinical trials can't answer all questions, they can answer all of the ones posed by Seligman in his American Psychologist article as suggestive of the limitations of experimental methods. If any of you are interested, Andy Christensen and I have contributed an article to be published by American Psychologist in October, which demonstrates: a) that the CR study tells us nothing that we didn't already know; b) that the new information provided is uninterpretable due to fatal methodological flaws in that study; and c) that despite their limitations clinical trials can elegantly answer all of the questions posed by Seligman, in a much more convincing way than surveys such as that published in CR. I would be happy to send out reprints

Neil S. Jacobson

Date: Sun, 21 Jul 1996 19:16:03 -0400 (EDT) From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: sscpnet@bailey.psych.nwu.edu Subject: Heads I win, Tails don't count
--

Neil seems to be making a "Heads I win, Tails don't count" argument. He has contended in numerous places that higher level of training doesn't produce better therapy outcomes. He says the CR study supports this because the social workers do as well as the doctoral providers. He discounts the fact that marriage counselors did worse than doctoral level providers, contending that mcs have a mixed bag of qualifications. Well, so do social workers.

But if I read Neil right, he is for the first time hedging his claim that degree doesn't matter, since now it is only social workers that he says do as well as doctoral providers. At any rate, I think it is an open empirical question in need of the right effectiveness study to test.

BTW I do not think Neil intended his claim mischievously. Rather, I believe that in a marketplace in which profit driven MCO's will grasp at any straw to justify providing cheaper providers, the claim has had mischievous EFFECTS.

Date: Sun, 25 Jul 1996 19:16:03 -0400 (EDT) From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: SSCP Network <sscpnet@bailey.psych.nwu.edu> Subject: Re: Long Term Therapy: good evidence

Lee is bothered by my claim that at last we have good evidence that long term therapy does better than short term therapy. This claim is based on two positive pieces of evidence: the CR study and Ken Howard's studies showing a dose-response curve for duration of therapy. These are the only two large scale studies I know which have investigated the question and both converge. Moreover, both use quite different methods and different samples.

Perhaps Lee can tell us about negative evidence that disconfirms the dose-response curve.

Date: Fri, 26 Jul 1996 10:56:49 -0900 (PDT)
From: Dianne Chambless <chambles@email.unc.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: Long Term Therapy: good evidence

Marty -

I won't stop to get references for you, but I can tell you with assurance that there are studies with random assignment for behavior therapy for various disorders that show brief treatment (7 or 8) sessions to be effective and as effective as longer treatment. I would argue that I need more time with some disorders, but I don't yet have any evidence that backs me up. So, we have the possibility that with some approaches to psychotherapy longer is better than shorter, or that the people who stuck out longer therapy in Ken Howard's and the CR study were different in some way.

There are a number of reasons that we don't have much controlled outcome data on long-term therapy. For example, NIMH is reluctant to fund (and my task force argued that APA needs to pressure NIMH to do so). But at a practical level we find that clients resist entering a study on long-term treatment and drop-out of long term treatment (perhaps the folks at Penn CPR can chime in here if they're lurking, because they've been battling these problems with studies of supportive-expressive psychotherapy). From over 20 years of practice, I can also say that my personal experience is that the long-term therapy crowd is different -- typically especially psychologically minded, interested in growth, and usually works as a psychotherapist!

At any rate, I do believe we need good studies of longer treatments, but that naturalistic studies are going to be hard to interpret.

Date: Fri, 26 Jul 1996 12:23:10 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>,
Dianne Chambless <chambles@email.unc.edu>
Subject: Re: Long Term Therapy: good evidence

Diane's memo points out all the interpretive problems when we allow people to choose whether to enter into short term versus long term therapy and then allow them to choose to stay or exit and then ask for a retrospectively based judgment of the degree to which it "helped."

Of course, allowing such selective entry and exit is "representative" of what goes on out there--but that's the whole point. Just looking at what goes on (or recreating it in miniture in order to claim ecological validity) leave us with uninterpretable results.

Church attendance has been suggested as an analiogy. I suggrst laetrile. People who have taken it for 20 years afterr their cancer was diagnosed will attest to its effectiveness.

Robyn

Date: Fri, 26 Jul 1996 15:40:52 -0400 (EDT)
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: sechrest@U.Arizona.EDU (Lee B Sechrest)
Subject: Re: Random assignment and control groups

Lee asserts:

>

> I have said this before without eliciting any reply from Marty nor any others of the soft methods that they cannot have it both ways. CR type studies, and other similar soft studies, cannot be accepted as "good evidence" for the effectiveness of some therapies and be rejected as evidence for others, e.g., New Age stuff, prayer, etc. How about it, Marty (and maybe Ken, I can't tell)? Are you willing to admit the legitimacy of such evidence for all therapies?>

Two comments:

1) The CR study asked about psychodynamic, cognitive, behavioral, feminist, and other. If memory serves, "other" was quite a small category. So I don't see what evidence was produced supporting a dose-response curve for prayer, New Age hot-tubs, and the like.

2) If Lee is asking what I would think if similar evidence was produced for a dose-response curve for prayer or nude marathons, I would say that I'm a Bayesian about evidence. More exactly, I'm partial toward a "coherence" theory of truth: To elicit equal belief increments, evidence of a given strength is modulated by how many other credible propositions I have to reject to believe it. I'd need stronger evidence for prayer, than I would for long term therapy of the four sorts CR asked about.

Date: Fri, 26 Jul 1996 15:30:49 -0400 (EDT)
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: jm-bailey@nwu.edu (michael bailey)
Subject: Re: Long Term Therapy: good evidence

>> Marty Seligman wrote:

It should also be said that no statistical modelling was needed to see the effects of long term therapy in the CR study. It was the biggest effect in the raw data. It surprised me. We then tried to eliminate it or reduce it with regression. The effect held up over every plausible model we could generate.>>

> Michael Bailey answered:

Marty, does the design eliminate or reduce the possibility that those who receive long term therapy were (a) those who had it together enough to stick it out (e.g., would one also get a strong "effect" of regular church attendance?), and (b) those whose stressors were long enough past so that their distress levels have regressed to their personal means? >

The people who had long term therapy differed from the rest in several ways:

They were more severe initially

More emotional disorders

Psychodynamic Treatment
More medication

It is also possible they differed in the ways Mike suggests. These were not measured. By the time of the survey however they were in better shape on every outcome variable than all the other durations of treatment. This robust finding clearly needs replication, explanation, and a search for mechanism, or artifacts.

Date: Fri, 26 Jul 1996 14:25:27 -0700 (PDT)
From: Neil Jacobson <njacob@u.washington.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: Long Term Therapy: good evidence

The differences between long-term and short-term therapy are totally uninterpretable, because they are confounded by differential probability for spontaneous remission. In other words, the longer you are in therapy, the more likely it is that you will get better for reasons having nothing to do with therapy. In our AP article, Andy and I argue that there are many reasons to suspect spontaneous remission as a confounding variable. I know that spontaneous simply means that it is one more variables unaccounted for rather than true spontaneity, but that is my point: the longer you are in therapy, the greater the opportunity for factors having nothing to do with therapy to produce change, which is then mistakenly attributed to therapy. Without proper experimental controls, there is no way to rule out this possibility. I think it is more than possible: for reasons spelled out in the article, I think that is what happened in the CR study.

Neil S. Jacobson

University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

Date: Fri, 26 Jul 1996 10:01:48 -0500
From: k-howard@nwu.edu (Ken Howard)
To: sscpnet@bailey.psych.nwu.edu
Subject: Mediocracy

I think that Brad Smith is on the right track when he states (with my slight editing) that: This post hoc reasoning (AS WELL AS THE RESULTS OF A SINGLE RANDOMIZED EXPERIMENT) can, with the help of latent variable models, reify garbage that does not replicate.

We need strong theory and programmatic research.

Ken Howard

Department of Psychology
Northwestern University

Evanston, IL 60208-2710
Phone: (708) 491-7373
Fax: (708) 491-7859

Date: Fri, 26 Jul 1996 17:44:29 -0400
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: sscpnet@bailey.psych.nwu.edu
Subject: Spontaneous Recovery?

Jacobson & Christensen (1996) claimed that because there is no control group in the CR study, one cannot infer that patients in long term therapy got better because of treatment rather than because of spontaneous remission. They are wrong in their claim. Spontaneous remission specifies the alternative cause--time--explicitly; so it is possible to test for the effect of time itself internally. Individuals who saw marriage counselors or family doctors for an equivalent length of time did not do as well as those who saw other mental health specialists. This means that spontaneous remission was not likely to be the cause of the superior outcome.

Now it is possible (but far-fetched) that marriage counselors and family doctors do something noxious (unknown and unspecified) that inhibits the healing effect of time. Such an ad hoc possibility hardly supports the claim that long term therapy had its benefits because of spontaneous recovery.

Date: Fri, 26 Jul 1996 18:49:25 -0400 (EDT)
From: Jonathan Huppert <huppert@acs.bu.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: Spontaneous Recovery?

The Jacobson and Christensen (1996) article deals with at least half of your rebuttal, as has Dr. Jacobson on the net. He has repeated that there is evidence to show that people do not accurately report the status of their clinicians, and that they are more likely to call their clinicians "Marriage counselors" when they don't like them or therapy doesn't work. This would explain why marriage counselors do worse than other therapists. The former are the cases that don't spontaneously remit. The latter may be. Is there something I am missing?

Jonathan Huppert
Boston University

Date: Fri, 26 Jul 1996 19:15:25 -0400 (EDT)
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: huppert@acs.bu.edu (Jonathan Huppert)
Subject: Re: Spontaneous Recovery?

What you are missing is that Jacobson's argument is ad hoc. He invokes his impression that somehow therapists get mislabeled marriage counselors when therapy doesn't work (but don't get mislabelled social workers or psychologists when therapy doesn't work> What is the nature of this "evidence." Do they get labeled family doctors as well, when therapy doesn't work?

Somehow only the cases that don't spontaneously remit happen to see family doctors and therapists they mislabel as marriage counselors? Too many epicycles for me.

Date: Fri, 26 Jul 1996 23:39:14 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: Spontaneous Recovery?
--

Marty,

Your claim that marriage counselors and family docs did more poorly than other mental health professionals addresses a different question than the comparison between short-term and long-term therapy. We have already discussed the marriage counselor issue, and I have pointed out the interpretive problems regarding their poorer performance. Family docs, as you yourself admit, saw clients for fewer sessions per unit time, and therefore do not control for anything. I guess these issues are too complicated to be debated with sufficient clarity in the internet, and I look forward to your rejoinder in the American Psychologist.

Neil S. Jacobson

Date: Sat, 27 Jul 1996 14:21:40 -0400 (EDT) From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: jwood@utep.edu (Jim Wood) Subject: CR's private dataset

CR still refuses to release the dataset. They released dribs and drabs of it to me, to Neil, to Crits-Christoph and others when we requested specific pieces for use in the October AP issue. This is a serious and vexing problem. They have ignored my repeated requests to release the whole dataset to any respectable group of scientists they chose.

Letters to Joel Gurin, editor of CR, from some of you might help.

This sad fact, along with the urgency of the policy issues at stake, should push us to design and carry out the "right"--methodologically as impeccable as we design--study of the long term therapy issue and the qualification of provider issue. So rather than prolong the debate on the pro's and con's of the CR article here (it's all in the October issue of AP), let me suggest the following more constructive course:

Is there a design we can agree on? If there is, I will do my best to try to make the study a reality.

Neil, Lee, Jim, George, Harry, Ken and others:

What is the right design--in detail or in schema--to ask

1)whether long term therapy works better than short term therapy
and if so, what is the cost benefit analysis?

2)whether qualifications of provider (education, experience, pre-existing skills) improve
outcome and if so, what is the cost benefit analysis?

Date: Sat, 27 Jul 1996 16:55:57 -0400 From: stricker@sable.adelphi.edu To: seligman@CATTELL.PSYCH.UPENN.EDU, SSCPNET@bailey.psych.nwu.edu Subject: Re: CR's private dataset
--

Dear Marty,

I hope you are able to make such a study a reality, and look forward to it. As to your question, before we can decide how to design a study comparing short term and long term therapy, we have to agree on outcome measures. I don't think the question is if short term or long term is better, any more than if an antibiotic or an anxiolytic is better - they do different things, and to different people. If we want to make circumscribed behavioral changes, short term should be superior; if we are interested in more sweeping personality change, or in working with people with major Axis II involvement, long term should be the choice - now the task is to demonstrate whether my guesses are so. And none of this addresses the issue of who should pay for whatever we choose to do, which is another important question, and one that depends on values rather than data. We can be guided in our choice by data, but the ultimate decision as to what is worth our money is an individual one. George Stricker

George Stricker

Internet address (STRICKER@SABLE.ADELPHI.EDU)
Telephone (516) 877-4803, Fax (516) 877-4805

Date: Sat, 27 Jul 1996 15:53:08 -0700 (MST) From: Lee B Sechrest <sechrest@U.Arizona.EDU> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: CR's private dataset

Lee Sechrest, Univ. of Arizona

I have an idea. Why don't we all treat the CR piece as the pariah "science" it really should be until the complete data set is released for examination by scientists. We do not let scientists get away with such strategies as the CR gang (of how many?) is employing. The fact of the matter is that the CR study is not science, no matter how scientific it may have been. It belongs in the same category as the proprietary studies of tobacco companies and the like. When they make it and its details widely available for peer review, then we might agree to incorporate its findings, whatever they are, into the recognized body of scientific knowledge.

So, perhaps starting with Marty, why don't we agree that we will not give further shrift to the CR study and its findings until such time as it is made properly available for peer review? Until then, it is simply another in the line of many product evaluations--

refrigerators, cars, movies, mouthwash--that CR has done. Caveat emptor, which goes both for emptying psychotherapy and CR.

Date: Sat, 27 Jul 1996 08:31:38 -0400 (EDT)
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: njacob@U.WASHINGTON.EDU (Neil Jacobson)
Subject: Re: Spontaneous Recovery?

It's not complicated at all, Neil. Spontaneous recovery says that TIME ALONE healed. Family docs and marriage counselors saw the patients for identical amounts of TIME. If long term therapy worked only because of time alone, any other professionals (or huka teachers) who saw their patients for the same amount of time should do just as well. MC and FD's did worse over just the same time. So it's not spontaneous recovery. QED.

Date: Sat, 27 Jul 1996 08:58:22 -0900 (PDT)
From: Jim Wood <jwood@utep.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: Spontaneous Recovery?

On Sat, 27 Jul 1996, Martin E. P. Seligman wrote:

> It's not complicated at all, Neil.see above.>

Actually, spontaneous recovery rates would not necessarily be the same if the patients treated by the different provider groups (psychologists, marriage counselors, family doctors, etc.) were different to start with (e.g. type of problems, severity of problems, background characteristics). It seems to me that Jacobson, Dawes, Sechrest and others keep making a very elementary point that just isn't sinking in: in the absence of randomization or unambiguous control groups, the between-groups differences observed in the CR study are probably uninterpretable.

Jim Wood

Department of Psychology
University of Texas at El Paso

Date: Sat, 27 Jul 1996 08:04:16 -0700 (PDT)
From: Neil Jacobson <njacob@u.washington.edu>
To: Jim Wood <jwood@utep.edu>
Subject: Re: Spontaneous Recovery?

Jim,

I have to agree with you on this one. It is an elementary point, straight out of Campbell and Stanley, and it is somewhat embarrassing to even be debating it.

Neil S. Jacobson

Date: Sat, 27 Jul 1996 09:25:53 -0700 (MST)
From: Lee B Sechrest <sechrest@U.Arizona.EDU>
To: Robyn Dawes <rd1b+@andrew.cmu.edu>
Subject: CR-methodology

Robyn,

I now have confirmation of the fundamental correctness of Marty's methodology for testing effects of interventions.

I have just returned from a week of instructing at workshops and learned that participants in the workshops were more satisfied if they stayed for the entire week. Those who left after a day or two reported lower satisfaction with the workshops and that the workshops were not terribly helpful to them. My own involvement in the workshops lasted all week, and it is gratifying to know that I did so much good.

Lee Sechrest

Date: Sat, 27 Jul 1996 10:33:39 -0900 (PDT)
From: Jim Wood <jwood@utep.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: Spontaneous Recovery?

On Sat, 27 Jul 1996, Martin E. P. Seligman wrote:
see above

Cook and Campbell (1977) provide a chapter that explains why statistically "partialing out" confounding variables doesn't work in these situations.

In most circumstances, "partialing out" results in "under-correction". That is, one is left with apparent "between group differences" that really don't exist -- as may well be true in the CR study. I really do think that this point and those made by Jacobson, Sechrest, Dawes etc. are pretty widely known.

While we're on the topic of the CR study: Is anyone else bothered by the fact that the American Psychologist article is based on data that are not accessible to the general scientific community? My own work on Exner's Comprehensive System for the Rorschach has made me pretty sensitive to this issue. Marty and I discussed this point on the net a few months ago, and my understanding was that he was "working on it." Has there been any progress in getting the data released? Is it true that the various commentaries coming out in the October American Psychologist had to be written without access to the data?

Interesting stuff!

Jim Wood

Department of Psychology
University of Texas at El Paso

Date: Sat, 27 Jul 1996 11:42:46 -0400 (EDT)
From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
To: jwood@utep.edu (Jim Wood)
Subject: Re: Spontaneous Recovery?

Marty replies to Jim:

Of course, Jim. That's why CR partialled out severity, type of problem, demographic differences in testing whether long term therapy did better than short term therapy. Further CR partialled out these variables in finding that family docs and marriage counselors did

worse than other mental health professionals, even when they saw patients over the same length of time.

So what Neil has to contend to defend spontaneous recovery as an explanation of why long term therapy worked better than short term therapy is:

1) Even though family docs and marriage counselors saw patients for just as long as other mental health professionals, they did worse because

2) FD's and MC's do something (unmeasured and unspecified) that inhibits the course of natural healing by time itself

or

3) Some internal subject characteristic, unmeasured and unspecified, of the subjects who saw FD's or MC's prevented spontaneous recovery in just these subjects.

Logically possible, but that's a lot of unspecified epicycles.

Jim went on to say:

> It seems to me that Jacobson, Dawes, Sechrest and others keep making a very elementary point that just isn't sinking in: in the absence of randomization or unambiguous control groups, the between-groups differences observed in the CR study are probably uninterpretable. >

Marty replies to Jim:

I like to think I understand the "very elementary" logic of control groups and randomization of subjects. What isn't sinking in is the trade-off between naturalistic designs and experimental designs:

Randomization of subjects to groups and "unambiguous" control groups (BTW, I have never seen an "unambiguous" control group) are fine if they can be carried out. But they cannot be carried out to test the effects of long term therapy because:

1) manuals cannot be written for the years of therapy or for plausible years-long placebo controls,

2) control groups of equal severity who talk to friends or pastors for an equal length of time cannot be found, 3) and subjects cannot be ethically assigned at random to placebo controls for years--even if 1 and 2 could be solved.

Unless you can solve those three problems, if you wish to investigate if long term therapy works and how much it is worth, you MUST use naturalistic designs and regression logic. In those designs, disputes get resolved by specifying competing variables and testing for them, just as CR did for spontaneous recovery. If you are unwilling to use naturalistic designs, you can only validate brief, manualized therapy.

Date: Sat, 27 Jul 1996 13:30:11 -0700 (PDT) From: Neil Jacobson <njacob@u.washington.edu> To: Jim Wood <jwood@utep.edu> Subject: Re: Spontaneous Recovery?

Jim,

Yes, it's true that we had to write the commentaries without access to the data. Marty tried to get it released, but was unsuccessful. However, if data are unreleased, they are unreleased. The playing field should be level in this debate. What bothers me is, the playing field is not level. When the study is criticized, new data get leaked to apparently debunk the criticism. I hope that, when rejoinders are written, they are based on the data to which we have all had access and a chance to absorb.

Neil S. Jacobson

Date: Sat, 27 Jul 1996 17:40:13 -0600 (MDT)
From: Stephen Ilardi <ssi@psych.colorado.edu>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject; CR replication study

On Sat, 27 Jul 1996, Martin E. P. Seligman wrote:

- > What is the right design--in detail or in schema--to ask
 - 1)whether long term therapy works better than short term therapy and if so, what is the cost benefit analysis?
 - 2)whether qualifications of provider (education, experience, pre-existing skills) improve outcome

It seems that the "right design", as Dr. Seligman puts it, for answering these questions would need to be constructed in such a way as to take into account the following interrelated issues:

1) Target syndrome.

It may well be that the effects of short-term/long-term treatment, or expert/rookie clinician, may vary considerably depending upon the symptom(s) which serve as the target of intervention. As George Stricker notes, amelioration of Axis II pathology may require a longer time frame (Linehan's DBT protocol is typically conducted for 12+ months) than, say, major depression. Pooling across all reported controlled CBT trials, Ilardi & Craighead (1994) found an average 60% reduction in initial HRSD scores within 4 weeks -- so we wouldn't expect big long-term vs. short-term advantages vis-a-vis acute MDE symptoms.

2) Severity of presenting symptoms.

In the Elkin et al (1995) reanalysis of TDCRP data, NO TREATMENT WAS SIGNIFICANTLY BETTER THAN PLACEBO among patients with initially mild (HRSD < 20) depression. It was only among more severely depressed patients that significant treatment effects emerged. Thus, it would be surprising to find that therapist experience is strongly related to outcome among less severely depressed patients.

3) Experience versus training.

I don't have data, but certainly lots of anecdotal material which suggests that quality of training may be vastly more important to patient outcomes than years of experience. Here in Boulder, we seem to have a plethora of experienced private practice clinicians who do lots of supportive therapy, but have no clue how to effectively treat OCD, panic, simple phobia, borderline PD, etc. Thus, I think it might be important to treat experience and training as separate constructs. Also, based on my own experience, I'm going to suggest that the relationship between #years in practice and patient outcomes is non-linear -- a very steep increase over the first 3-5 years, which then tapers off very rapidly.

4) Long-run versus short-run costs.

Approximately 50% of remitted depressed patients will relapse within 12 months post-treatment if they receive no follow-up treatment. Since we know that maintenance (i.e., long-term) therapies reduce the risk of relapse, they might look quite cost-effective if costs are projected out over a 1-2 year period.

But what sort of study (or studies) could provide more compelling answers to Dr. Seligman's questions (keeping in mind the aforementioned principles)? Dave Miklowitz and I were discussing this earlier today (evidence of his mental powers is found in the fact that he came up with much of the following study design while he was lifting massive weights at the gym), and we came up with the following as at least a modest beginning:

For each syndrome of interest (MDE, BPD, OCD, panic, GAD, ETOH abuse, etc.), consider a 4(number of treatments -- 2 different psychosocial, 1 pill placebo, 1 med)-by-2(initial syndrome severity)-by-2(short-term vs. long-term)-by-3(therapist training: rookie, post-post doc, expert) design. We're assuming that such a study, to have modest power, would require an unprecedented n (10 in each group would require $n=480$). Actually, sample size could be reduced by treating severity as a continuous variable and making sure to sample across a full range of initial values. Also, the study would employ multiple baseline measures of symptomatology, which would be assessed at regular intervals for all subjects throughout a minimum 24-month period (this would permit analysis of relapse/recurrence, among other things). And of course, we're assuming initial random assignment of patients across all conditions.

One important characteristic of such a study (aside from its cost, which could rival the annual salary of certain NBA stars) would be its ability to shed light on interaction effects which might exist among the aforementioned constructs. For example, it may turn out that the only borderline patients with good outcomes would be found in the following condition: DBT, expert therapist, long-term therapy. Or, it might be that rookie CBT therapists do rather poorly with MDE in the short-term (I think Steve Hollon made this argument regarding the TDCRP study), but that they end up getting adequate long-term results by virtue of OJT effects.

The above is not intended as *the* definitive study, by any means, but I believe it might represent an improvement over the CR study (which, by the way, seems to deserve much credit even, among its detractors, for catalyzing interest in such important questions) in terms of addressing the questions at hand.

Stephen Ilardi, Ph.D.

Department of Psychology
Box 345
University of Colorado
Boulder, CO 80309-0345
(303) 492-4159

Date: Sat, 27 Jul 1996 09:21:31 -0700 (MST) From: Lee B Sechrest <sechrest@U.Arizona.EDU> To: Ken Howard <k-howard@nwu.edu> Subject: Re: Mediocrity is so easily acheived...

On Fri, 26 Jul 1996, Ken Howard wrote:

> to quote Lee Sechrest "mediocrity is so easily acheived that it is not worth planning for".If its so easy to achieve, why can't Lee do it? >

Ken, I did not plan on even that level of competence.

Lee

Date: Sat, 27 Jul 1996 20:07:07 -0400 From: DF Klein@ To: seligman@CATTELL.PSYCH.UPENN.EDU Subject: long vs short
--

dear all
please take a look at these relevant ancient articles

Rosen, B, Katzoff, A, Carillo, C & Klein, DF: Clinical effectiveness of "short" vs "long" psychiatric hospitalization I: Inpatient results. Arch Gen Psychiatry 33(6):1316- 1322, 1976.

Mattes, J, Rosen, B & Klein, DF: Comparison of the clinical effectiveness of "short" vs. "long" stay psychiatric hospitalization II: Results of a three year post hospital follow-up. J Nerv & Ment Dis 165(6):387-394, 1977.

Mattes, JA, Rosen, B, Klein, DF & Millan, D: Comparison of the clinical effectiveness of "short" vs "long" stay psychiatric hospitalization III: Further results of a three year post hospital follow-up. J Nerv & Ment Dis 165(6):395-401, 1977.

briefly i was director of research at hillside hospital a 200 bed psychoanalytic hospital-major modality psychodynamic psychotherapy -3X weekly supervised by ny psychoanalytic institute staff - with an average length of stay of 10 months supported by the forward thinking citizenry of NYC. plainly this was the jurassic era.

the city (which then supported a maximum of a mere 90 days inpatient psychiatric care at any other hospital) grew restless and threatened to cut us back to this cruelly abbreviated mockery of necessary intensive and extensive treatment. this was vigorously protested by our enlightened citizenry so the city used salami tactics. their ukase was that 1/2 the beds were now maximum 90 day beds and the rest were still ad lib.

"eureka" i quoth,"its a study."

firmly backed by the great open minded analyst lew robbins who was medical director i took over the admissions screening and Randomly (that word) assigned patients to the 90 day and long term units. patients were evaluated at admission,discharge,and three years later.

our results--well you really should read the articles but i'll give you a hint. when the city cut out the ad lib units we had no reason to disagree nor did our follow up indicate any differential prophylactic effects.

but to return to marty's question-its premature to design a comparative trial since we dont have the preliminary phase 2 data ie a consecutive series of patients ,well diagnosed and evaluated longitudinally of sufficient magnitude to pick out what looks like a substantial sample of specific responders . first we need a good open study.there are very few of those. even the menninger study was of completed treatments and forget randomization.

once its determined that say highly separation anxious,frequently fired, single women who are poor romantic pickers and prone to develop nonautonomous reactive depressions to interpersonal rejection are in sufficient supply and that they do well in long term care and further maintain their gains for 1-2 years post treatment (this is all a fantasy) you can then move to phase 3.

such patients are told that our clinic offers two treatments - long and short. a random 1/2 are randomly assigned to rx --the rest are given their choice. if you want to add some bells and whistles you can randomize within rxs as to whether a manual is used as opposed to a generally training the therapist in principles and practices.this is all followed up by appropriately blinded independent,therapist and self evaluations thru treatment and followup with attention to objective indicators of success eg holding a job longer than 3 months,paying the rent etc

such a design should illuminate whether choice by patient of treatment matters or length of treatment or manualization along with a raft of interactive explorations.

none of this should be taken too seriously unless it was co-directed by mutually critical adherents of long and short,choice and assignment,and manualization vs flexible application of principle --with the statistical analysts blinded to treatment assignment thus avoiding toxic allegiance effects. and of couse replications(2-3?) are essential.

it may turn out that in a phase 2 properly studied consecutive group of attempted long term patients that given a high proportion of drop out or termination in an unsatisfactory state or rapid relapse post treatment termination that one might not feel fired up to press on-- but we dont know do we? thats why we need phase 2 as i have been trying unsuccessfully to get the psychoanalytic institutes to understand before the deluge. cordially,
donald f. klein

Date: Sun, 28 Jul 1996 07:32:22 -0400 (EDT) From: John F Kihlstrom <john.kihlstrom@yale.edu> To: SSCP Network <sscpnet@bailey.psych.nwu.edu> Subject: Re: CR's private dataset

I agree with Lee Sechrest.

The CR dataset should be strongly and publicly disavowed by all concerned with it. Perhaps such a disavowal, or even the threat of a disavowal, by the President-Elect of APA, until now among its strongest supporters, would force CR to release the data.

At least, it would give the public a good example of the self-correcting nature of science.

John F. Kihlstrom, Professor

Editor, Psychological Science
Department of Psychology, Yale University
P.O. Box 208205, New Haven, Connecticut 06520-8205
Department Address: Kirtland Hall, 2 Hillhouse Avenue
Telephone (203) 432-2596 Facsimile (203) 432-7172

Date: Sun, 28 Jul 1996 08:44:59 -0400 (EDT) From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: john.kihlstrom@yale.edu (John F Kihlstrom)

Subject: Re: CR's private dataset

I assume that it the dataset that sscp members want, and not posturing which I have good reason to think will make CR even less likely to release the dataset. My judgement is that John Kihlstrom and Lee Sechrest and any others should first take the trouble to write or call Joel Gurin, editor of CR, and Rhoda Karparkin, the publisher, and express their feelings about it. They should offer to be custodians of the dataset or to form a disinterested consortium to have such custody.

The fax is 914 378 2912.

John and Lee, will you do that?

If that fails, we can take the next measured step.

Date: Sun, 28 Jul 1996 08:12:34 -0700 (PDT)
From: Neil Jacobson <njacob@u.washington.edu>
To: Lee B Sechrest <sechrest@U.ARIZONA.EDU>
Subject: Re: CR's private dataset

Lee,

The problem is that the data get leaked when the CR study is criticized. And there are commentaries already completed for the October issue of AP. As I said in my last message, it is too late to call a moratorium on the debate, given that there is one already in press. But I hope we all stick to the data released by CR publicly and don't resort to data not publicly accessible to refute arguments.

Neil S. Jacobson

University of Washington
Department of Psychology
Center for Clinical Research
1107 NE 45th, #310
Seattle, WA 98105
Phone: (206) 543-9871
Fax: (206) 685-3293

Date: Sun, 28 Jul 1996 10:47:40 -0700 (MST)
From: Lee B Sechrest <sechrest@U.Arizona.EDU>
To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU>
Subject: Re: CR's private dataset

Lee Sechrest, Univ. of Arizona

I will be very glad to write to CR about the release of their data set. I must say, however, that I am a bit mystified about what is at stake. Why does a data set need "custodians?" A data set is just numbers along with reasonable documentation. What care and feeding does it require? There are data sets that are ongoing and developing that probably do require

curators in order to maintain them. But any complete data set ought to be a truly low-maintenance operation.

Unless there is concern that somehow a data set might be misused? Any data set can be poorly analyzed and reported on, but that is why we have peer review. Or a data set may involve problems of confidentiality. I cannot see any such problem in the CR data.

Does anyone else imagine any problems that would require any intervention once the data set is released?

In my view, the interests of science would be best served by simply making the data set available to anyone who wants to have a crack at it.

Lee

Date: Sun, 28 Jul 96 20:17:11 EDT From: "George F. Ronan" <33Q4773@CMUVM.CSV.CMICH.EDU> To: SSCPNET@BAILEY.PSYCH.NWU.EDU Subject: A "CR-TYPE" Study
--

The proposed resolution for the CR Study is very reasonable. Kudos!

Dr. Seligman asked for ideas regarding the next survey type study.

Instead of recommending a specific study to test some of the ideas that have been proposed (which I agreed is also important), I wonder whether APA could forge relationships with the companies that actually reimburse for psychological services (e.g., B X BS, Etna, Prudential, Medicaid, Medicare, etc.). APA could offer to store and analyze some of the large data sets that are collected on the reimbursement of psychological services. APA could help the insurer, the consumer, and the profession get a handle on what is going on nationally. Then we would have a representative data set. Although less likely to happen, we might even suggest a standard assessment protocol that insurance companies could administer pre and posttreatment (e.g., SCL-90 & a Problem Checklist or a different one for each Dx). Or better yet, maybe APA could fund the development of a standard assessment protocol and give it away free to insurance companies that work with APA to evaluate the efficacy of the services being reimbursed.

I think Bill Bermann was hinting at something similar in a post he made about a week ago. George.F.Ronan@CMUVM.CSV.CMICH.EDU

Date: Sun, 28 Jul 1996 16:26:40 -0400 From: Stephen Soldz@ To: seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman) Subject: Re[2]: Spontaneous Recovery?
--

Marty,
you wrote: >"What isn't sinking in is the trade-off between naturalistic designs and experimental designs: Randomization of subjects to groups and "unambiguous" control groups (BTW, I have never seen an "unambiguous" control group) are fine if they can be carried out. But they cannot be carried out to test the effects of long term therapy because:

1) manuals cannot be written for the years of therapy or for plausible years-long placebo controls, 2) control groups of equal severity who talk to friends or pastors for an equal length of time cannot be found, 3) and subjects cannot be ethically assigned at random to placebo controls for years--even if 1 and 2 could be solved. Unless you can solve those three problems, if you wish to investigate if long term therapy works and how much it is worth, you MUST use naturalistic designs and regression logic. In those designs, disputes get resolved by specifying competing variables and testing for them, just as CR did for spontaneous recovery. If you are unwilling to use naturalistic designs, you can only validate brief, manualized therapy.">

Excellent points! Untill people can come up with reasonable and fundable designs for experimental study of long-term therapy (which is probably never), we'll be limited to naturalistic designs with statistical controlls. Multiple studies will need to rule out various competing hypotheses.

We aren't much worse than epidemiology here, except that they frequently get 10,000 to several hundred thousand subjects to examine the relation between dietary collesterol and death, so they can do more statistical manipulations. An unfortunate aspect of this analogy is that we may end up as confused as they are in this area. It seems to me that ultimately, regardless of design issues, only a good understanding of causal mechanisms leads to final resolution of issues. Thus, research on human change processes may ultimately be more important than a hundred conflicting outcome studies.

Cheers,
Stephen

Stephen Soldz
Director of Research and Evaluation Department of Psychiatry

Date: Mon, 29 Jul 1996 09:13:13 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: njacob@U.WASHINGTON.EDU (Neil Jacobson), seligman@CATTELL.PSYCH.UPENN.EDU (Martin E. P. Seligman) Subject: Re: Spontaneous Recovery?

Marty,

Please read the Kahneman areticle. The fact that we "control" for type of problem by asking about it and receiving a response contaminated by retrospection, random error, and (additionally) unreliability of response scale (more error) does not "control" in any adequate way.

The upshot is that we can't say ANYTHING (without looking for compatibility from other sources, and even then only tentatively).

There are very compelling logical reasons for the random assignment study. Incomparibility of problem when people self-select type of therapist (and time in therapy) is just one.
Robyn

Date: Mon, 29 Jul 1996 09:33:51 -0400 (EDT)
From: Robyn Dawes <rd1b+@andrew.cmu.edu>
To: jwood@utep.edu (Jim Wood),
seligman@cattell.psych.upenn.edu (Martin E. P. Seligman)
Subject: Re: Spontaneous Recovery?

Marty,

If you would like, I will share the actual reports of the 65 studies Janet Landman and I independently coded in 1978 from the Smith/Glass list for true randomization versus not. (About 1/3 were concluded to be "not"--with very high independent inter-rater reliability.)

Perhaps there's a vogue in the past 20 years to make all or most therapies evaluated by randomized trials brief and "manualized." Neither of these descriptions characterized the therapies Janet and I examined; sure, some were, and some weren't.

I'm concerned that you keep repeating this assertion without providing evidence. Pretty soon, like all those kids who are going to go into poverty as the result of constant repetition of the need for "welfare reform," those RCT studies that are not brief or manualized are going to be repeated out of existence.

Robyn

Date: Mon, 29 Jul 1996 19:50:16 -0700 (MST)
From: Lee B Sechrest <sechrest@U.Arizona.EDU>
To: ssoldz@warren.med.harvard.edu
Subject: RCT

Stephen,

Right you are.

We are not going to do randomized trials of cryogenic preservation of bodies either, but we don't need to. In fact, is there anyone out there who is game for even a structural equations model?

Lee Sechrest, Univ. of Arizona

Date: Mon, 29 Jul 1996 09:55:58 -0900 (PDT)
From: Jim Wood <jwood@utep.edu>
To: sscp <sscpnet@bailey.psych.nwu.edu>
Subject: A Question of Standards

This weekend I read the pre-print by Jacobson and Christensen that critiques Seligman's Consumer Reports article in American Psychologist. I thought the critique was devastating, like many of the commentaries that have appeared on this net.

I'm left with two disturbing questions:

First, why does the APA "science candidate," now president-elect, keep minimizing the very serious methodological flaws in the CR study -- flaws so elementary that my second-year graduate students recognize them right away?

Second, why was Seligman's article accepted into American Psychologist in the first place? Certainly it's a great shot in the arm for professional psychology. But is this the quality of study that we expect to find in our premier journal? And what about the fact that the data is inaccessible to the scientific community -- didn't that give pause to the editor of AP?

After some reflection, I've concluded that behind all the methodological debate there really lies a more fundamental question: Should we psychologists and our journals "relax" our current requirements regarding little niceties such as accessibility of data, selection bias, and valid measurement? Are we willing to go that far to promote the financial and political interests of professional psychology?

As I read it, Marty Seligman is really inviting us to lower our standards, not only for the CR study but more generally. We can use "statistical partialing out" instead of randomization, and "causal modeling" instead of experimentation. I'm afraid that the American Psychologist article, with its many weaknesses and dubious assertions, may be just a small preview of the "science" we can expect, once our journals and researchers all learn to "loosen up" about these things.

There are some fields of social science where the standards for research are very lax. Social Work and some areas of Education come immediately to mind. However, I don't think we want to imitate such examples. The famous "split" between scientific and non-scientific psychologists cannot destroy clinical psychology as a legitimate scientific enterprise. But a general retreat from standards in our research and journals could do just that. If that's where the "scientific leadership" of APA is headed, I for one do not intend to follow.

Jim Wood

Department of Psychology
University of Texas at El Paso

Date: Mon, 29 Jul 1996 14:05:22 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: jwood@utep.edu, sscpnet@bailey.psych.nwu.edu Subject: scientific standards
--

Jim:

It is most certainly not my intention to lower scientific standards and were that the result of my apa presidency or of my own scientific contribution, I would judge both failures. SSCP will, I hope, play a constructive role in both keeping standards impeccable and at the same time investigating empirically a set of issues crucial to practice. I trust that open debate in this forum, such as your exhortation, will keep clinical science on course--although I do think that illumination is best achieved when ad hominem attacks are minimized.

First my comments on the substance of where you take issue with me:

1) You say the Jacobson & Christensen critique of my AP article is 'devastating.' I have already responded on the net to its main critiques--that the CR article supports the idea that qualifications do not improve outcome, that the CR article found nothing new, and the the benefits of long term therapy were spontaneous recovery. Each of these arguments is I believe incorrect. I have sent you an electronic preprint of my response in the October AP, and I will be happy to send it to others on request. (You wonder why AP published my December 1995 article in the first place. You should be aware that both Neil Jacobson and

Lee Sechrest generously reviewed drafts for me, and as I recall, all of their suggestions were incorporated.)

2) You say I "minimize" the flaws of the CR article. Both in the December AP and in the forthcoming October issue, I have been at pains to point out the problem of possible selection bias, the need for valid and more complete measurement, and the flaw of having only cross-sectional retrospective report. I end my October article by calling for studies of long term therapy free of such selection bias, longitudinal in nature, and measured by blind diagnosis and full assessment battery, as well as patient report. And I take exception in print to CR's withholding the dataset, and I will do the same in my remarks at the convention. 3) You equate using "naturalistic methods," regression and causal modelling with a lowering of standards. Here you are dead wrong. Both naturalistic methods and experimental methods can be used sloppily or rigorously. I have consistently advocated using both experimental and naturalistic methods--rigorously--to investigate therapy. They complement each other, and neither is inherently superior methodologically. Perhaps you can explain how astronomy has done rigorous science without experimental methods. Perhaps you can explain the belief that cigarette smoking causes lung cancer--the vast bulk of the evidence uses only naturalistic methods. Perhaps you can explain how we have come to believe that a good part of psychopathology and personality are heritable--with no experimental methods.

Lastly my view of where the 'scientific leadership' of APA is heading. I am looking to SSCP to help provide the scientific leadership for APA. At best I can help organize that leadership and make its force felt. That is why I asked SSCP to read this debate carefully and then generate the designs--both experimental and naturalistic--that will answer the questions pressing most heavily on practice--at the highest level of scientific standards.

I have concerned myself with both highlighting the methodological flaws of the CR article, as well as defending its method when the arguments against it were poor, so that better studies--using higher standards of evidence--will emerge. It is my conviction that science is the bedrock on which practice rests and the best friend that practice has. At this critical juncture, science can become the critical ally of practice. It can do this in two ways: first by continuing the debate about efficacy versus effectiveness that we are having. Second by designing and carrying out the studies needed to speak to the survival issues raised by practice's confrontation with MCO's.

Date: Mon, 29 Jul 1996 13:54:15 -0700 (PDT) From: Larry Beutler <beutler@education.ucsb.edu> To: Jim Wood <jwood@utep.edu> Subject: Re: A Question of Standards
--

I think in all fairness that we need to put some of the issues that Jim and others have raised into perspective. Aside from the fact that I Personally don't find the J & C critique either that devastating or the CR study so lacking as some (I'm sure Lee will have a comment or two about my failure), there are two issues that Jim raised should be given another look.

The failure of the CR Editors to open their data to inspection is not unique. While APA has encouraged scientists to open up their data sets to scrutiny, that has not been a requirement for some very good reasons. For example, some IRBs do not allow data, even masked, to be distributed except when a question arises about the ethics of the research and then, only with special oversight. Their concerns have been subject/participant protection. There are also many other private organizations that maintain a clear restriction on release of their data because of proprietary reasons. These include drug companies, health care companies, etc. Editors must contend with this issue frequently during the review process. The CR study shouldn't be singled out for a

problem that characterizes a good deal of research in the field, whether we like it or not or agree with it or not. The fact remains that a great deal of the data gathered is not and will not be available for inspection by other scientists unless there are charges of fraud or ethical fault.

Second, the ad hominem attack on Marty is a bit beyond the limits of good dialogue. There are some legitimate questions being raised here, and because one's undergraduate students (bred and trained by a given standard) can see the problems in the same way as their professor, does not make those issues unimportant. The CR study has now undergone a good deal of peer review. It went through peer review in its design. The fact that it was not published in a peer reviewed journal is beside the point. Is it good science? Some would say no and others would say yes. That type of disagreement is the basis of scientific argument and we need not resort to attacking the credentials and good sense of those who differ in their opinions from us.

It seems to me that we are addressing some very strong and rigid opinions here. Those on both sides are prone to use a double standard of evidence. Those of my ilk are accused of arguing away ns findings and being soft on positive ones, when we discuss either effectiveness research, ATI, or therapist experience. But, double standards are also present when one ignores what many of us believe are compelling lines of evidence to support (1) the value of therapist experience among the most severely disturbed patients, (2) the necessity of using effectiveness designs in studying long term treatments because of all the reasons Marty detailed, (3) the roles of other patient moderators in therapy effectiveness, etc. We see a logical flaw of arguing only from the negative. We find evidence of a double standard when every instance of a positive effect is dismissed, seemingly out of hand, because of what are often inevitable design flaws in order to accept a null hypothesis that derives from very weak studies that find ns results. The issue of experience of therapist again comes to mind.

LEBeutler

Date: Mon, 29 Jul 1996 17:09:18 -0500 From: michael bailey <jm-bailey@nwu.edu> To: sscpnet@bailey.psych.nwu.edu Subject: value of Consumer Reports study

It seems to me that neither extreme interpretation of the CR study--that it is scientifically worthless or that it (even nearly) clinches the case for the value of longer versus shorter therapy, the value of training, etc.--neither of these is tenable.

I do not know what Marty Seligman has said in public; I haven't read the CR study. I am generally impressed with his SSCPNET pronouncements though. I understand his position as basically, e.g., "the data sure look the way they should if longterm therapy were especially beneficial, though of course there are possible confounds we haven't yet examined that could account for the findings. More and better controlled studies are needed. To the extent that those are consistent with the CR results, there is progress. To the extent that they are inconsistent, there are enough limitations of both CR and most experimental studies, that we don't automatically know which to believe."

I do not read Marty as saying either that (a) randomized designs aren't needed or (b) CR definitively answers any important question (i.e., no more studies are needed). If I'm wrong, and he does say those things, I disagree.

Now a more difficult issue is, what should policy be given the substantial limitations in our knowledge. For example, should we reimburse PhDs more than "marriage counselors" treating the same type of patient?

Michael Bailey

Department of Psychology
Northwestern University
Evanston, IL 60208-2710
Office: 847-491-7429
Fax: 847-491-7859
jm-bailey@nwu.edu

<http://www.psych.nwu.edu/psych/people/faculty/bailey/bailey.html>

Date: Mon, 29 Jul 1996 17:04:39 -0900 (PDT) From: Jim Wood <jwood@utep.edu> To: "Martin E. P. Seligman" <seligman@CATTELL.PSYCH.UPENN.EDU> Subject: Re: Scientific Standards & SSCP
--

Dear Marty,

Thank you for your spirited response. It grapples directly and eloquently with the issues that I raised concerning scientific standards.

A couple messages have referred to the issue of "ad hominem" arguments. If anyone has the impression that I'm attacking you personally, I would like to clarify and apologize. Your scientific credentials are above reproach, and you clearly are trying to lead psychotherapy outcome research in a direction that you consider worthwhile.

My concern, however, is that in the name of "naturalistic research" you sometimes seem to be advocating a relaxation of scientific standards. As you point out, astronomy is a well-established science based on naturalistic, non-experimental data. However, it did not reach its present level of maturity by cutting corners, or by relying on weak evidence like the Consumer Reports study.

With best wishes,

Jim Wood

Date: Mon, 29 Jul 1996 16:30:50 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: sscp <sscpnet@bailey.psych.nwu.edu>, Jim Wood <jwood@utep.edu> Subject: Re: A Question of Standards

Jim,

Having resigned from APA in 1988, I can no longer "work from within" to try to enforce such requirements as sharing data whenever an assertion is made on their basis.

On the other hand, I found--primarily from serving on the APA Ethics committee and also as a state president--that "working from within" means simply being listened to politely before being told that you "just don't get it" and then having whatever little clout you have being added to enhance the legitimacy of those whom you do not understand.

It is perhaps a retrospective illusion that it is no surprise that all sorts of rules--like data sharing--are violated if the purported result is supposedly good for the profession (guild).

If an illusion, it is a strong one. And maybe it's not an illusion.

Robyn

Date: Tue, 30 Jul 1996 04:37:13 -0700 (MST) From: Lee B Sechrest <sechrest@U.Arizona.EDU> To: Michael Bailey@ Subject: Re: value of Consumer Reports study

Lee Sechrest, Univ. of Arizona

On Mon, 29 Jul 1996, michael bailey wrote:
see above

Mike, you are correct in your views, which means, of course, no more than that I agree with you.

If Marty took only the position you outlined above, I could agree with him. But he goes a long way beyond that, and publicly too. As one example, there is a *considerable* disconnect between the view of the CR-study as you have stated it and the use of the term "greed" to describe the decisions of HMOs and other payers not to reimburse for long term therapy.

That is the position that Marty took in his public statement in asking for votes for APA president.

To use the term greed in that context, and he has done so in other places, requires a much stronger interpretation of the CR study than you have outlined.

It is particularly disturbing that Marty lodges the "greed" charge so readily when the scientific community, let alone the health maintenance community, has not had a chance to digest the "good evidence" that would justify the charge of greed. The response to the CR study has not yet been published, and yet the results of the study apparently should somehow already have been known to mental health insurance payers and accepted by them as a basis for payment.

In fact, I sort of remember Marty having made the greed and related statements on this network almost from the very beginning of the CR controversy, suggesting that his position was already established and that he expected that everyone else should already have adopted it.

In the inimitable words of Cool Hand Luke, what we have here is a failure of communications.

Date: Thu, 1 Aug 96 02:49:49 UT From: "John Kline" <KlineJP@msn.com> To: "Robyn Dawes" <rd1b+@andrew.cmu.edu> Subject: RE: value of Consumer Reports study

>" Marty's arguments are not matters of emphasis, approach, or shading. They are extraordinarily strong claims. If all--or even any--of these claims are correct, they should lead to a total revolution in the way research is conducted, perhaps even a change to not bothering to conduct it at all."

Robyn

Robyn,

This is slippery-slope reasoning, not to mention setting up Marty's arguments as straw men. Marty makes some legitimate points, which may constitute methodological frustrations, but so be it. This is a far cry from saying that research should not be conducted at all..

We have to face the possibility that efficacy studies face some of the same problems as other experimental studies. For one, Internal validity does not guarantee external validity, and in some instances can undermine it. There seems to be a tradeoff between external and internal validity of studies, such that the more experimental control is exercised, the less that the experiment captures the phenomenon of interest. Furthermore, it is often difficult to

connect outcomes of laboratory studies with those of field studies.

John Kline

Date: Wed, 31 Jul 1996 16:33:25 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: sscpnet@bailey.psych.nwu.edu, michael bailey <jm-bailey@nwu.edu> Subject: Re: value of Consumer Reports study

Michael,

"I do not read Marty as saying"--you, July 29.

Many of the rest of us read Marty as saying quite a lot, which--I suspect--is some of the reason for the heat.

Marty's heat--quotes from the AP article.

(pg.966). "I no longer believe that efficacy studies are the only [TEPID], or even the best [WARM], way of finding out what treatments actually work in the field."

(pg. 966, later). "The efficacy study is the wrong method for empirically validating psychotherapy as it is actually done [HOT]."

(pg. 973). "Random assignment of patients to a modality or to a particular therapist not only undercuts the remoralizing effects of treatment but also undercuts the nonrandom decision of therapists in choice of modality for a particular patient [HOT--especially because why such choices themselves could not be a focus of an "efficacy" study employing randomizations is never explained]."

(pg. 974). "The upshot of this is that random assignment, the prettiest of the methodological niceties [PUTDOWN] in efficacy studies, may turn out to be worse than

useless [SIZZLING] for the investigation of the actual treatment of mental illness in the field."

Taking him at his word, we reach a rather strong conclusion. He is asserting not that the CU study is an adjunct or complement of randomized trials studies, but is a superior replacement--thereby denegrating all the reasons that have been advanced for decade after decade for the necessity (or at least the desirability) of the randomized trials study. Moreover, if he is correct, there is no point in doing any other type of outcome study, because we can rely not only on subjects' retrospective evaluations of their subjective states at past times but on their attributions about what led to changes of state. Question #1, which accounts for 0-100 out of a total of 0-300 points on the criterion, reads (pg. 967): ""How much did treatment help with the specific problem that led you to therapy."

And it isn't even treated separately--and since the data are not "open" to other researchers, we have no idea how much variance in the total variance in the 0-300 total scale is accounted for by this attribution.

Marty's arguments are not matters of emphasis, approach, or shading. They are extraordinarily strong claims. If all--or even any--of these claims are correct, they should lead to a total revolution in the way research is conducted, perhaps even a change to not bothering to conduct it at all.

Robyn

Date: Tue, 30 Jul 1996 04:37:13 -0700 (MST) From: Lee B Sechrest <sechrest@U.Arizona.EDU> To: michael bailey <jm-bailey@nwu.edu> Subject: Re: value of Consumer Reports study
--

Lee Sechrest, Univ. of Arizona

On Mon, 29 Jul 1996, michael bailey wrote:
see above

Mike, you are correct in your views, which means, of course, no more than that I agree with you.

If Marty took only the position you outlined above, I could agree with him. But he goes a long way beyond that, and publicly too. As one example, there is a *considerable* disconnect between the view of the CR study as you have stated it and the use of the term "greed" to describe the decisions of HMOs and other payers not to reimburse for long term therapy. That is the position that Marty took in his public statement in asking for votes for APA president. To use the term greed in that context, and he has done so in other places, requires a much stronger interpretation of the CR study than you have outlined.

It is particularly disturbing that Marty lodges the "greed" charge so readily when the scientific community, let alone the health maintenance community, has not had a chance to digest the "good evidence" that would justify the charge of greed. The response to the CR study has not yet been published, and yet the results of the study apparently should somehow already have been known to mental health insurance payers and accepted by them as a basis for payment. In fact, I sort of remember Marty having made the greed and related statements on this network almost from the very beginning of the CR controversy, suggesting that his position was already established and that he expected that

everyone else should already have adopted it. In the inimitable words of Cool Hand Luke, what we have here is a failure of communications.

Date: Mon, 29 Jul 1996 20:47:13 -0700 (MST) From: Lee B Sechrest <sechrest@U.Arizona.EDU> To: Jim Wood <jwood@utep.edu> Subject: Re: value of Consumer Reports study

Jim,

you sort of stole my message, but here is what I was all ready to post. We have had proposals that (at least sound like) we should relax our standards of evidence. "Good" evidence can be constituted by nonrandomized designs with appropriate statistical or other allowances for what might otherwise seem to be deficiencies in the data. That is an especially interesting proposition, for it opens up possibilities for re-examination of enormous quantities of research that was not looked at with sufficient care because it "lacked" something. Those overly cautious concerns for the cleanliness of scientific research stemmed from perceptions of what it lacked, its "lacks" if you will. Research "lacked" randomization; comparison groups "lacked" comparability; measures "lacked" validity. But with new lax standards, those are clearly "ex-lacks." Which, perhaps, is fortunate, for the movements of science are often very slow and labored, resulting in only small amounts of really hard data. The delivery of such hard data, moreover, being accompanied often by a fair amount of noise, attesting fully to its infrequency. Think of the way scientists appear proudly at press conferences to display what they have managed to produce and for which they demand accolades, lest they decide not to do it again. But now, with all those deficiencies being taken care of by ex-lacks interventions, we can anticipate a veritable explosion of scientific increment. Scientists will be able, at last, to get their, ah, stuff together and let everyone see how much they can really produce when they get their behinds in gear. Now, for those of you who are worried about all this, I have just one bit of advice. Bring those standards back. Insist on them. Re-lacks. Lest you think I am kidding. Why should we suddenly decide that our standards of evidence, of what constitutes "good" evidence have changed? If we have changed in our standards, then we have a tremendous body of "old" research that is now subject to reinterpretation. All those studies with nonequivalent comparison groups are now available for reinterpretation. I recall reading any number of studies that were regarded as "lacking" because comparison groups were not regarded as equivalent. Well, all we need to do is go back and make some reasonable estimates of parameters involved in the putative differences, and we can produce a wild explosion of "new" knowledge. Think about it. We have all those studies on race differences, flawed only by the fact that one cannot, obviously assign subjects randomly to re of one race or the other. So we just model out all those irrelevant differences and whatever is left over, perforce, must be an effect of race. Same for sex differences. Males and females differ in productivity after we have "modeled out" all the differences on "control" variables? Well, then the sexes, bless them, must differ fundamentally in productivity--scientific, or any other kind. We will now have "good" evidence that specific therapist training and experience do not matter. We will have good evidence that short term therapy is as effective as long-term therapy (the CR study is not the only study with a nonrandomized design that can be brought to bear on the problem. Hell, we can probably even show that psychoanalysis should become the national sport. I keep trying to make this point, but it seems that almost no one will listen (well, Robyn, Jim Wood, Mike Bailey, and a few others will), we cannot have a science that persistently changes the rules of evidence according to whose argument is being made and who is making it. Larry Beutler may be right that we are not completely consistent in our standards of evidence, but that is just exactly what I have in mind. We *do*, we *must* have standards of evidence. (Without them we would not be able to know, as Larry does,

that we violate them, fudge them, now and then and that we owe apologies for doing so.) I STILL have never had a reply to my question how we will distinguish "good" evidence for the effectiveness of psychotherapy from "good" evidence for effectiveness of using a crystal as a suppositor intended to bring on enlightenment.

Date: Mon, 29 Jul 1996 13:54:15 -0700 (PDT) From: Larry Beutler <beutler@education.ucsb.edu> To: Jim Wood <jwood@utep.edu> Subject: Re: value of Consumer Reports study
--

On Mon, 29 Jul 1996, Jim Wood wrote:
see above

I think in all fairness that we need to put some of the issues that Jim and others have raised into perspective. Aside from the fact that I Personally don't find the J & C critique either that devastating or the CR study so lacking as some (I'm sure Lee will have a comment or two about my failure), there are two issues that Jim raised should be given another look. The failure of the CR Editors to open their data to inspection is not unique. While APA has encouraged scientists to open up their data sets to scrutiny, that has not been a requirement for some very good reasons.

For example, some IRBs do not allow data, even masked, to be distributed except when a question arises about the ethics of the research and then, only with special oversight. Their concerns have been subject/participant protection. There are also many other private organizations that maintain a clear restriction on release of their data because of proprietary reasons. These include drug companies, health care companies, etc. Editors must contend with this issue frequently during the review process.

The CR study shouldn't be singled out for a problem that characterizes a good deal of research in the field, whether we like it or not or agree with it or not. The fact remains that a great deal of the data gathered is not and will not be available for inspection by other scientists unless there are charges of fraud or ethical fault.

Second, the ad hominem attack on Marty is a bit beyond the limits of good dialogue. There are some legitimate questions being raised here, and because one's undergraduate students (bred and trained by a given standard) can see the problems in the same way as their professor, does not make those issues unimportant. The CR study has now undergone a good deal of peer review. It went through peer review in its design. The fact that it was not published in a peer reviewed journal is beside the point. Is it good science? Some would say no and others would say yes. That type of disagreement is the basis of scientific argument and we need not resort to attacking the credentials and good sense of those who differ in their opinions from us. It seems to me that we are addressing some very strong and rigid opinions here. Those on both sides are prone to use a double standard of evidence. Those of my ilk are accused of arguing away ns findings and being soft on positive ones, when we discuss either effectiveness research, ATI, or therapist experience. But, double standards are also present when one ignores what many of us believe are compelling lines of evidence to support (1) the value of therapist experience among the most severely disturbed patients, (2) the necessity of using effectiveness designs in studying long term treatments because of all the reasons Marty detailed, (3) the roles of other patient moderators in therapy effectiveness, etc. We see a logical flaw of arguing only from the negative. We find evidence of a double standard when every instance of a positive effect is dismissed, seemingly out of hand, because of what are often inevitable design flaws in order to accept a null hypothesis that derives from very weak studies that find ns results. The issue of experience of therapist again comes to mind.

LE Beutler

Date: Mon, 29 Jul 1996 14:05:22 -0400 From: seligman@cattell.psych.upenn.edu (Martin E. P. Seligman) To: sscpnet@bailey.psych.nwu.edu Subject: Scientific Standards & SSCP

Jim:

It is most certainly not my intention to lower scientific standards and were that the result of my apa presidency or of my own scientific contribution, I would judge both failures. SSCP will, I hope, play a constructive role in both keeping standards impeccable and at the same time investigating empirically a set of issues crucial to practice. I trust that open debate in this forum, such as your exhortation, will keep clinical science on course--although I do think that illumination is best achieved when ad hominem attacks are minimized.

First my comments on the substance of where you take issue with me:

1) You say the Jacobson & Christensen critique of my AP article is 'devastating.' I have already responded on the net to its main critiques--that the CR article supports the idea that qualifications do not improve outcome, that the CR article found nothing new, and that the benefits of long term therapy were spontaneous recovery.

Each of these arguments is I believe incorrect. I have sent you an electronic preprint of my response in the October AP, and I will be happy to send it to others on request. (You wonder why AP published my December 1995 article in the first place. You should be aware that both Neil Jacobson and Lee Sechrest generously reviewed drafts for me, and as I recall, all of their suggestions were incorporated.)

2) You say I "minimize" the flaws of the CR article. Both in the December AP and in the forthcoming October issue, I have been at pains to point out the problem of possible selection bias, the need for valid and more complete measurement, and the flaw of having only cross-sectional retrospective report. I end my October article by calling for studies of long term therapy free of such selection bias, longitudinal in nature, and measured by blind diagnosis and full assessment battery, as well as patient report. And I take exception in print to CR's withholding the dataset, and I will do the same in my remarks at the convention.

3) You equate using "naturalistic methods," regression and causal modelling with a lowering of standards. Here you are dead wrong. Both naturalistic methods and experimental methods can be used sloppily or rigorously. I have consistently advocated using both experimental and naturalistic methods--rigorously--to investigate therapy. They complement each other, and neither is inherently superior methodologically. Perhaps you can explain how astronomy has done rigorous science without experimental methods. Perhaps you can explain the belief that cigarette smoking causes lung cancer--the vast bulk of the evidence uses only naturalistic methods. Perhaps you can explain how we have come to believe that a good part of psychopathology and personality are heritable--with no experimental methods.

Lastly my view of where the 'scientific leadership' of APA is heading.

I am looking to SSCP to help provide the scientific leadership for APA. At best I can help organize that leadership and make its force felt. That is why I asked SSCP to read this debate carefully and then generate the designs--both experimental and naturalistic--that will answer the questions pressing most heavily on practice--at the highest level of scientific standards.

I have concerned myself with both highlighting the methodological flaws of the CR article, as well as defending its method when the arguments against it were poor, so that better studies--using higher standards of evidence--will emerge.

It is my conviction that science is the bedrock on which practice rests and the best friend that practice has. At this critical juncture, science can become the critical ally of practice. It can do this in two ways: first by continuing the debate about efficacy versus effectiveness that we are having. Second by designing and carrying out the studies needed to speak to the survival issues raised by practice's confrontation with MCO's.

Date: Tue, 30 Jul 1996 07:50:33 -0700 (PDT) From: Larry Beutler <beutler@education.ucsb.edu> To: Robyn Dawes <rd1b+@andrew.cmu.edu> Subject: A Question of Standards

On Tue, 30 Jul 1996, Robyn Dawes wrote:
see above

Robyn,

You're right! The default position should be to share data. But, to insist on throwing out a study specifically because it doesn't share data certainly introduces a different standard for these data than for others. If that were among the criteria of truth that we decided to accept in response to Lee's appeal, we would be introducing a serious non-scientific contaminant. Having masked data does not solve the problem for many IRBs. I know of several instances in which an IRB determined that to share data with another investigator required that each participant (in one case several hundred archival patients) be recontacted to get permission since the new analysis went beyond that for which the participants had given permission. I think Lee is right-- what we lack is some specific protection against using a double standard. We can and do all find ways of interpreting data to fit our private and personal predilections. It is a bias about which we know much but seldom recognize as being embedded within our own arguments. The standard or criteria for accepting evidence of there being differences in long and short term psychotherapy should be the same as that for accepting EMDR, Tarot card reading, pre-cognition, etc.

I argue, however, that those criteria should address the very sticky issue that not all evidence is not defined by a $p < .05$ statistic. They should also include some way of including the context of the data within a history of research, the logic of the results, the fit with other (e.g., sources of evidence (such as clinical observation), and even their social value (dare I say that?). Those abstract concerns are the ones on which we differ and that give rise to different interpretations from one scientist to another.

I don't believe that the answer is to exclude these aspects of understanding and evaluating evidence, as we currently try to do. Instead, I think there is need to find ways to incorporate these sources of interpretation, balance or weight them, and use them consistently and intentionally. We have failed, as have most other fields, in creating a value free science. It's time to accept the value differences and use them or study them, not to pretend they are unimportant as sources of evidence.

Larry

Date: Tue, 30 Jul 1996 09:43:44 -0400 (EDT) From: Robyn Dawes <rd1b+@andrew.cmu.edu> To: Jim Wood <jwood@utep.edu>, Larry Beutler <beutler@education.ucsb.edu> Subject: Re: A Question of Standards
--

Larry,

But the "default" option is and should be to share data.

Where is the compelling reason for CU not to do so? Responses to their polls can be entirely anonymous. (I know, as a responder, that names are not required.)

Robyn

Date: Tue, 30 Jul 1996 10:43:43 -0900 (PDT) From: Jim Wood <jwood@utep.edu> To: Larry Beutler <beutler@education.ucsb.edu> Subject: A Question of Standards
--

I'd like to recommend a very thoughtful article regarding accessibility of data. Its specific focus is on research findings that feed into the policy making process, as the CR study has:

Terry E. Hedrick, Robert F. Borch, & Jerry Ross. (1978). On ensuring the availability of evaluative data for secondary analysis. *Policy Sciences*, 9, 259-280.

The article was also re-printed in Lee Sechrest et al. (eds). (1979). *Evaluation Studies Review Annual*, volume 4, pp. 385-407.

For list members who are interested, here are the opening paragraphs of the article:

In 1975 an economist at a prominent university published an article which claimed that between 1935 and 1969 each execution in this country may have deterred between seven and eight murders (Ehrlich, 1975). The claim of a deterrence effect for capital punishment was itself noteworthy in diverging from earlier investigators' findings. The research was even more remarkable for its rapid introduction into policy making process. At the time the results were published, the Supreme Court was reconsidering its 1972 decision declaring capital punishment unconstitutional. In an amicus curiae brief filed by the Justice Department in *Fowler v North Carolina*, Solicitor General Bork used the economist's results to argue for reinstitution of capital punishment. During this period, the data underlying the original article were not manifestly available to other interested researchers. Consequently, it was impossible for outside analysts to determine how well (or how poorly) the deterrence claim was justified based on reanalysis of identical data.

The preceding example, along with other cases, illustrates the need for an open access policy to evaluative data. Outside analysts should have access to any data which feeds into the policy making process so that the initial findings can be reexamined, and in particular, verified or refuted as quickly as possible. This paper focuses on that need and on mechanisms for assuring that competing analyses of policy relevant data are possible....

(End of quote)

My own feeling is that Hedrick, Boruch & Ross make a good point. Because the CR study was quickly and forcefully introduced into policy debates about managed care and funding

for psychotherapy, there were particularly strong reasons to want the data available for immediate scrutiny.

Jim Wood
University of Texas at El Paso

Date: Tue, 30 Jul 1996 12:08:32 -0800 From: "Gregory J. Meyer" <afgjm@UAA.ALASKA.EDU> To: sscpnet@bailey.psych.nwu.edu Subject: Dose-Response and Self-report Bias

Several people have noted how the dose-response effect in the CR study may be an artifact of self-report bias.

Simply as a function of cognitive dissonance, patients who invest more time and more money in their treatment may be more prone to say treatment was effective. This may be a particular problem for retrospective designs, although it is also a problem with prospective designs as patients may feel similarly compelled to report an overly rosy view of symptom reductions. As Seligman pointed out in the original CR paper, optimal scientific knowledge would be served by using pre- and post- measures that are not biased by self-reports.

Rightfully, he indicated interview generated ratings would not fit the bill, since self-reports form the "blood and guts" of this source of information too. Thus, unless one uses the kind of "hard" criteria alluded to by Don Klein, namely the ability to keep a job, rates of rehospitalization, etc., one is in a real bind when it comes to evaluating the extent of therapeutic change uncontaminated by self-report. Unfortunately, "hard" criteria are not equally applicable to or available for all patient groups.

To get around this problem, I think it would be wise to use the Rorschach procedure as an alternative source of data. The reason for this is simply because Rorschach constructs are independent of similarly named self-report constructs in heterogeneous patient samples. That is, in general patient samples, Rorschach depression is uncorrelated with self-reported depression, Rorschach-measured interpersonal problems are uncorrelated with self-reported interpersonal problems, etc. #

I believe the reasons for this independence are complex and won't go into them here. Regardless of the reasons, however, the independence of these methods allows a test of the dose-response hypothesis without the potential confounds that may be introduced by self-report. Although some have argued that the independence of Rorschach and self-reported constructs should be seen as an indication of Rorschach invalidity, I'm suggesting this independence is a strength that can be put to good use.

Two recent studies provide Rorschach data on the relationship between therapy length and effectiveness. Although its a bit lengthy for a post to the list, I organized this data for other purposes and I'll summarize the findings reported in Weiner & Exner, 1991, JPA, 56, 453-465, and Exner & Andronikof-Sanglade, 1992, JPA, 59, 59-71 for those who may not be familiar with the data.

Both of these studies are longitudinal investigations of naturalistic treatment and consequently suffer from the limitations that have been discussed here over the last couple of weeks. Further, the data analysis was oriented towards understanding the Rorschach's sensitivity to treatment changes. Consequently, the results simply assessed changes in the frequency of certain scores which are believed to indicate psychological trouble.

In the first study, Weiner and Exner (1991) recruited patients seeing 53 U.S. therapists. Rorschachs were obtained at treatment outset and three follow-up periods, the last about four years after treatment began. At baseline, patients were classified as long-term if they were to be seen more than once per week in dynamically oriented treatment, and short-term if they were to be seen once per week in rational emotive, gestalt, modeling, or assertiveness types of therapy. All 88 patients in long-term therapy were used in the data analysis and a random sample of 88 short-term patients were drawn from a pool of several hundred. Testing was done independent of treatment by 61 examiners, none of whom tested the same patient more than once.

In terms of baseline characteristics, patients in short-term treatment more often had a low SES classification (43% vs 20%), and more often reported depression as their primary complaint (30% vs 20%) although there were no other differences in 5 other symptom areas (anxiety, tension, emotional dyscontrol, interpersonal difficulties, somatic problems). Marital status, age, gender, source of payment, and setting of treatment also did not differ between groups.

The data reported below represent rough effectiveness rates for 22 conditions. Each condition is quantified by a Rorschach score which has a specified cut-off point, beyond which disturbance is generally indicated. Averaged across all scores, 8.3% of the CS normative sample falls above each cut-off. Thus, on average, these cut points would be about one and a half standard deviations from the mean (i.e., T-score of about 65) if the scores were normally distributed, which most are not.

The first two columns of data reported below indicate the incidence of the problem at baseline testing. The last two columns indicate the proportional reduction in incidence at follow-up testing four years after baseline testing (i.e., roughly the percent of initially disturbed patients who achieved average functioning at retest). At follow-up, short-term patients had received an average of 62.1 sessions and long-term patients had received an average of 452 sessions.

If effectiveness is defined as a statistically significant change in the incidence of a problem ($p < .01$), then long-term treatment is effective for 21 conditions while short-term therapy is effective for 12.

Table 1. Rorschach scores indicating the differential effectiveness of short-term and long-term therapy after four years: Rates of proportional improvement to normal functioning.

In Incidence Problematic months Condition	% Symptomatic at		% Reduction	
	Baseline		at 46-50	
	Short	Long	Short	Long
1. Chronic difficulty managing internal reactions	23	31	50	85*
2. limited resources	24	34	43	90*
3. generalized coping deficits	35	32	61*	82*
4. inconsistent coping	43	36	32	78*
5. hasty processing	32	39	69*	71*
6. narrow awareness	25	29	36	80*
7. impaired perceptions	39	30	54*	73*
8. affective distress > ideational stress	32	32	72*	97*

9. depression	59	60	67*	77*
10. emoti withdrawal	35	34	55*	70*
11. uncontrolled affect	65	60	59*	79*
12. disjointed thinking	24	31	24	63*
13. impaired perception of people	38	47	39*	85*
14. excessive fantasy	35	42	39	73*
15. intellectualization	16	24	-07	71*
16. intrinsic narcissi.	11	14	00	50
17. self-absorption	33	27	62*	67*
18. low self-regard	34	32	67*	54*
19. passivity	26	34	48	67*
20. lack of relatedness	23	31	24	74*
21. lonely distress	29	22	73*	89*
22. poor identification w/ people	42	31	38*	93*

Note: Ns are 88 in each group. All data derived from Weiner & Exner (1991). Rorschach constructs were operationally defined as follows: 1 = Adj D < 0; 2 = EA < 7; 3 = CDI > 3; 4 = Ambitient EB; 5 = Zd < -3.0; 6 = Lambda > .99; 7 = X-% > .20; 8 = Sum of all shading > FM + m (typo in original corrected); 9 = DEPI > 4; 10 = Afr. < .5; 11 = CF + C > FC + 1; 12 = Sum of Six Special Scores > 6; 13 = M- > 0; 14 = Mp > Ma; 15 = Intellectualization Index > 5; 16 = reflections > 1; 17 = Egocentricity Index > .43; 18 = Egocentricity Index < .33; 19 = p > a; 20 = T = 0; 21 = T > 1; and 22 = human content < 2. * = p < .01 (chi-square contrasting baseline and retest incidence)

The second study (Exner & Andronikof-Sanglade, 1992) used 35 patients seen by 16 therapists in various forms of brief therapy, described as supportive, eclectic, or directive. Rorschach protocols were obtained at baseline, at termination 3-4 months later, and again 8-12 months after treatment had started. These protocols were compared to 35 patients selected from the large pool of patients who had received short-term therapy. The pool of short-term patients was initially stratified to match the brief therapy patients for presenting complaints and key demographic variables and then patients were randomly selected from within these strata. The short-term patients were seen by 22 different therapists. All testing for both groups was done by 22 examiners and no patient was tested by the same examiner twice.

Since the brief group was obtained with different sampling procedures than the short-term group, and since both groups only contain 35 subjects, results should be considered cautiously.

The following table reports data for baseline testing and the one year follow-up. Patients in brief treatment had received an average of 14.2 sessions, while those in short-term treatment received an average of 47 sessions.

Using the criterion of a statistically significant change ($p < .01$) in incidence, the data suggest short-term treatment is effective for eight conditions while brief therapy is effective for only four.

Table 2. Rorschach scores indicating the differential effectiveness of brief and short-term therapy after one year: Rates of proportional improvement to normal functioning.

In Incidence Problematic months Condition	% Symptomatic at		% Reduction	
	Baseline		at 8-12	
	Brief	Short	Brief	Short
1. Chronic difficulty managing internal reactions	14	26	00	66
2. limited resources	31	43	09	47
3. generalized coping deficits	23	31	13	73
4. inconsistent coping	51	54	06	58*
5. hasty processing	43	34	73*	75
6. narrow awareness	34	29	33	80*
7. impair. perceptions	17	26	17	67
8. affective distress > ideational stress	71	57	84*	60*
9. depression	54	51	37	61*
10. emot. withdrawal	40	43	57*	67*
11. uncontroled affect	66	60	44*	48*
12. disjointed thinking	20	20	14	43
13. impaired perception of people	29	40	10	57
14. excessive fantasy	40	43	14	60
15. intellectualization	26	23	-11	87
16. intrinsic narcissism	11	14	00	20
17. self-absorption	37	40	46	36
18. low self-regard	49	46	29	62
19. passivity	31	37	09	69*
20. lack of relatedness	51	43	06	67*
21. lonely distress	20	23	71	87
22. poor identification w/ people	23	31	13	73

Note: Ns are 35 in each group. All data derived from Exner & Andronikof-Sanglade (1992). Rorschach constructs were operationally defined as in the previous table.

* = $p < .01$ (chi-square contrasting baseline and retest incidence)

Importantly, this study also provided self-report data.

Unlike Rorschach scores, these data did not portray the same pattern of differential effectiveness. Instead, both the brief and short-term groups reported significantly fewer problems at the follow-up across all domains evaluated, including anxiety, depression, interpersonal functioning, emotional control, participation in activities, and satisfaction with life activities.

While these studies are certainly limited and the samples are much smaller than the CR sample, the studies provide an important counterpoint to the CR survey because they are prospective and use a DV that is generally uncorrelated with self-ratings.

Mirroring the CR results, a dose-response relationship is evident in both studies at the final retest. Rorschach scores indicate the length of therapy affects both the degree and scope of effectiveness. Compared to short-term treatment, long-term dynamic therapy appears to be more effective for more people, and more effective across a broader range of personality problems. In contrast, brief supportive or directive treatment does not appear to help most patients in most areas of functioning.

I find the discrepancy between self-reported improvement and Rorschach indices of improvement to be the most interesting aspect of these studies. It may be that the Rorschach is less sensitive to the change that is consciously experienced by patients. Alternatively, it may well be that self-ratings give a global and undifferentiated picture of change that is biased by selective perception or other demand characteristics that are engaged in the research setting.

In either case, I believe the data demonstrate we could improve our science if we regularly used multiple methods of assessment to quantify the effectiveness and efficacy of treatment.

Greg

Gregory J. Meyer
Department of Psychology
University of Alaska Anchorage

phone: 907-786-1741
fax: 907-786-4898
afgjm@uaa.alaska.edu

Date: Tue, 30 Jul 1996 08:41:35 -0700 (PDT) From: Gerald Rosen <groser@u.washington.edu> To: Lee B Sechrest <sechrest@U.ARIZONA.EDU> Subject: Tushiology

On behalf of the entire Board for the Institute of Tushiology I want to congratulate Dr. Sechrest for his analysis and offer him an honorary position on the Board. I know it's been a few months since any of you thought about our Institute, but believe you me, with contributions like those of Dr. Sechrest, the Institute is alive and well.

I know it's kind of tacky to promote the Institute this way, but I do want to take the opportunity to let all of you know that we have a self-instructional packet that allows you to take a photo of your derriere which can be mailed in to the Institute for analysis of your personality. We call the method the Hind-Look Maneuver, as suggested by a loyal follower.

Sincerely,
Jerry Rosen
Director, Institute of Tushiology
Editor, Anals of Tushiology

Letters
from
SSCP net_{No 10 Juli 98}

A heated discussion
EMDR

Francine Shapiro, Larry Beutler,
RichMc Nally and others

ed
horst u kaechele

Am Freitag, den 7. Juli 1995, polemisierte der inzwischen bei uns hoch geschätzte Martin Seligman zum Thema "TFT and EMDR" im SSCP Netz.

At the risk of offending some dear friends and valued colleagues, I confess to bemusement and annoyance at the abuse heaped on Charles Figley for spelling out the details of TFT, and suggesting people try to validate it scientifically.

Both TFT and EMDR have poor face validity--you have to be made of stone not to laugh at the hokey-pokey jibe--but you all know that face validity is not a very good criterion of what works.

When I was a psychiatric resident--that's not a typo--the residents roared with laughter about the following Arnold Lazarus story:

"Did you hear what Lazarus did in therapy recently?" "He had a patient who thought his penis was too small..." "So he had him measure it!"
Ho Ho Ho.

It's not as if we have loads of face-valid treatments that work on PTSD.

It's not as if there is a good theory about the mechanism of PTSD.

It's not as if Figley and Shapiro are not seasoned clinicians with intriguing, preliminary evidence of efficacy. And one seemingly good outcome study, BTW, Wilson, Becker, and Tinker, in press, JCCP.

So let's hold our guffaws, try some outcome studies as Figley and Shapiro urge, and then decide--over martinis.

Kürzlich hat sich erneut eine interessante Diskussion ergeben, die ich den deutschen Lesern der Letters from SSCNet nicht vorenthalten möchte.

horst u kaechele

Date: Wed, 22 Jul 1998 14:13:56 -0700 (PDT) From: fshapiro@emdr.org (Francine Shapiro) To: sscpnet@listserv.acns.nwu.edu Subject: Exner System/EMDR and exposure

As noted in a review of the EMDR literature by Feske (1998) the component analyses of EMDR are at best equivocal regarding the eye movement or other stimulation. Unfortunately, one of the problems with the Lohr, Tolin & Lilienfeld (1998) review is that they cite component analyses but fail to take into account external/clinical validity factors. I have pointed out these types of problems in a previous commentary (Shapiro, 1996b). To stipulate that the eye movements have been "proved superfluous" one must argue that the following are suitable conditions to test the components of a complex clinical method:

- 1) Students with disturbing memories (i.e., analogues)
- 2) Multiply traumatized combat veterans (using only two sessions and/or addressing only one or two memories)
- 3) An insufficient n to have the conventional 80% power to detect even medium, let alone fine, differences.

All of the controlled component analyses of EMDR (with positive or negative findings regarding stimulation) fit into these categories.

EMDR includes a wide variety of elements in addition to a form of dual stimulation (see Shapiro, 1995, in press). Rigorously controlled component analyses that test the method as it is used in clinical practice with a suitable PTSD population and number of subjects will be a welcome addition to the literature. Obviously, replications by the originator are unconvincing. Instead, I've delineated parameters of investigation repeatedly since 1995 (Shapiro, 1995, 1996a, 1997, 1998, in press).

In the meanwhile, the literature does not support the notion that the effects of EMDR can be attributed to exposure as it is traditionally employed.

EMDR's effectiveness with diagnosed civilian PTSD has been documented in eight controlled studies. The four most recently published rigorously controlled studies use the same standard psychometrics and independent assessors as the exposure studies. Two waitlist (Rothbaum, 1997; Wilson et al., 1995, 1997) and two comparative (Marcus et al., 1997; Scheck et al., 1998) controlled studies document that 80-100% of the subjects lose the PTSD diagnosis in 3-5 hours of treatment. Comparable results with exposure techniques have been documented for PTSD only with a combination of imaginal and in vivo exposure after 50-100 hours of combined treatment sessions and homework (Richards et al., 1994; Marks et al., 1998). The one peer-reviewed controlled study of flooding alone with civilian PTSD (Foa et al., 1991) found that after 7 sessions and homework (@25 hours of exposure) 28.6% had dropped out and 45% still had PTSD.

Also, as noted in previous posts, "In strict exposure therapy the use of many of [a host of EMDR-essential treatment components] is considered contrary to theory. Previous information also found that therapists and patients prefer this procedure over the more direct exposure procedure" (Boudewyns & Hyer, 1996, p. 192). I have included a reference list below for those who are interested in coming to their own conclusions.

References

- Boudewyns, P.A. & Hyer, L.A. (1996) Eye movement desensitization and reprocessing (EMDR) as treatment for post-traumatic stress disorder (PTSD). *Clinical Psychology and Psychotherapy*, 3, 185-195.
- Carlson, J.G., Chemtob, C.M., Rusnak, K., Hedlund, N.L., & Muraoka, M.Y. (1998). Eye movement desensitization and reprocessing for combat-related posttraumatic stress disorder. *Journal of Traumatic Stress*, 11, 3-24.
- Feske, U. (1998) Eye movement desensitization and reprocessing treatment for posttraumatic stress disorder. *Clinical Psychology: Science and Practice*, 5, 171-181.
- Foa, E.B., Olasov Rothbaum, B., Riggs, D.S. Murdock, T.B. (1991) Treatment of posttraumatic stress disorder in rape

victims: a comparison between cognitive behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology*, 59, 715-723.

Lohr, J. M., Tolin, D. F., & Lilienfeld, S. O. (1998). Efficacy of Eye movement desensitization and reprocessing: Implications for behavior Therapy. *Behavior Therapy*, 29, 123-156.

Marcus, S. , Marquis, P. & Sakai, C. (1997) Controlled study of treatment of PTSD using EMDR in an HMO setting. *Psychotherapy*, 34, 307-315.

Marks, I, Lovell, K., Noshirvani, H., Livanou, .& Thrasher, S. (1998) Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring. *Archives of General Psychiatry*, 55, 317-325.

Richards, D.A., Lovell, K. & Marks, I.M. (1994) Post-traumatic stress disorder: Evaluation of a behavioral treatment program. *Journal of Traumatic Stress*, 7, 669-680.

Rothbaum, B.O. (1997). A controlled study of eye movement desensitization and reprocessing for posttraumatic stress disordered sexual assault victims. *Bulletin of the Menninger Clinic*, 61, 317-334.

Scheck, M.M., Schaeffer, J.A. & Gillette, C.S. (1998) Brief psychological intervention with traumatized young women: The efficacy of eye movement desensitization and reprocessing. *Journal of Traumatic Stress*, 11, 25-44.

Shapiro, F. (1995). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures*. New York: Guilford Press.

Shapiro, F. (1996a). Eye movement desensitization and reprocessing (EMDR): Evaluation of controlled PTSD research. *Journal of Behavior Therapy and Experimental Psychiatry*, 27, 209-218.

Shapiro, F. (1996b). Errors of context and review of eye movement desensitization and reprocessing research. *Journal of Behavior Therapy and Experimental Psychiatry*, 27, 313-317.

Shapiro, F. (1997). Eye movement desensitization and reprocessing (EMDR): Research and clinical significance. In P. Gosselin and W. Matthews (Eds.) Current thinking and research in brief therapy: Solutions, strategies, narratives. New York: Brunner Mazel.

Shapiro, F. (1998). Eye movement desensitization and reprocessing (EMDR): Historical context, recent research, and future directions. In L. Vander creek & T. Jackson (Eds.) Innovations in Clinical Practice: A Source Book (Vol. 16). Sarasota, FL: Professional Resource Press.

Shapiro, F. (in press) EMDR and the Anxiety Disorders: Clinical and Research Implications of an Integrated Psychotherapy Treatment. Journal of Anxiety Disorders.

Wilson, D., Silver, S.M, Covi, W., & Foster, S. (1996). Eye movement desensitization and reprocessing: Effectiveness and autonomic correlates. Journal of Behavior Therapy and Experimental Psychiatry, 27 , 219-229.

Wilson, S.A., Becker, L.A., & Tinker, R.H. (1995). Eye movement desensitization and reprocessing (EMDR) treatment for psychologically traumatized individuals. Journal of Consulting and Clinical Psychology, 63, 928-937.

Wilson, S.A., Becker, L.A., & Tinker, R.H. (1997). Fifteen-month follow-up of eye movement desensitization and reprocessing (EMDR) treatment for PTSD and psychological trauma. Journal of Consulting and Clinical Psychology, 65, 1047-1056.

Francine Shapiro, Ph.D. Senior Research Fellow Mental Research Institute 555 Middlefield Road Palo Alto, CA 94301 fax: 650/728-0654 FShapiro@emdr.org

Date: Wed, 22 Jul 1998 19:01:25 -0400 (EDT) From: "Richard J. McNally" <rjm@wjh.harvard.edu> To: Francine Shapiro <fshapiro@emdr.org> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Francine:

The burden of proof lies on those claiming that eye movements enhance the efficacy of imaginal exposure. That is, it is not a matter of "proving" that eye movements are superfluous; rather it is a matter of "proving" that they actually do something positive.

Skepticism about EMDR constituting anything more than a variant of exposure therapy will be dispelled only by evidence showing that its distinctive ingredient enhances the efficacy of desensitization. Extant studies, reviewed by Lohr et al., fail to support the hypothesis that eye movements possess this special therapeutic power. You question the probative import of these studies on methodological grounds. But are studies adduced in support of EMDR of superior methodological quality than those failing to confirm it as a "breakthrough therapy" and "miracle" (quotes from dust jacket of Shapiro & Forrester, 1997 book)?

Finally, in your original EMD study (Shapiro, 1989) a 100% success rate for reduction of distress associated with traumatic memories was reported in a single session in combat veterans (among others). Hence, I do not understand why you say that studies on "Multiply traumatized combat veterans (using only two sessions and/or addressing only one or two memories)" are inappropriate for testing the efficacy of EMDR. Indeed, it was **precisely** such a study that launched the method into the limelight in the first place! Moreover, what is even more puzzling is that these outstanding results, accomplished with the presumably obsolete EMD method, have never been replicated with the new, improved EMDR.

We agree that statistical power issues are important in psychotherapy evaluation. But EMDR is described as a "miracle", not as an ordinary run-of-the-mill therapy. Hence, one might expect it to surmount the pesky power problems that mask the effects of nonmiraculous psychotherapies.

Rich

From: fshapiro@emdr.org (Francine Shapiro) To: "Richard J. McNally" <rjm@wjh.harvard.edu> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure
--

Rich:

-----answers below. However, first let me clarify for those not familiar with PTSD treatment that the theory, research, and practice of exposure therapy mandates prolonged exposure to the highest level of anxiety possible without distraction and the need for daily homework as part of the treatment. This is explicitly stated in the two imaginal/in vivo exposure studies (Richards et al., 1994; Marks et al., 1998) I mentioned that match the effects of 3-5 hours of EMDR treatment with 50-100 exposure hours of combined treatment and homework.

Francine:

>

>The burden of proof lies on those claiming that eye movements enhance the efficacy of imaginal exposure.

-----The "burden of proof" in science came first in showing that EMDR was effective for PTSD. That has now been done. Component analyses generally come after that has occurred. I'd rather believe that dispassionate researchers would conduct appropriate component analyses to assist in refining the clinical procedures and protocols--rather than an "I and thou" situation of proponents and antagonists.

>That is, it is not a matter of "proving"
>that eye movements are superfluous; rather it is a matter of "proving"
>that they actually do something positive.

-----EMDR is a complex clinical method and only appropriately conducted component analyses will settle that question. Until that time, however, conclusions such as your on-line statements are misleading and unwarranted (i.e., "... it appears that the defining element of EMDR - eye movements - do not add to the efficacy of imaginal exposure (Lohr, Tolin, & Lilienfeld, 1998). Hence, what is effective in EMDR is not new, and what is new is not effective.") Further, the distinctive combination of elements that achieve 84-100% PTSD remission in only 3-5 hours is in itself "new"--particularly compared to the only civilian PTSD imaginal exposure study (Foa et al., 1991) which used @25 hours of exposure to achieve a 55% success rate.

>Skepticism about EMDR constituting anything more than a variant of exposure therapy will be dispelled only by evidence showing that its

>distinctive ingredient enhances the efficacy of desensitization.

-----As I've repeatedly stated, EMDR is a complex method that integrates elements of many of the major psychological orientations (see Shapiro, 1995). The targets looked for in the history-taking, the form of preparation, the delineation of components in the assessment, the procedures used during processing, etc. are all "distinctive" ingredients--in a unique combination. However, if you believe that EMDR is merely a "variant of exposure therapy," then by all means, use it without the stimulation component. The bottom line is the alleviation of suffering, not dogma.

I've always been curious though: How is EMDR to be defined as an exposure therapy if the rapidity of effects cannot be predicted by the extinction/habituation model and many of the procedures are in direct contradiction to it? For instance, in contrast to EMDR, prolonged (rather than brief) exposure is considered to be the most effective form of treatment (Chaplin & Levine, 1981; Foa, Steketee, & Rothbaum, 1989; Lyons & Scotti, 1995; Marks, 1972). As noted by Marks et al. (1998), "Invertebrates and invertebrates, exposure gradually reduces defensive responses to cues to which the subject is exposed; this habituation depends on the dose of exposure. Continuous stimulation in neurons and immune and endocrine cells tends to dampen responses, and intermittent stimulation tends to increase them" (p. 324). According to this criterion, EMDR's intermittent exposure to the traumatic cue should sensitize, rather than desensitize, the client. Since the EMDR research contradicts that notion, I'd appreciate understanding your reasoning.

>Extant studies, reviewed by Lohr et al., fail to support the hypothesis that eye

>movements possess this special therapeutic power. You question the

>probative import of these studies on methodological grounds.

But are

>studies adduced in support of EMDR of superior methodological quality than

>those failing to confirm it as a "breakthrough therapy" and "miracle"

>(quotes from dust jacket of Shapiro & Forrest, 1997 book)?

----- There are component analyses that are both positive and negative--they all fall short. Once again, there is a need for rigorous science rather than invective. I don't recall any civilian

PTSD EMDR studies that failed to show positive effects more rapidly (with fewer exposure hours) than traditional exposure techniques--do you? As far as the most rigorous research is concerned: Isn't a 3-5 hour EMDR effect compared to a 50-100 hour prolonged exposure effect reasonably called a "breakthrough"? After all, that's the way a past president of AABT described it when he introduced an EMDR symposium. I'm sorry if Basic Book's selection of media quotes bothers you. I believe it was Barbara Walters who waxed eloquent about the rapidity after a 20/20 investigative report and said something like "miracle though it may seem." It seems for others "too good to be true" is more acceptable parlance. However, neither classification should deflect rigorous scientific study.

>
 >Finally, in your original EMD study (Shapiro, 1989) a 100% success rate
 >for reduction of distress associated with traumatic memories was reported
 >in a single session in combat veterans (among others).
 Hence, I do not >understand why you say that studies on "Multiply traumatized combat
 >veterans (using only two sessions and/or addressing only one or two
 >memories)" are inappropriate for testing the efficacy of EMDR. Indeed, it
 >was *precisely* such a study that launched the method into the limelight
 >in the first place! Moreover, what is even more puzzling is that
 >these outstanding results, accomplished with the presumably obsolete EMD
 >method, have never been replicated with the new, improved EMDR.

-----I'm happy to clarify the situation for you. All studies of combat veterans with EMDR have shown a reduction in SUD levels. The two combat veterans included in my original study evinced a reduced level of anxiety on the SUD scale--but no other standard psychometrics were used and only one symptom was evaluated (and within the original article I noted the need for more comprehensive training to attain the reported results--"EMD" was left to refer to the 1989 article's 2.5 pages of procedural instructions). The other EMDR studies with combat veterans have additionally tested the method with global psychometrics that have never been found to change when a small number of memories are addressed (Fairbank &

Keane, 1992). It is clinically inappropriate to assume that global psychometrics will substantially change (or be maintained) when only one or two out of @12-24 memories are addressed. *Certainly,* this is not the appropriate condition for the fine discriminations necessary for component analyses. Do you disagree?

The single-trauma (civilian) EMDR literature shows that three sessions are standard for the 84-100% success rate on standard psychometrics and full symptomatology. Clearly, combat veterans need more treatment. The sole EMDR study using the number of sessions recommended in my 1995 text (Carlson et al., 1998) found that after 12 EMDR sessions, 75% no longer had PTSD. This is in contrast to the exposure combat studies, using the same number of sessions, that found only a 30% symptom reduction rate--with no elimination of PTSD diagnosis reported. The need for appropriate clinical parameters to insure external validity is of grave concern.

>

>We agree that statistical power issues are important in psychotherapy

>evaluation. But EMDR is described as a "miracle", not as an ordinary

>run-of-the-mill therapy. Hence, one might expect it to surmount the pesky

>power problems that mask the effects of nonmiraculous psychotherapies.

-----Invective again? I'm not Barbara Walters. I don't use the word "miraculous" to describe EMDR's effects. I consider them predictable, consistent, and fortunate for those who are suffering. And I've always supported the notion that scientifically rigorous studies were necessary. Are you contending that they aren't?

Francine

From: beutler@education.ucsb.edu To: rjm@wjh.harvard.edu, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Rick provides a nice critique of Shapiro. However, it is more of Shapiro than of EMDR. The major thrust is persuasive but scientifically weak. The issue is not whether Shapiro's advertisements are valid (an ethical issue) but whether the method is scientifically sound. Thus, the issue of length of

treatment effects can't rest on the early studies that were clearly flawed, but on direct comparisons of EMDR with other treatments in terms of length. My reading, and it is extensive of this literature at this point, suggests that it may be somewhat shorter than CT and may retain effects better, though long term outcomes are sparse. It is also not true (Rick did not comment on this, but Lohr did) that the literature is devoid of standard outcome data, including physiological measures. The measure are both consistent with extant psychotherapy literature and quite extensive, both behavioral, self-report, clinician report and psychophysiological. Likewise, while the comparison treatments are not standard CT, etc., they do include standard and traditional care as well as no treatment and placebo treatment groups.

Larry

From: Gerald Rosen <groser@u.washington.edu> To: beutler@education.ucsb.edu Cc: rjm@wjh.harvard.edu, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

One day someone should do a sociological analysis of the EMDR phenomenon.

Meanwhile, Beutler suggested that McNally's comments were more on Shapiro than on the technique. I see no reason for Beutler to do this, but it is not the first time that EMDR proponents have countered sound argument with the diversionary tactic that comments were ad hominem. And by the way, this observation involves an analysis of communication tactics, and is not ad hominem either.

McNally makes a valid point when he says that EMDR has been accompanied by extraordinary claims. If Shapiro didn't use the word "Miracle" herself, perhaps she did get to read the book jacket as is routinely done by publishers. Even if she didn't, her own words of praise for EMDR and the claims that have been made on behalf of EMDR, have been extraordinary.

Extraordinary claims do require extraordinary proof. One must demonstrate that eye movements do something before colleagues have to demonstrate that they don't. If EMDR has helped to demonstrate that guided exposure can reduce the anxiety associated with "PTSD symptoms" that's fine, but let's remember that guided imaginal exposure was around, like flooding techniques, before EMDR.

EMDR is presenting a great challenge to clinical psychologists and it is taking a good while to rid ourselves of this hocus pocus. I say this point as strongly as I do because I believe the data are in. Eye movements are very silly. We don't have to wave fingers in front of a patient's face to increase the effects of antibiotics for bronchitis, and we don't have to wave fingers in front of a patient's face to increase the effects of guided imaginal exposure or flooding. The data are very clear about this in relation to exposure therapies, and I doubt to many physicians are going to check out the eye movements with antibiotics.

If Beutler and others wish, they can maintain this silliness. But for some of us, there exists enough data already.

Meanwhile, when McNally made his points in an earlier message, he was sticking to the issues on EMDR. The notion that a piercing critique of EMDR is an attack on Shapiro is even sillier than waving fingers.

-Jerry Rosen

Date: Thu, 23 Jul 1998 16:00:59 -0400 (EDT) From: Carolyn Pepper <cpepper@binghamton.edu> To: beutler@education.ucsb.edu, rjm@wjh.harvard.edu, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure
--

beutler@education.ucsb.edu wrote:

> Thus, the issue of length
> of treatment effects can't rest on the early studies that were clearly
> flawed, but on direct comparisons of EMDR with other treatments in terms of
> length. My reading, and it is extensive of this literature at this point,
> suggests that it may be somewhat shorter than CT and may retain effects
> better, though long term outcomes are sparse.

I'm not sure I follow this. Are you (and Dr. Shapiro) advocating comparing across studies that were designed to have different treatment lengths? Can we really say that one treatment method is shorter than another unless outcomes were assessed at the same time? Have there been studies that directly

compare EMDR to a CBT protocol with an exposure component?

Carolyn M. Pepper, Ph.D. Office: (607)777-2426
Department of Psychology fax: (607)777-4890
Binghamton University
Binghamton, NY 13902-6000

Date: Thu, 23 Jul 1998 14:19:42 -0700 From: beutler@education.ucsb.edu To: Gerald Rosen <groesen@u.washington.edu>, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Jerry's comments reprove me for being ad hominum. If that is the way my comments were percieve, either by Rick or any one else, I am sincerely sorry. I in fact appreciated, and think I said that, Rick's comments, noting only that I thought he had gotten off track on a couple of points, failing to distinguish between the scientific issues and the obvious hype.

There is certainly no rancor here. But I do have some at being called (ad hominem) an advocate of EMDR and with the implication from one who does not know me at all, that I have no integrity. I most profoundly am not. I have never taken a course, have only met Shapiro once, have never seen it practiced except by tape, and have no interest in becoming an EMDR Therapist. I have not written about it, though one of these days, I do plan to publish some data on integrity of EMDR to which I've referred several times on this net.

In any case, in the meantime, Rick I'm sorry if indeed I owe you an apology or have misrepresented you in any way.

Larry

At 11:39 AM 7/23/98 -0700, you wrote:
>One day someone should do a sociological analysis of the EMDR phenomenon.

>Meanwhile, Beutler suggested that McNally's comments were more on Shapiro
 >than on the technique. I see no reason for Beutler to do this, but it is
 >not the first time that EMDR proponents have countered sound argument with
 >the diversionary tactic that comments were ad hominem. And by the way,
 >this observation involves an analysis of communication tactics, and is not
 >ad hominem either.
 >
 >McNally makes a valid point when he says that EMDR has been accompanied by
 >extraordinary claims. If Shapiro didn't use the word "Miracle" herself,
 >perhaps she did get to read the book jacket as is routinely done by
 >publishers. Even if she didn't, her own words of praise for EMDR and the
 >claims that have been made on behalf of EMDR, have been extraordinary.
 >
 >Extraordinary claims do require extraordinary proof. One must demonstrate
 >that eye movements do something before colleagues have to demonstrate that
 >they don't. If EMDR has helped to demonstrate that guided exposure can reduce the anxiety associated with "PTSD symptoms" that's
 >fine, but let's remember that guided imaginal exposure was around, like
 >flooding techniques, before EMDR.
 >
 >EMDR is presenting a great challenge to clinical psychologists and it is
 >taking a good while to rid ourselves of this hocus pocus. I say this
 >point as strongly as I do because I believe the data are in. Eye
 >movements are very silly. We don't have to wave fingers in front of a
 >patient's face to increase the effects of antibiotics for bronchitis, and
 >we don't have to wave fingers in front of a patient's face to increase the

>effects of guided imaginal exposure or flooding. The data are very clear
>about this in relation to exposure therapies, and I doubt to many
>physicians are going to check out the eye movements with antibiotics.
>
>If Beutler and others wish, they can maintain this silliness. But for
>some of us, there exists enough data already.
Meanwhile, when McNally made
>his points in an earlier message, he was sticking to the issues on EMDR.
>The notion that a piercing critique of EMDR is an attack on Shapiro is
>even sillier than waving fingers.
>
> -Jerry Rosen

Date: Thu, 23 Jul 1998 18:11:10 -0400 (EDT) From: "Richard J. McNally" <rjm@wjh.harvard.edu> To: Francine Shapiro <fshapiro@emdr.org> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Francine, Larry, & Jerry:

Many thanks for your comments! (WARNING: long-winded reply!!)

Jerry correctly states that the EMDR phenomenon is a fascinating topic worthy of study by sociologists (and economists and anthropologists as well). Freud once wrote a book entitled "History of the Psychoanalytic Movement," and when someone finally writes a book entitled "History of the EMDR Movement," he or she will have the chance to unravel these knotty, but absorbing, issues.

As Prolegomena to Any Future History of EMDR, consider the following:

1) Francine (Shapiro, 1989) describes an imaginal desensitization technique whereby Wolpe's progressive muscle

relaxation is replaced by therapist-induced eye movements. When Wolpe and I discussed it at the New York City AABT in 1991, that is precisely how he conceptualized it: eye movements simply replaced muscle relaxation as the "reciprocal inhibitor" of anxiety.

Nearly ten years have elapsed and there is **still** no convincing evidence that eye movements add anything to imaginal exposure. Persuaded by Lohr, Tolin, and Lilienfeld's (1998) review, I concluded that "what is effective in EMDR is not new [imaginal exposure], and what is new [eye movements] is not effective." I respect that fact that Larry and Francine disagree.

Francine informs us that only two of the 22 patients in her study were Vietnam combat veterans. This is what the article says "One subject had undergone severe mental and physical abuse, while all others had experienced either rape/molestation or Vietnam War combat incidents" (p. 203). No breakdown of vets versus rape survivors is provided in the article.

But if studies on repeatedly traumatized, often compensated, veterans with combat-related PTSD are inappropriate for treatment studies on EMDR, why do EMDR therapists continue to treat these people in the VA system? Are they not inappropriate candidates for EMDR?

Francine was careful to emphasize that eye movement therapy, as described in her article, was not intended to be a treatment for **all** PTSD symptoms. It was specifically targeted for distress associated with traumatic memories.

2) But by 1994, a metamorphosis had occurred as announced in an article delightfully entitled "EMDR: In the Eye of a Paradigm Shift" (Shapiro, 1994, the Behavior Therapist). By now she had reconceptualized the method in the jargon of information processing (adding an "R" to EMD), and proclaimed in the preface to her 1995 book, "We went from Kitty Hawk to a man on the moon in little more than 50 years, yet we have not had a major paradigm shift in psychology since Freud, nearly a century ago" (p. xii).

Not merely a variant of systematic desensitization, EMDR was now described as a complex eight-stage intervention. But close reading of her book reveals that these "stages" include such routine generic elements as taking the patient's history,

patient preparation, assessment, reevaluation, and so forth. One wag even wisecracked, "Shapiro's method is really a *ten*-stage complex intervention: she forgot to include 'scheduling the appointment' and 'billing the patient'."

3) By the spring of 1998, Shapiro had affirmed that contemporary EMDR now incorporates elements from behavior therapy, psychodynamic therapy, interpersonal therapy, family therapy, and cognitive therapy. How the eye movement therapy manages to incorporate all this is not entirely clear. It seems akin to a goldfish swallowing a whale. If the expanding perimeter of EMDR now incorporates all other psychotherapies, is it now simply "eclectic psychotherapy"?

4) Unbeknownst to most listmembers, EMDR enthusiasts continue to expand into new markets. The irrepressibly cheerful Ricky Greenwald, Psy.D., one of the EMDR community's most innovative practitioners, has recently discussed ways that EMDR can be incorporated into the treatment of Asberger's Syndrome. Other enthusiasts have discussed how EMDR can be used in the treatment of epilepsy, premature ejaculation, and encopresis (i.e., fecal soiling due to impaction).

It is unclear to me how having seizure-prone patients flick their eyes back and forth is helpful, nor do I understand how moving one's fingers before the patient's eyes ought to foster bowel movements in the encopretic, while delaying ejaculation in the sexually dysfunctional.

Enthusiasts have also added EMDR to effective treatments for bulimia (e.g., SSRI, interpersonal psychotherapy, CBT) which reminds me of an old Russian folktale that I paraphrase from one of David Navon's articles in "Psychological Review" entitled: "Resources: A Theoretical Soup Stone?".

"The Soup Stone"

Once upon a time, a scamp taught a fool how to make a delicious soup. The scamp said, "It's easy; all you need to do is to take this here soup stone, put it in some water, and boil it for three minutes." The fool was impressed. The scamp continued, "But, of course, it doesn't hurt to toss in a little meat, and maybe add a few vegetables. A little salt and pepper doesn't hurt either." The fool remained impressed.

Now Navon told this parable in reference to the construct of "cognitive resources." But when EMDR is added to otherwise established treatments, and assumed to have boosted efficacy without evidence, is it not possible that eye movement is yet another soup stone? Inquiring minds want to know.

In any event, if EMDR is shown to be the breakthrough treatment that many sincerely believe it is, I will gladly toast the pioneers who developed and validated it. I will happily reverse my now-skeptical view, because like Ricky Greenwald, I, too, am irrepressibly cheerful!

I join Francine in urging interested listmembers to read Lohr et al. and the articles cited therein.

Ricky McNally

P.S. Francine's dustjacket describes the technique as a "miracle" and my mentioning that fact does not seem to qualify as "invective". Moreover, I find this advertisement neither "unethical" nor "bothersome"; I find it amusing.

Date: Thu, 23 Jul 1998 18:24:30 -0400 (EDT) From: "Richard J. McNally" <rjm@wjh.harvard.edu> To: beutler@education.ucsb.edu Cc: Gerald Rosen <groser@u.washington.edu>, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure
--

Larry:

Gosh, no need to apologize! I did not take your remarks in a negative light. In fact, I appreciate learning about your thoughts on this fascinating topic.

Regarding the fidelity issue, I am delighted that you are studying this important issue. In my opinion, the most

penetrating analysis ever written on this topic (in reference to EMDR) is by none other than Jerry Rosen himself! (Jerry being a shy sort of guy, I will toot his horn for him! ;-)).

Jerry's article is one of 12 (!!), both "pro" and "con", slated to appear in the January 1999 of the Journal of Anxiety Disorders. Francine and I also have articles in this special series. If nothing else, this issue, done by guest editor, Ron Aceirno, documents the salience of EMDR in today's clinical community.

Rich

P.S. to listmembers: My article concerns the sociological dimensions of EMDR, and I am happy to mail reprints to interested individuals who email me their snail mail addresses backchannel.

On Thu, 23 Jul 1998 beutler@education.ucsb.edu wrote:

- > Jerry's comments reprove me for being ad hominum. If that is the way my
- > comments were percieve, either by Rick or any one else, I am sincerely
- > sorry. I in fact appreciated, and think I said that, Rick's comments,
- > noting only that I thought he had gotten off track on a couple of points,
- > failing to distinguish between the scientific issues and the obvious hype.
- >
- > There is certainly no rancor here. But I do have some at being called (ad
- > hominem) an advocate of EMDR and with the implication from one who does not
- > know me at all, that I have no integrity. I most profoundly am not. I
- > have never taken a course, have only met Shapiro once, have never seen it
- > practiced except by tape, and have no interest in becoming an EMDR
- > Therapist. I have not written about it, though one of these days, I do
- > plan to publish some data on integrity of EMDR to which I've referred

> several times on this net.
>
> In any case, in the meantime, Rick I'm sorry if indeed I owe
you an apology or have misrepresented you in any way.
>
> Larry

Date: Thu, 23 Jul 1998 15:39:03 -0700 (PDT) From: Gerald Rosen <groser@u.washington.edu> To: beutler@education.ucsb.edu Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Beutler stated in his original comment regarding McNally:

"Rick provides a nice critique of Shapiro. However, it is more
of Shapiro than of EMDR".

It seems to me that Beutler was suggesting McNally had taken
an ad hominum/adhominem tact toward Shapiro. In my
suggesting otherwise, I was not reproving Beutler for being ad
hominum/adhominem. That is all ridiculous. And I see
nothing in my message that implies integrity problems for
Beutler.

As regards the scientific issues, I look forward to seeing the
data Beutler refers to that suggests EMDR has longer lasting
effects and less relapse than other exposure therapies. This is a
new claim that can now be added to the many made on behalf
of EMDR, and I imagine psychologists will be spending
another 10 years of our profession's life chasing this one down.

-Jerry Rosen

Date: Fri, 24 Jul 1998 07:13:09 -0500 (CDT) From: "Jeffrey M. Lohr" <jlohr@comp.uark.edu> To: beutler@education.ucsb.edu Cc: rjm@wjh.harvard.edu, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

The measures that are consistently used are verbal report indices, either questionnaires completed by research participants or interviews conducted infrequently by clinicians who are unaware of assignment to treatment condition. Very rarely are psychophysiological measures used, and even more rarely are overt behavioral measures used. The application of measurement and intervention procedures do not conform well to Foa and Meadows (1997) gold standard criteria. Multi-modal and multi-content assessment has not been well established in EMDR research. These issues are clearly portrayed in Tables 1 and 2 of Lohr, Tolin, and Lilienfeld (1998).

Moreover, direct comparisons of emdr and in vivo or imagery exposure treatments by the Muris and Merckelbach group show no differential benefit of emdr. Indeed, emdr does not do as well as the exposure treatments. There is no direct evidence that emdr "lasts longer" or "is easier to take" (see Cahill and Freuh, 1997). I suggest that we wait until the first issue of the 1999 Journal of Anxiety Disorders is published before we rush to judgment.

Jeffrey M. Lohr, Ph.D.
Department of Psychology
216 Memorial Hall
University of Arkansas
Fayetteville, Arkansas 72701
Phone (501) 575-5813
Fax (501) 575-3219

"Forma omni, atqui re sine"

Date: Fri, 24 Jul 1998 07:19:01 -0500 (CDT) From: "Jeffrey M. Lohr" <jlohr@comp.uark.edu> To: Carolyn Pepper <cpepper@binghamton.edu> Cc: beutler@education.ucsb.edu, rjm@wjh.harvard.edu, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

There has been one study that yoked the "amount" of treatment (Dunn et al. 1996) between emdr and a control procedure, and the results for emdr were not very impressive, and certainly not extraordinary.

Jeffrey M. Lohr, Ph.D.
Department of Psychology
216 Memorial Hall
University of Arkansas
Fayetteville, Arkansas 72701
Phone (501) 575-5813
Fax (501) 575-3219

"Forma omni, atqui re sine"

Date: Fri, 24 Jul 1998 09:42:50 -0700 (PDT) From: fshapiro@emdr.org (Francine Shapiro) To: "Richard J. McNally" <rjm@wjh.harvard.edu> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Rich:

Since you haven't answered my direct questions I'll reiterate points to refer back to from my previous post before attending to your latest message:

- 1) . . . the distinctive combination of elements [that comprise EMDR] that achieve 84-100% PTSD remission in only 3-5 hours is in itself "new"--particularly compared to the only civilian PTSD imaginal exposure study (Foa et al., 1991) which used @25 hours of exposure to achieve a 55% success rate.
- 2) How is EMDR defined as an exposure therapy if the rapidity of effects cannot be predicted by the extinction/habituation model and many of the procedures are in direct contradiction to it?
- 3) No controlled civilian PTSD EMDR studies failed to show positive effects more rapidly (with fewer exposure hours) than traditional exposure techniques
- 4) A 3-5 hour EMDR effect compared to a 50-100 hour prolonged exposure effect can reasonably be called a "breakthrough"
- 5) The situation with my initial study of EMD has now been clarified. Hopefully, you will make sure that the

misinformation doesn't show up in future articles (e.g., Rosen, in press; Rosen, McNally, Lohr, Devilly, Herbert & Lilienfeld, in press).

6) The three conditions I enumerated (analogues/combat veterans/small n) as problematic in component analyses represent all of the EMDR component analyses. Therefore, any statement that the eye movement, or other stimulation, has been proven "superfluous" is incorrect. Hopefully, you will make sure that the misinformation doesn't show up in future articles (e.g., Rosen, McNally, Lohr, Devilly, Herbert & Lilienfeld, in press).

Now for your new observations:

...

>

>1) Francine (Shapiro, 1989) describes an imaginal desensitization

>technique whereby Wolpe's progressive muscle relaxation is replaced by

>therapist-induced eye movements. When Wolpe and I discussed it at the New

>York City AABT in 1991, that is precisely how he conceptualized it: eye

>movements simply replaced muscle relaxation as the "reciprocal inhibitor"

>of anxiety.

-----Yes, Wolpe did define it as a reciprocal inhibition/desensitization technique. He considered it to be achieving results far beyond systematic desensitization, bless his heart. I was impressed by his non-territoriality in his acceptance of EMDR. He provided a worthy model in many ways.

>

>Nearly ten years have elapsed and there is *still* no convincing evidence

>that eye movements add anything to imaginal exposure.

-----Rather, there is no real evidence to show it does not. I remind you of my uncontested statement that EMDR has produced effects (80-100% remission of PTSD) in 3-5 hours in four controlled studies only matched by 50-100 hours of imaginal and in vivo exposure. What do you attribute that to? As noted, even the simplistic "EMD" was more effective than

imaginal exposure alone -- Wolpe merely added eye movements to previously unresponsive systematic desensitization clients and observed the positive effects as have many other behavioral clinicians. Now the EMDR procedures have become so robust that rigorous component analyses are vital. However, my primary suggestion, as previously stated, is that if you don't like the stimulation component (e.g., find it "silly"), then use the rest of the method without it. However, recognize that both clinicians and clients generally prefer the eye movement or other stimulation (e.g., Boudewyns & Hyer, 1996) since at the very least it appears to titrate the distress during treatment (see Andrade, J., Kavanagh, D., & Baddeley, A. (1997). Eye-movements and visual imagery: A working memory approach to the treatment of post-traumatic stress disorder. *British Journal of Clinical Psychology*, 36, 209-223.)

>Persuaded by Lohr,
>Tolin, and Lilienfeld's (1998) review, I concluded that "what is effective
>in EMDR is not new [imaginal exposure], and what is new [eye movements] is
>not effective." I respect that fact that Larry and Francine disagree.

-----Rather than review the studies that support the eye movement, I'll just refer you to #1-4 above regarding imaginal exposure and say that it is unfortunate that you were persuaded by the Lohr et al (1998) review to make your declaration. Particularly when I've published a detailed commentary (Shapiro, 1996, JBTEP--which I'm happy to send along to anyone on request) decisively documenting four full pages of data citation errors in their previous review (Lohr et al., 1995). (It is noteworthy that my findings were never rebutted, but rather were followed with a gale of attacks on my "credentials".) This underscores an admonition during controversy: Read the original research rather than literature reviews.

>

>Francine informs us that only two of the 22 patients in her study were
>Vietnam combat veterans. This is what the article says "One subject had
>undergone severe mental and physical abuse, while all others had
>experienced either rape/molestation or Vietnam War combat incidents" (p.

>203). No breakdown of vets versus rape survivors is provided in the
>article.

-----Glad to have clarified the situation for you. If you note the grid of studies in chapter 12 of my 1995 text I listed the subjects as primarily single-trauma.

>

>But if studies on repeatedly traumatized, often compensated, veterans with

>combat-related PTSD are inappropriate for treatment studies on EMDR, why

>do EMDR therapists continue to treat these people in the VA system? Are

>they not inappropriate candidates for EMDR?

-----You have misstated my point. Certainly combat veterans are worthy candidates, but only if a sufficient number of sessions are given. Once again, the sole EMDR study using the number of sessions recommended in my 1995 text (Carlson et al., 1998) found that after 12 EMDR sessions, 75% no longer had PTSD. This is in contrast to the exposure combat studies, using the same number of sessions, that found only a 30% symptom reduction rate--with no elimination of PTSD diagnosis reported. The 75% EMDR success-rate with vets in the Carlson study, during a time period comparable to usual in-patient treatment, mirrors the success-rate previously reported by practicing V.A. clinicians using EMDR. That is why a panel of directors of V.A. PTSD programs advocated the use of EMDR a number of years ago at the International Society for Traumatic Stress Studies conference. It is unfortunate that studies that attempt only two sessions of EMDR treatment with this population are given any credence (e.g., Devilly et al., in press; Jensen, 1994). These types of studies simply muddy the waters and prevent vets from being helped.

>Francine was careful to emphasize that eye movement therapy, as described

>in her article, was not intended to be a treatment for *all* PTSD

>symptoms. It was specifically targeted for distress associated with

>traumatic memories.

-----Yes, and I thought that would have helped to dispel the misconceptions you relayed regarding the initial study's success-rate. But since 1989 the treatment has been refined and made more robust, as described in my 1995 text.

>

>

>2) But by 1994, a metamorphosis had occurred as announced in an article

>delightfully entitled "EMDR: In the Eye of a Paradigm Shift" (Shapiro,

>1994, the Behavior Therapist).

-----Thank you. I also enjoyed the title. I was describing EMDR's shift in paradigm from a desensitization conception to the information processing model. At the time I considered the article informative, not controversial. Live and learn.

By now she had reconceptualized the method

>in the jargon of information processing (adding an "R" to EMD), and

>proclaimed in the preface to her 1995 book, "We went from Kitty Hawk to a

>man on the moon in little more than 50 years, yet we have not had a major

>paradigm shift in psychology since Freud, nearly a century ago" (p. xii).

-----The term paradigm shift in the article did not have a relation to its use in the preface--it was, as you note, another publication and at least a year later. As to the use of the term in the preface, do you disagree that there has not been a paradigm shift since Freud? Wolpe often decried the fact that psychodynamic therapy still reigned despite all his attempts. I believe a series in the journal of Behavior Therapy also indicated "promises unkept." Unfortunately, while I was recently in Canada, a radio program discussed the efficacy of cognitive therapy for depression, but decried the fact that so few clinicians practiced it. As Kuhn used the term it indicated to pervasive shift. Do you disagree with Wolpe's point of view?

>

>Not merely a variant of systematic desensitization, EMDR was now described

>as a complex eight-stage intervention. But close reading of her book

>reveals that these "stages" include such routine generic elements as

>taking the patient's history, patient preparation, assessment,

>reevaluation, and so forth. One was even wisecracked,
"Shapiro's method
>is really a *ten*-stage complex intervention: she forgot to
include
>'scheduling the appointment' and 'billing the patient'."

-----Obviously not a close enough reading of the book. As
previously noted, the targets looked for in the history-taking,
the form of preparation, the delineation of components in the
assessment, the procedures used during processing, etc. are all
"distinctive" ingredients--in a unique combination. "Wags"
aside, the "generic" name of a phase should not be confused
with the clinical procedures employed. As an example,
perhaps you would be willing to define the EMDR assessment
phase and note any other method that delineates and arranges
these elements in the precise order for subsequent processing?

>
>3) By the spring of 1998, Shapiro had affirmed that
contemporary
>EMDR now incorporates elements from behavior therapy,
psychodynamic
>therapy, interpersonal therapy, family therapy, and cognitive
therapy.
>How the eye movement therapy manages to incorporate all
this is not
>entirely clear. It seems akin to a goldfish swallowing a whale.
>If the expanding perimeter of EMDR now incorporates all
other
>psychotherapies, is it now simply "eclectic psychotherapy"?

-----Actually, if you reread my 1995 textbook, I made the
statement then, with concrete examples. As I've repeatedly
stated, EMDR was refined over the past ten years into a
complex method incorporating many elements in order to
achieve robust effects across the clinical spectrum. Remember,
"EMD" was initially applied primarily to single-trauma
victims. Unfortunately, EMDR's critics seem to insist on
continuing to define it as a simplistic technique. Nor is it
"eclectic psychotherapy"-- I suggest you read Fensterheim's
article in the Journal of Psychotherapy Integration (1996):
"Unlike many other integrative methods that have been
proposed, the different aspects of this method [EMDR] are not
applied separately but form a cohesive whole" (p.27). Or
carefully read again my 1995 text or a new EMDR book
(Manfield, 1998) for further clarification. >
>4) Unbeknownst to most listmembers, EMDR enthusiasts
continue to expand

>into new markets. The irrepressibly cheerful Ricky Greenwald, Psy.D., . . .

-----Sorry, had to start ellipsing. Shall we invite Greenwald to address your concerns? In my writing I have been very clear. EMDR is to be used to address the experiential contributors of complaints--not organic issues. Again, a careful reading of my text makes the parameters clear.

But when EMDR is added to otherwise established treatments, >and assumed to have boosted efficacy without evidence, is it not possible
>that eye movement is yet another soup stone? Inquiring minds want to
>know.

-----Perhaps inquiring minds should take another look at the "empirically supported treatments" lists. Approximately twelve methods were determined to be "well-established" for specific disorders. While I haven't used EMDR for anything you mentioned, exactly which of the disorders have well-established treatments that appear on the list? And the further mention of "eye movement" as a pejorative is, as previously noted, unwarranted given the present state of component analyses. All that can be appropriately stated based upon the controlled research is that EMDR produces effects in 3-5 hours of treatment comparable to 50-100 hours of imaginal and in vivo exposure. Further, EMDR entails a distinct cognitive component not found in exposure therapies (Pitman et al, 1996; Shapiro, 1995). Therefore, wherever cognitive and exposure therapies have proved effective, EMDR may well be integrated into the treatment. Only appropriately controlled research will provide the definitive answers.

>In any event, if EMDR is shown to be the breakthrough treatment that many
>sincerely believe it is, I will gladly toast the pioneers who developed
>and validated it. I will happily reverse my now-skeptical view, because
>like Ricky Greenwald, I, too, am irrepressibly cheerful!

-----And I in turn would toast you, the practicing clinicians who were willing to stand up to ridicule, and the dispassionate researchers that rigorously evaluate the method. Nothing

would be more refreshing than collapsing an "I and thou" mentality into a joint effort.

>

>I join Francine in urging interested listmembers to read Lohr et al. and

>the articles cited therein.

-----Actually, the Feske review (1998) is a useful companion to the Lohr et al (1998) article. But please do read the articles cited therein. Further, three of the most recent controlled EMDR/PTSD outcome studies don't appear in the Lohr et al review so here is the complete list of recent controlled PTSD outcome research: EMDR: Carlson et al., 1998; Marcus et al., 1998; Rothbaum, 1997; Scheck et al., 1998; and Wilson et al., 1995, 1997. Exposure: Marks et al., 1998 As previously noted, during times of controversy, I'd suggest relying on your own judgment.

>Ricky McNally

>

>P.S. Francine's dustjacket describes the technique as a "miracle" and my

>mentioning that fact does not seem to qualify as "invective".

>Moreover, I find this advertisement neither "unethical" nor "bothersome";

>I find it amusing.

-----Some time in the future, it would be useful to concentrate solely on the scientific issues. I have to admit my "amusement" threshold is a great deal lower than yours. Due to the constant drumbeat of misinformation, clinicians are being told that combat veterans and other chronic PTSD suffers cannot be treated (Shalev et al. 1996). Unfortunately, they chose to review all known PTSD treatments except EMDR. As noted in this post, according to those employing EMDR in the VA system--and the one EMDR study to use a sufficient (comparable) number of treatment sessions-- it simply isn't so. Unfortunately, many don't know that and the substance abuse, violence, and suicides reported by many VA clinicians continue. To paraphrase a client of mine, "People are dying, and that you can't reverse."

//////////

Francine Shapiro, Ph.D.

FShapiro@emdr.org

Senior Research Fellow
Mental Research Institute
555 Middlefield Road
Palo Alto, CA 94301
fax: 650/728-0654

Executive Director
EMDR Institute
P.O. Box 51010
Pacific Grove, CA 93950

tel: 408/372-3900
fax: 408/647-9881

Date: Fri, 24 Jul 1998 14:27:28 -0400 (EDT) From: "Richard J. McNally" <rjm@wjh.harvard.edu> To: Francine Shapiro <fshapiro@emdr.org> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure

Francine: (Warning: Yet another long-winded reply!!)

Many thanks again for your informative comments! I appreciate your efforts!

1) I think we will have to "agree to disagree" on some points. Whatever components, generic or otherwise, that constitute contemporary EMDR, eye movements surely must count as the defining core. I believe that researchers need to show that eye movements enhance the efficacy of imaginal exposure, whereas you believe (I think) that researchers need to show that eye movements do **not** enhance the efficacy of imaginal exposure. You have also stated that other forms of bilateral stimulation can replace eye movements, and you noted that "forced fixation" (of the eyes, presumably) also works [Shapiro, 1995, p. 25]. I do not understand how moving eyes and fixed eyes produce the same therapeutic effect.

Also, I did not call the stimulation component "silly." I said that there was no convincing evidence that it helped.

2) In fairness to Jensen (1994) and Davilly et al. (in press), their two-session treatment studies had twice as many sessions as your landmark study. It is unfortunate that you did not specify how many patients in your original study were combat veterans; this would have helped prevent what you term the "misinformation" reported by others when they discuss your report. And what "misconceptions" have I relayed about this study? You cited a "100% success rate" in reducing distress associated with traumatic memories, and that's what I said.

3) In fairness to Lohr, Tolin, and Lilienfeld (1998), your critique of Lohr et al.'s early review is irrelevant to the one that just appeared in Behavior Therapy. Indeed, the point of their review article was to survey developments that had occurred since their previous article had appeared. I respect the fact that your "take" on the studies covered by Lohr, Tolin, and Lilienfeld (1998) may differ from theirs, and I urge you to submit a critique to Behavior Therapy. After all, criticism is the oxygen of science.

4) Yes, I agree that interested listmembers should read the Feske review, too. And yes Dr. Fensterheim kindly sent me a reprint of his article, but it only seems to underscore my sense that the expanding perimeter of EMDR seems to render it another eclectic psychotherapy. But I hasten to add that you or others may see it differently. (Another point where we may need to "agree to disagree").

5) Is EMDR the first new paradigm in the field of psychology since the advent of psychoanalysis, as you suggest?

Well, part of the problem turns on how one defines "paradigm" (you'll recall Margaret Masterman's famous essay on Kuhn's book where she lists 20+ different ways in which Kuhn used the word!!). Some people hold that Behaviorism - either the Hull/Spence version or the Radical Behaviorism of Skinner - count as post-Freudian paradigms in psychology, and others refer to the "cognitive revolution" of George Miller, Noam Chomsky, and others as a new paradigm that preceded the advent of EMDR.

In the clinical field, one often hears of post-Freudian paradigms like behavior therapy, cognitive therapy, and neo-Kraepelinian biological psychiatry. Pioneers advancing these new approaches appeared on the scene in the period between Freud and yourself (and one of them, Don Klein, is an active member of SSCP).

In any event, your title about being in the eye of a paradigm shift was one of the coolest I've seen! Although Dan Wegner's recent article entitled "The Putt and Pendulum" (cognitive overcontrol in golf) comes pretty darn close!

6) One of the articles slated to appear in the Journal of Anxiety Disorders special issue is a randomized trial that compares Foa-style CBT for PTSD with EMDR. The patients are civilians, plenty of sessions were used, etc., and in fact the authors were especially attentive to the methodological issues that you say trouble previous studies (*including*, I might add, the several new "positive" studies you cite, that are quite flawed, in my opinion).

The results: Both EMDR and CBT produced significant improvements in these civilian PTSD cases, with CBT doing better than EMDR. Both treatments were tolerated equally well. At follow-up, EMDR cases had started to relapse, whereas CBT cases continued to improve.

7) What Wolpe told you about eye movement therapy differed from what he told me (the conversations, however, may have occurred quite separate in time). Interestingly, his last article was on EMDR, and we will see whether you continue to "bless his heart" when it appears.

8) There are many competing theories of why exposure therapy works, and you oversimplify matters quite a bit. As Jerry Rosen noted, brief exposures can be therapeutic as well as long ones, and nothing in "the" extinction/habituation model indicates that EMDR is anything but another exposure variant.

Rather than bore listmembers any more than I already have by recounting the ins-and-outs of various theories of fear reduction, I *strongly* recommend that you (and everyone else, for that matter) consult an outstanding new book (I've read the galley proofs, it appears in next month) by UCLA's Michelle Craske. Michelle covers the entire anxiety disorders field, and goes into great detail in her critiques of extant theories of fear reduction, exposure therapy, and so forth. If a person were to read one book on anxiety disorders, this is the one to read.

Here's the reference:

Craske, M. G. (1998). Anxiety disorders: Psychological approaches to theory and treatment. Boulder, CO: Westview Press.

9) Voltaire once exclaimed, "Woe is the philosopher who cannot laugh away his wrinkles! I regard solemnity as a disease."

Suitable APA-style copyediting of Voltaire yields:

"Woe is the psychologist who cannot laugh away his or her wrinkles. I regard solemnity as a syndrome."

The moral: Don't lose your sense of humor, Francine. It is one of your most endearing characteristics.

Rich

Date: Fri, 24 Jul 1998 12:30:01 -0700 (PDT) From: Gerald Rosen <groser@u.washington.edu> To: "Richard J. McNally" <rjm@wjh.harvard.edu> Cc: Francine Shapiro <fshapiro@emdr.org>, sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure
--

The dialogue between Shapiro and McNally has been helpful. Perhaps one point everyone can agree on is that EMDR has generated a good deal of research that demonstrates the effectiveness of guided imagery to produce exposure. As indicated once before to this list, Barry Wolfe developed a guided imagery approach quite identical to that used in EMDR and he published on this the same year as Shapiro. The article, "Phobias, Panic and Psychotherapy Integration" was published in the Journal of Integrative and Eclectic Psychotherapy, 1989, vol 8, pp. 264-??. Consider the following excerpt from Wolfe's article:

"The cornerstone of this integrative treatment model is actually a product of a serendipitous discovery; that the application of imaginal exposure regularly elicits the tacit, catastrophic conflicts that are directly related to the phobic or panic symptoms under treatment. The imaginal procedure begins with a breathing/induction exercise which helps the patient focus his or her attention inwardly while experiencing an

increasing level of relaxation. I then invite the patient to be receptive to his or her internal productions, that is, to any thoughts, feelings, images or ideas that arise automatically. The patient is asked to imagine the feared object or situation, and while intensively focusing his or her attention on the object to notice any automatically arising feeling or thought. Typically, within the first few sessions, this procedure results in the appearance of several thematically related and emotionally laden images. The themes are usually themes of conflict which feel catastrophic to the patient."

Wolfe continued in this article and spoke of rapid therapeutic effects obtained with the imagery method. Also pointed out to this list previously, there are scattered papers on the use of imagery to produce rapid cures of PTSD related symptoms. For example, there is the paper by Kellner et al, *Am J. Psychiatry*, 1992, 149, 659-662 entitled "Changes in Chronic Nightmares After One Session of Desensitization or Rehearsal Instructions."

Neither Wolfe nor Kellner claimed paradigm shifts and neither took their methods on the road to train over 25,000 clinicians in fairly costly workshops. Neither claimed anything beyond the known effects of guided imagery and exposure. And neither helped to spur a sizeable number of controlled outcome studies with PTSD diagnosed patients.

It is wonderful if people with PTSD can be helped with guided imagery and exposure methods. No one is going to disagree with that. And if EMDR has helped to document those effects that's a very positive development. What isn't very positive concerns all the other unsupported claims for the uniqueness and breakthrough power of EMDR, especially when study after study fails to support the benefits of eye movements, and as McNally pointed out, even fixed eye conditions are claimed to be EMDR.

Of course, if this all keeps up I may need to try EMDR on myself. That's because I keep getting anxious over the image of mental health professionals in the year 2025 laughing their heads off when they read about psychologists waving their fingers in front of their patients' faces to accelerate information processing and increase interhemispheric coherence. On the other hand, I imagine Shapiro images those same professionals laughing at my papers!!!

-Jerry Rosen

Date: Fri, 24 Jul 1998 15:48:46 -0400
From: "John Winston Bush" <jwb@alumni.stanford.org>
To: <groesen@u.washington.edu>,
<beutler@education.ucsb.edu>
Cc: <sscpnet@listserv.acns.nwu.edu>
Subject: Re: Exner System/EMDR and exposure

OK, folks. The phrase is "ad hominem." Two words; penultimate letter an "e." Please everybody save this to an indestructible file.

Date: Fri, 24 Jul 1998 17:15:26 -0700 (PDT)
From: fshapiro@emdr.org (Francine Shapiro)
To: "Richard J. McNally" <rjm@wjh.harvard.edu>
Cc: sscpnet@listserv.acns.nwu.edu
Subject: Re: Exner System/EMDR and exposure

Rich:

Glad we are winding this down on such a positive note-- "agreeing to disagree" is a fine conclusion. Perhaps some of the researchers/students on the list have gotten some thoughts for future study. I also appreciate your efforts. Simply for clarity I'll comment below.

>

>1) I think we will have to "agree to disagree" on some points.

... I

>do not understand how moving eyes and fixed eyes produce the same

>therapeutic effect.

-----Explanations have been offered that "dual attention" is the primary function of the stimulation. Having the client concentrate on a non-emotional task is a common denominator to both. Further, perhaps part of our disagreement is based upon how we view the function of component analyses. I think that since EMDR's efficacy with PTSD has been established-- and repeated studies show that 3-5 hours of treatment are sufficient---the function of component analyses should be to help refine the protocols. Component analyses have to be sensitive to all clinical dimensions of care to be useful--who needs what form of treatment, parameters of treatment and interaction, client and clinician distress, etc. You seem to feel

that the component analyses are needed to confirm the efficacy of the method itself. I don't. I personally don't care if the eye movements prove to be superfluous. If they do, we'd stop using them. The overall procedures are effective with clients and that's what I care about most. I think if I had more wisely called the therapy "Reprocessing Therapy" or "Exposure-Mediated Dual Reprocessing" (as I'd mentioned in the 1995 text) we simply wouldn't be having this conversation.

>

>2) In fairness to Jensen (1994) and Davilly et al. (in press), their

>two-session treatment studies had twice as many sessions as your landmark

>study. It is unfortunate that you did not specify how many patients

>in your original study were combat veterans; this would have helped

>prevent what you term the "misinformation" reported by others when they

>discuss your report. And what "misconceptions" have I relayed about this

>study? You cited a "100% success rate" in reducing distress associated

>with traumatic memories, and that's what I said.

-----As previously noted, statements saying that the success rate has never been duplicated are in error. Many studies have reported an immediate drop in distress during the first session with many more subjects with higher disturbance than mine. And the results have been far surpassed by the elimination of full symptomatology. As for Jensen (1994) and Davilly et al.(in press), my point is that even twice as many sessions (or any other small number) would not possibly allow a change as assessed by the standard psychometrics used. This is simply a clinical validity factor that needs to be taken into account--if not by the researchers, then the research reviewers. Now that we are all clear, however, perhaps future reviews can reflect these shortcomings and more realistic studies emphasized.

>

>3) In fairness to Lohr, Tolin, and Lilienfeld (1998), your critique of

>Lohr et al.'s early review is irrelevant to the one that just appeared in

>Behavior Therapy. Indeed, the point of their review article was to survey
>developments that had occurred since their previous article had appeared.
>I respect the fact that your "take" on the studies covered by Lohr, Tolin,
>and Lilienfeld (1998) may differ from theirs, and I urge you to submit a
>critique to Behavior Therapy. After all, criticism is the oxygen of
>science.

-----Since I've already put on record my evaluation of the caliber of their work, I'll leave the task to others. But I'm sorry to say that I've been informed that Behavior Therapy doesn't accept these types of critiques. >

>
>5) Is EMDR the first new paradigm in the field of psychology since the
>advent of psychoanalysis, as you suggest?

-----Sorry Rich, you misunderstood me. EMDR is absolutely not "the first new paradigm." All of the ones you mentioned are obviously true paradigms, as are many, many more. That's why I've never understood the problem with offering EMDR as governed by an information processing paradigm instead of its earlier desensitization formulation. There are tenets that predict effects and guide its use in clinical practice. I had no intention of stating it was "a paradigm shift" in the Kuhnian sense that seems to have generated so much heat--*obviously* since it had just come on the scene. My previous point was, according to Wolpe and others, the field had not *pervasively* shifted from psychodynamic thought to behaviorism, or cognitive therapy, etc.

>

>

>6) One of the articles slated to appear in the Journal of Anxiety
>Disorders special issue is a randomized trial that compares Foa-style CBT
>for PTSD with EMDR. The patients are civilians, plenty of sessions were
>used, etc., and in fact the authors were especially attentive to the
>methodological issues that you say trouble previous studies (*including*,

>I might add, the several new "positive" studies you cite, that are quite

>flawed, in my opinion).

>

>The results: Both EMDR and CBT produced significant improvements in these

>civilian PTSD cases, with CBT doing better than EMDR.

Both treatments

>were tolerated equally well. At follow-up, EMDR cases had started to

>relapse, whereas CBT cases continued to improve.

-----Thank you for bringing it to my attention. Before I make any judgment about it I'd want to check out a variety of things including the number of homework hours, how fidelity was established, etc. The latter is especially significant because no other EMDR civilian PTSD study has reported any relapses. I think we are in for a very interesting new decade of studies since three others I know of (that are being prepared for publication) compare EMDR to CBT and found just the opposite results to the one you cite.

>

>7) What Wolpe told you about eye movement therapy differed from what he

>told me (the conversations, however, may have occurred quite separate in

>time). Interestingly, his last article was on EMDR, and we will see

>whether you continue to "bless his heart" when it appears.

-----Yes, I know. Wolpe always resisted my adding the "R" and abandonment of the desensitization model and inclusion of the cognitions. We are all entitled to our biases.

>

>8) There are many competing theories of why exposure therapy works, and

>you oversimplify matters quite a bit. As Jerry Rosen noted, brief

>exposures can be therapeutic as well as long ones, and nothing in "the"

>extinction/habituation model indicates that EMDR is anything but another

>exposure variant.

-----Sorry, but what Jerry Rosen describes in Wolfe's article is nothing like EMDR, e.g., clients are not asked to "intensely concentrate", "themes" are not elicited in multiple sessions--and EMDR processing/desensitization effects occur within the first minutes of treatment. However, rather than belaboring the point, I'll recommend a reading of my review of what I call the "dosed exposure" element of EMDR (Shapiro, 1995, pp. 311-312). Obviously exposure is an element of EMDR treatment. It's simply not the only element--nor is it conducted in any way recommended by prolonged exposure advocates in the treatment of PTSD. Just as an aside, its interesting that arriving at one agreed upon theory of how exposure works is not made a prerequisite for acceptance of the utility of exposure methods. I'd suggest the same be acknowledged for EMDR.

Thanks for the Craske reference. And I've always enjoyed Voltaire. Wasn't Candide the premier volume for overcoming the potential for PTSD?

Have a nice weekend,

Francine

//////////

Francine Shapiro, Ph.D.
Senior Research Fellow
Mental Research Institute
555 Middlefield Road
Palo Alto, CA 94301
fax: 650/728-0654

FShapiro@emdr.org

Executive Director
EMDR Institute
P.O. Box 51010
Pacific Grove, CA 93950

tel: 408/372-3900
fax: 408/647-9881

Date: Fri, 24 Jul 1998 21:06:24 -0400 (EDT) From: "Richard J. McNally" <rjm@wjh.harvard.edu>

To: Francine Shapiro <fshapiro@emdr.org> Cc: sscpnet@listserv.acns.nwu.edu Subject: Re: Exner System/EMDR and exposure
--

Francine:

Thanks, once again, for your thoughts! Given the relative brevity of your note, I will interpolate in CAPS.

On Fri, 24 Jul 1998, Francine Shapiro wrote:

> Rich:

>

> Glad we are winding this down on such a positive note--
"agreeing to

> disagree" is a fine conclusion. Perhaps some of the
researchers/students

> on the list have gotten some thoughts for future study. I also
appreciate

> your efforts. Simply for clarity I'll comment below.

YES, AGREEING TO DISAGREE *IS* PROGRESS
BECAUSE IT SHARPENS THE ISSUES FOR FURTHER
STUDY.

>>

>>1) I think we will have to "agree to disagree" on some
points. . . . I

>>do not understand how moving eyes and fixed eyes produce
the same

>>therapeutic effect.

>

> -----Explanations have been offered that "dual attention" is
the primary

> function of the stimulation. Having the client concentrate on
a

> non-emotional task is a common denominator to both.

Further, perhaps part

> of our disagreement is based upon how we view the function
of component

> analyses. I think that since EMDR's efficacy with PTSD has
been

> established--and repeated studies show that 3-5 hours of
treatment are

> sufficient---the function of component analyses should be to
help refine

> the protocols. Component analyses have to be sensitive to all clinical
 > dimensions of care to be useful--who needs what form of treatment,
 > parameters of treatment and interaction, client and clinician distress,
 > etc. You seem to feel that the component analyses are needed to confirm
 > the efficacy of the method itself. I don't. I personally don't care if
 > the eye movements prove to be superfluous. If they do, we'd stop using
 > them. The overall procedures are effective with clients and that's what
 > I care about most. I think if I had more wisely called the therapy
 > "Reprocessing Therapy" or "Exposure-Mediated Dual Reprocessing" (as I'd
 > mentioned in the 1995 text) we simply wouldn't be having this conversation.
 >

YES, I THINK WE DO SEE THE COMPONENT ANALYSES DIFFERENTLY. REQUESTS THAT RESEARCHERS SHOW THAT EYE MOVEMENTS DO NOT ADD ANYTHING TO IMAGINAL EXPOSURE STRIKE ME AS A REQUEST FOR RESEARCHERS TO PROVE THE NULL HYPOTHESIS. WELL, IN ANY EVENT, THE "DUAL STIMULATION HYPOTHESIS" *IS* TESTABLE.

YOUR COMMENTS ABOUT THE BEST TITLE FOR WHAT IS NOW CALLED EMDR REMIND ME OF WOLPE'S COMMENTS IN 1982 (WHEN I WAS A INTERN IN HIS BEHAVIOR THERAPY UNIT IN PHILADELPHIA). UPON REFLECTION, HE WISHED HE HAD TERMED HIS APPROACH "DECONDITIONING THERAPY," NOT "BEHAVIOR THERAPY."

>
 >
 > >
 > >2) In fairness to Jensen (1994) and Davilly et al. (in press), their
 > >two-session treatment studies had twice as many sessions as your landmark

> >study. It is unfortunate that you did not specify how many patients
 > >in your original study were combat veterans; this would have helped
 > >prevent what you term the "misinformation" reported by others when they
 > >discuss your report. And what "misconceptions" have I relayed about this
 > >study? You cited a "100% success rate" in reducing distress associated
 > >with traumatic memories, and that's what I said.
 >
 > -----As previously noted, statements saying that the success rate has
 > never been duplicated are in error. Many studies have reported an
 > immediate drop in distress during the first session with many more subjects
 > with higher disturbance than mine. And the results have been far surpassed
 > by the elimination of full symptomatology. As for Jensen (1994) and
 > Devilly et al.(in press), my point is that even twice as many sessions (or
 > any other small number) would not possibly allow a change as assessed by
 > the standard psychometrics used. This is simply a clinical validity factor
 > that needs to be taken into account--if not by the researchers, then the
 > research reviewers. Now that we are all clear, however, perhaps future
 > reviews can reflect these shortcomings and more realistic studies
 > emphasized.
 > >
 > >3) In fairness to Lohr, Tolin, and Lilienfeld (1998), your critique of
 > >Lohr et al.'s early review is irrelevant to the one that just appeared in
 > >Behavior Therapy. Indeed, the point of their review article was to survey
 > >developments that had occurred since their previous article had appeared.
 > >I respect the fact that your "take" on the studies covered by Lohr, Tolin,

> >and Lilienfeld (1998) may differ from theirs, and I urge you to submit a

> >critique to Behavior Therapy. After all, criticism is the oxygen of
> >science.

>

> -----Since I've already put on record my evaluation of the caliber of their

> work, I'll leave the task to others. But I'm sorry to say that I've been

> informed that Behavior Therapy doesn't accept these types of critiques.

> >

WELL, AGAIN, WE WILL HAVE TO "AGREE TO DISAGREE" ON THE CALIBER OF THE LOHR, TOLIN, & LILIENFELD (1998) ARTICLE. ACTUALLY, THERE WAS AN EXCHANGE ON FLOODING THERAPY FOR TRAUMA PATIENTS BACK IN THE MID-1980S IN BEHAVIOR THERAPY. BUT I AGREE THAT CRITIQUES AND REJOINDERS HAVE BEEN UNCOMMON IN THIS JOURNAL.

> >

> >5) Is EMDR the first new paradigm in the field of psychology since the

> >advent of psychoanalysis, as you suggest?

>

> -----Sorry Rich, you misunderstood me. EMDR is absolutely not "the

> first new paradigm." All of the ones you mentioned are obviously true

> paradigms, as are many, many more. That's why I've never understood the

> problem with offering EMDR as governed by an information processing

> paradigm instead of its earlier desensitization formulation.

There are

> tenets that predict effects and guide its use in clinical practice. I had

> no intention of stating it was "a paradigm shift" in the Kuhnian sense that

> seems to have generated so much heat--*obviously* since it had just come on

> the scene. My previous point was, according to Wolpe and others, the field

> had not *pervasively* shifted from psychodynamic thought
to behaviorism, or
> cognitive therapy, etc.
> >

WITH ALL DUE RESPECT, FRANCINE, IT IS VERY,
VERY DIFFICULT FOR THE READER OF YOUR
PREFACE *NOT* TO CONCLUDE THAT EMDR IS NOT
THE "FIRST PARADIGM SINCE FREUD." IN ANY
EVENT, YOUR CLARIFICATION IS MUCH
APPRECIATED. YES, I KNOW WOLPE EXPRESSED
CONSIDERABLE EXASPERATION AT WHAT HE
REGARDED AS THE PERSISTENCE OF
PSYCHOANALYTIC APPROACHES IN PSYCHIATRY,
AND I WAS STARTLED SEVERAL MONTHS AGO
WHEN DON KLEIN TOLD ME HOW ENTRENCHED THE
PSYCHOANALYSTS WERE AMONG THE CLINICAL
SUPERVISORS AT *COLUMBIA UNIVERSITY* -
SURELY A WORLD-CLASS CENTER OF
EXPERIMENTAL PSYCHOPATHOLOGY AND
PSYCHOPHARMACOLOGY IF THERE EVER WAS ONE!

REGARDING KUHN AND PARADIGMS IN OUR FIELD:
BEFORE HE GOT REALLY SICK, HE GAVE A GREAT
TALK IN OUR HISTORY OF SCIENCE DEPARTMENT,
AND DURING THE QUESTION AND ANSWER PERIOD,
SOMEONE ASKED HIM WHETHER PSYCHOLOGY OR
THE OTHER SOCIAL SCIENCES HAD ACHIEVED
PARADIGMATIC STATUS (SINCE THE PUBLICATION
OF HIS 1962 BOOK). HE WAS QUITE DISMISSIVE, AND
FLATLY STATED "NO."

>
> >
> >6) One of the articles slated to appear in the Journal of
Anxiety
> >Disorders special issue is a randomized trial that compares
Foa-style CBT
> >for PTSD with EMDR. The patients are civilians, plenty of
sessions were
> >used, etc., and in fact the authors were especially attentive
to the
> >methodological issues that you say trouble previous studies
(*including*,
> >I might add, the several new "positive" studies you cite, that
are quite
> >flawed, in my opinion).

> >
> >The results: Both EMDR and CBT produced significant improvements in these
> >civilian PTSD cases, with CBT doing better than EMDR. Both treatments
> >were tolerated equally well. At follow-up, EMDR cases had started to
> >relapse, whereas CBT cases continued to improve.
>
> -----Thank you for bringing it to my attention. Before I make any
> judgment about it I'd want to check out a variety of things including the
> number of homework hours, how fidelity was established, etc. The latter is
> especially significant because no other EMDR civilian PTSD study has
> reported any relapses. I think we are in for a very interesting new decade
> of studies since three others I know of (that are being prepared for
> publication) compare EMDR to CBT and found just the opposite results to the
> one you cite.
>
> >
> >7) What Wolpe told you about eye movement therapy differed from what he
> >told me (the conversations, however, may have occurred quite separate in
> >time). Interestingly, his last article was on EMDR, and we will see
> >whether you continue to "bless his heart" when it appears.
>
> -----Yes, I know. Wolpe always resisted my adding the "R" and
> abandonment of the desensitization model and inclusion of the cognitions.
> We are all entitled to our biases.
> >
>
> >8) There are many competing theories of why exposure therapy works, and
> >you oversimplify matters quite a bit. As Jerry Rosen noted, brief
> >exposures can be therapeutic as well as long ones, and nothing in "the"

> >extinction/habituation model indicates that EMDR is anything but another
 > >exposure variant.
 >
 > -----Sorry, but what Jerry Rosen describes in Wolfe's article is nothing
 > like EMDR, e.g., clients are not asked to "intensely concentrate", "themes"
 > are not elicited in multiple sessions--and EMDR processing/desensitization
 > effects occur within the first minutes of treatment. However, rather than
 > belaboring the point, I'll recommend a reading of my review of what I call
 > the "dosed exposure" element of EMDR (Shapiro, 1995, pp. 311-312).
 > Obviously exposure is an element of EMDR treatment. It's simply not the
 > only element--nor is it conducted in any way recommended by prolonged
 > exposure advocates in the treatment of PTSD. Just as an aside, its
 > interesting that arriving at one agreed upon theory of how exposure works
 > is not made a prerequisite for acceptance of the utility of exposure
 > methods. I'd suggest the same be acknowledged for EMDR.
 >
 > Thanks for the Craske reference.

YES, MICHELLE CRASKE'S BOOK IS VERY, VERY GOOD.

>And I've always enjoyed Voltaire. Wasn't Candide the premier volume for overcoming the potential for PTSD?

YES! SURELY, IF ANYONE HAD EVER BEEN EXPOSED TO "MULTIPLE CRITERION A EVENTS" IT WAS CANDIDE!

YOU HAVE A GREAT WEEKEND, TOO!

RICH

Letters
from
SSCP net No 11 January
99

Evidenced-based
practice and
psychology

*ed
horst u kaechele*

Date: Wed, 27 Jan 1999 13:55:06 -0600
From: "Dr. Douglas J. Tataryn" <tatarat@cc.umanitoba.ca>
To: <SSCPNET@listserv.acns.nwu.edu>
Subject: Evidenced-based practice and psychology - a request

Within the last week, someone made a posting regarding a talk or paper addressing the issue that psychologists might want to think twice before we, as a profession, fully embrace evidence-based practice as a foundation of our clinical discipline.

Can someone provide me with more details, as it may be relevant to my feedback on proposed draft of "Canadian National Standards of Practice in Psychosocial Oncology". The document makes many references the construct.

Thanks in advance

Douglas J. Tataryn, Ph.D.
Asst. Professor, University of Manitoba/
Psychologist, Manitoba Cancer Treatment and Research Foundation

Date: Wed, 27 Jan 1999 15:14:56 -0500
From: Bruce Thyer <bthyer@arches.uga.edu>
To: tatarat@cc.umanitoba.ca
Subject: Re: Evidenced-based practice and psychology - a request

I would also like to be sent posts pertaining to the pros and cons of evidence based practice in psychology.

This summer, the week of July 12th, I will be offering a five day, 9-5, NIMH-sponsored workshop tentatively titled "Research on Evidence-based Practice in Mental Health". About 20 faculty will be invited to participate. Those selected will be paid about \$800 by the NIMH to attend. Details and brochure remain to be worked out, but if list members care to send me their regular postal mailing address, when a brochure and application is available I will mail you one.

I plan to highlight the research on evidence-based therapies in a variety of disorders (schizophrenia, depression, alcoholism, cocaine addiction, panic disorder, etc.). If any listmembers care to send me recommended articles or recent reviews along these lines I would be most grateful. Feel free to recommend your own works, if appropriate.

This will be a research-focused workshop, not a clinical one. Its operative assumption is that evidence-based practice (and its synonyms) is the only defensible model of scientific therapy (re Dick McFall), and should be encouraged via graduate training ,service delivery, and empirical research. Some lecture, but hopefully lots of interactive discussion, and some time for literature searches using the vast UM resources.

See you in A-squared.

--

Bruce A. Thyer, Ph.D.
Research Professor of Social Work
School of Social Work
University of Georgia
Athens, GA 30602
(706) 542-5440-voice
(706) 542-3282-FAX
(706) 783-3853-home
<Bthyer@arches.uga.edu> email

Date: Wed, 27 Jan 1999 15:48:01 -0500
From: John Hunsley <hunch@uottawa.ca>
To: tatarat@cc.umanitoba.ca, <SSCPNET@listserv.acns.nwu.edu>
Subject: Re: Evidenced-based practice and psychology - a request

Dr. Tataryn:

I believe the posting you are referring to was one about Alan Kazdin's address at APA this summer.

As to the issue of evidence-based practice in psychology, I cannot see any other defensible alternative. If practice is not based on scientific evidence, then just what is it based on? On the other hand, I am sure there is much to be debated about how best to develop and implement evidence-based standards and practices.

John Hunsley, Ph.D., C.Psych.
Professor, School of Psychology
Director, Centre for Psychological Services
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
Phone: (613) 562-5881
Fax: (613) 562-5169
Email: hunch@uottawa.ca

Date: Wed, 27 Jan 1999 20:58:12 +0000

From: Paul Salkovskis <paul.salkovskis@psychiatry.oxford.ac.uk>
To: bthyer@arches.uga.edu, tatarat@cc.umanitoba.ca
Subject: Re: Evidenced-based practice and psychology - a request

Hi Bruce,

The current fad for "evidence based practice" has alarming elements in my view.

In particular, I fear that it equates controlled outcome trials (and even more especially the meta analyses of such trials) with science. I am currently writing a short critical paper (which I'll send when I finish) entitled "Evidence based mental health: one dimensional science". I find myself seriously worried at the way this development, which owes a great deal more to the "sound bite" than to serious science, has taken root and is uncritically accepted as providing a "gold standard". (cf the Cochrane Collaboration stuff). Many find it hard to articulate why they disagree and even more problematic to do so; to disagree with something called "evidence based practice" is rather like disagreeing with something called "freedom and democracy". I don't disagree with "evidence based practice"; however, I do disagree with the uncritical acceptance of all "evidence based practice" practices.

Yours glumly
Paul Salkovskis

Paul Salkovskis,
University of Oxford Department of Psychiatry,
Warneford Hospital,
Oxford, OX3 7JX, UK

Tel +44 (0)1865 226475
Fax +44 (0)1865 226234

email: paul.salkovskis@psych.ox.ac.uk

Date: Wed, 27 Jan 1999 16:21:17 EST
From: BOBAPPIC@aol.com
To: paul.salkovskis@psychiatry.oxford.ac.uk,
Subject: Re: Evidenced-based practice and psychology - a request

I am a staunch supporter of the notion of basing practice upon the best available concepts and processes, as defined by empirical support. But I agree that the practice is more than a catchword, and more than choosing interventions from any "pre-approved" list. We need to get off our duffs and articulate just what good evidence- based practice is, how it's done, and how it's taught.

Bob Klepac
bobappic@aol.com

Date: Wed, 27 Jan 1999 14:11:15 -0800 (PST)
From: "Jacqueline B. Persons" <persons@itsa.ucsf.edu>
To: BOBAPPIC@aol.com
Subject: Re: Evidenced-based practice and psychology - a request

I agree with Bob that EBP seems vital. But I am intrigued by Paul Salkovskis' objections; I'm sure they are intelligent and thoughtful and look forward to hearing them in detail when his paper is done.

I agree with Bob (and Michael Addis who made this point recently on the sscpnet) that many of the steps required to make links between results of randomized trials and evidence-based treatment of the individual case have not yet been spelled out. I'd like to take this opportunity to say that I have some ideas about this issue and I'll tackle this subject in my sscp presidential address, at the APA in Boston in August. My current title is "The Nuts and Bolts of Evidence-based Practice."

Jackie Persons

Jacqueline B. Persons, Ph.D., Director
Center for Cognitive Therapy
5435 College Avenue
Oakland, CA 94618
(510) 652-4455
fax (510) 652-3872

Date: Wed, 27 Jan 1999 22:52:47 +0000
From: Paul Salkovskis <paul.salkovskis@psychiatry.oxford.ac.uk>
To: persons@itsa.ucsf.edu, BOBAPPIC@aol.com
Subject: Re: Evidenced-based practice and psychology - a request

At 14:11 27/01/99 -0800, Jacqueline B. Persons wrote:

>I agree with Bob that EBP seems vital. But I am intrigued by Paul
>Salkovskis' objections; I'm sure they are intelligent and thoughtful and
>look forward to hearing them in detail when his paper is done.

Whoa, everyone. As I said (repeated below) I'm not arguing against basing one's practice on evidence; as a dyed-in-the-wool scientist/practitioner/clinical scientist this is my primary orientation. However, I consider it crucial that clinical science be multifaceted, and not exclusively focussed on the best recent meta analysis. My point about the difficulty in suggesting that "evidence based mental health" may be flawed is crucial. The Soviet Union believed in freedom and democracy; so why didn't we give in to it, being such a good thing? One man's convincing evidence is another's partial answer.

Witness the discussions of EMDR which went on not too long ago. Anyway, thanks to those of you who back-channeled asking for a preprint; seems like a good stimulus to stop emailing and get writing papers!!

best wishes less glumly
paul salkovskis

Paul Salkovskis,
University of Oxford Department of Psychiatry,
Warneford Hospital,
Oxford, OX3 7JX, UK

Tel +44 (0)1865 226475
Fax +44 (0)1865 226234

email: paul.salkovskis@psych.ox.ac.uk

Date: Wed, 27 Jan 1999 18:27:21 -0500
From: "John W. Bush" <jwb@alumni.stanford.org>
To: BOBAPPIC@aol.com

Bob,

I would think that, among other things, an adequate definition of evidence-based practice would have to take account of these facts:

1. The scientific findings at our disposal have to be adapted to the treatment of particular clients at particular times. They are not self-applying. This is true even when there is a well-designed protocol available, e.g., such as we now have for most of the anxiety disorders.
2. We often have to go beyond the available scientific evidence and simply use our best judgment and common sense.
3. The evidence is sometimes contradictory (or at least appears to be so), yet action needs to be taken now.
4. We do not, however, want to let considerations such as these be used -- as they quite typically are -- as excuses for ignorance and undisciplined intuition. It will be quite a challenge finding where to draw the line.

Physicians, surgeons and dentists are faced with similar ambiguities. However, there is a difference between conducting psychotherapy and choosing an antibiotic or filling teeth. The latter tasks are usually better understood, defined and structured. Consequently our job is tougher in this respect.

There's an article in the current issue (Feb. 1) of The New Yorker that I think has some relevance here. The author, a surgeon, makes the point that most

medical and surgical errors -- even the worst ones -- are not made by lazy or unethical doctors, but by fallible good ones, often working in poorly designed environments. Of particular interest is his account of the radical reduction over the past several years in the morbidity and mortality rates in anesthesiology. Well worth a read, IMO.

--John

John W. Bush
207 Berkeley Place
Brooklyn, NY 11217-3801

Phone: 718 636-5071
Fax: 718 636-5166
Email: jwb@alumni.stanford.org
Web: <http://www.cognitivetherapy.com>

Date: Thu, 28 Jan 1999 07:47:02 -0600 (CST)
From: "Jeffrey M. Lohr" <jlohr@comp.uark.edu>
To: Paul Salkovskis <paul.salkovskis@psychiatry.oxford.ac.uk>

I agree with Paul about the limits of meta-analysis to tell us what the evidence is. In the past, it has been used erroneously (see Eysenck, 1994) to give lie to the Dodo Bird verdict. I, for one, tend to view it as a statistical technology, at best. Science it is not. Much more careful attention to procedural and experimental controls in a series of programmatic studies will inform us much more clearly of the limits of reliability, validity, and efficacy.

Jeffrey M. Lohr, Ph.D.
Department of Psychology
216 Memorial Hall
University of Arkansas
Fayetteville, Arkansas 72701
Phone (501) 575-5813
Fax (501) 575-3219

"Forma omni, atqui re sine"

Date: Thu, 28 Jan 1999 10:20:30 -0500
From: "James Coyne" <jcoyne@umich.edu>
To: bthyer@arches.uga.edu, tatarat@cc.umanitoba.ca,
paul.salkovskis@psychiatry.oxford.ac.uk
Subject: Re: Evidenced-based practice and psychology - a request

My complaint against evidence-based practice is that the quality and quantity of evidence does not yet put us in a situation to make very many useful distinctions.

1. As I argue in a forthcoming "Treatment and Prevention" piece, Evidenced-based practice advocates/e.v.t. folks talk effectiveness but largely limit themselves to efficacy research conducted with unrepresentative samples. Evidence of generalization to more typical patients is meager at best. No one has ever shown that under conditions of routine care, anything works for depression in primary care, for instance, although I have tried twice myself.

2. Results of the bulk of the available studies are better predicted by investigator allegiance than the approach being evaluated. Note Luborsky et al's forthcoming paper indicating allegiance explains 70% (!) of the variance in outcome in 29 such studies.

3. How empirical evidence has been accumulated in favor of a particular approach (as in Dodson's claim that cognitive therapy is the most validated therapy for depression) is at this very early stage of the game a matter of politics, not favorable versus unfavorable evaluations. See Helay's "Antidepressant era" for this argument.

4. Current evidenced-based practice standards are ineffective block nonsense claims about particular approaches and may even selectively be invoked to support nonsense claims (such as De Jongh's repetitious, just-like-a-Colonel-Saunders-of-the-Internet's 'scientifically speaking, finger wiggling good, scientifically speaking, finger wiggling good' and the outrageous hype of EMDR workshops and its paid (\$500 a pop) bellowing hitmen.

Date: Thu, 28 Jan 1999 11:01:05 -0500
From: John Hunsley <hunch@uottawa.ca>
To: jcoyne@umich.edu
Subject: Re: Evidenced-based practice and psychology - a request

Jim (James Coyne w):

Although I agree that we need to be cautious about the conditions under which we generalize from efficacy trials to routine practice, I think that the situation is changing rapidly with regards to the "generalizability dilemma", as Goldfried and Wolfe have called it. Certainly until a few years ago there were relatively few studies conducted until "normal" clinical conditions (i.e., types of clients, referral sources, types and training of therapists), as witnessed by the Shadish et al. (1997) meta-analysis in JCCP (56 studies that were deemed clinically representative). However, open any issue of JCCP or

Behavior Therapy in the last couple of years and you are likely to encounter studies that use clients/patients in hospital or institutional settings. Furthermore, although there are relatively few published studies to date, what has been published suggests that there are reasons for being fairly optimistic that many efficacious treatments can be transported to routine practice and be effective, at least for some treatments for adults.

John

John Hunsley, Ph.D., C.Psych.
Professor, School of Psychology
Director, Centre for Psychological Services
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
Phone: (613) 562-5881
Fax: (613) 562-5169
Email: hunch@uottawa.ca

Date: Thu, 28 Jan 1999 09:17:42 -0700 (MST)
From: Lee B Sechrest <sechrest@U.Arizona.EDU>
To: sscpnet@listserv.acns.nwu.edu
Subject: Re: Evidenced-based practice and psychology - a request

OK, so there are problems with the evidence. No question. So what is the solution?

Evidence based nonpractice? That is, on the evidence there is no basis for practice, so don't do it.

Non-evidence based practice? We have that now, e.g., all those fringe therapies.

Non-evidence based nonpractice? Collapse into nihilism?

I am serious, you know. We either slink away and let every individual practitioner do whatever he or she imagines might be best, or we begin with the evidence we do have (Cochrane-type committees in the mental health area would not be all that bad an idea) and proceed iteratively with modest claims and immodest effort to bolster them.

I vote for regular and rigorous reviews of literature and recommendations based on the best evidence we have. Even if that may make Paul glum.
Lee

Date: Thu, 28 Jan 1999 11:41:01 -0500
From: "John W. Bush" <jwb@alumni.stanford.org>
To: jcoyne@umich.edu
Subject: Re: Evidenced-based practice and psychology - a request

I assume Luborsky's paper covered only published studies. If so, unless investigators try equally hard to publish results that reinforce their allegiances and those that don't, the 70% figure may be spuriously high. A further possible source of inflation in the explained-variance statistic would be classical publication bias, i.e., the tendency to favor studies that produce clear-cut or novel results.

Luborsky's own paper may be a case in point. His allegiance is largely psychodynamic, and his finding of strong allegiance effects tends to undermine the current pre-eminence of cognitive and behavioral therapies in efficacy research. Would he have sought publication of this study if it had revealed weak or negative allegiance effects? And if he had, would it have been accepted?

John W. Bush
207 Berkeley Place
Brooklyn, NY 11217-3801

Phone: 718 636-5071
Fax: 718 636-5166
Email: jwb@alumni.stanford.org
Web: <http://www.cognitivetherapy.com>

Date: Thu, 28 Jan 1999 10:59:01 -0600 (CST)
From: "Jeffrey M. Lohr" <jlohr@comp.uark.edu>
To: Lee B Sechrest <sechrest@U.Arizona.EDU>
Subject: Re: Evidenced-based practice and psychology - a request

Sounds like good, traditional, academic science to me.

Call us old fashioned, but.....

Jeffrey M. Lohr, Ph.D.

Date: Thu, 28 Jan 1999 12:18:26 -0500
From: "David A. F. Haaga" <dhaaga@american.edu>
To: "John W. Bush" <jwb@alumni.stanford.org>
Subject: Re: Evidenced-based practice and psychology - a request

John W. Bush wrote:

.....>

> Luborsky's own paper may be a case in point.

--Sure it would. See Gaffan et al. (1995) JCCP 63 966-980 Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. This study led to conclusion that allegiance effects (in specific case of CT for depression) existed but were in decline in more recent literature.

In general, I think therapy outcome research is, among all realms of psychological research, about the least susceptible to "prejudice against the null hypothesis" (A. Greenwald's term). This applies to primary studies (e.g., Gortner, Gollan, Dobson, & Jacobson 1998 Cognitive-behavioral treatment for depression: Relapse prevention, JCCP, 66, 377-384) and reviews (see Wampold et al. 1997 A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes" Psychological Bulletin, 122, 203-215, plus two peer commentary papers and author reply.

I haven't seen the Luborsky et al. paper, but assuming (based on the standards of the editor of the journal in which it will appear, the author's competence, and Jim Coyne's endorsement) it is a methodologically sound study of allegiance effects, I am confident it could have been published in a reputable journal with large, moderate, small, or even no allegiance effect as the finding.

Best wishes,

Dave Haaga
Assoc. Editor, JCCP

Date: Thu, 28 Jan 1999 14:06:49 -0500
From: "James Coyne" <jcoyne@umich.edu>
To: jwb@alumni.stanford.org, dhaaga@american.edu
Subject: Re: Evidenced-based practice...replies

Of course, if there is good solid empirical evidence for a particular treatment approach, follow the advice of it, particularly when there is evidence against alternative approaches.

The problem is the evidence for differential effectiveness is limited. Much of the clamor for evidence-based approaches is a politically motivated effort to advance stronger claims than the evidence really justifies. Or to make nonsense claims such as for EMDR. It is not typically a way to eliminate false claims, at least at this point. to cut away. Those are my beefs.

I love EMDR in the one respect of it having clearly demonstrated that a move to evidence-based practice is insurance against bullshit.

Six available studies show detected depressed patients do not get better in primary care quicker than missed patients. Among them:

1. Coyne, J.C., Klinkman M.S., Gallo S.M., & Schwenk, T.L. (1997). Short-term outcome of detected and undetected depressed primary care patients and depressed psychiatric patients. *Gen Hosp Psychiatry*, 19, 333-343.

2. Rost, K., Zhang, M., Fortney, J., Smith J., & Coyne, J. C. (1998). Persistent poor outcomes of undetected major depression in primary care: implications for intervention. *General Hospital Psychiatry*.

I was surprised to find out that most of a representative sample of depressed patients recruited in a psychiatric research setting would not qualify for recruitment into a clinical trial. And that their level of depression according to the HAM-D was in a range in which pill-placebo is as good as antidepressants or psychotherapy. See

Schwenk, T.L., Coyne, J.C. & Fechner-Bates, S. (1996). Differences between detected and undetected depressed patients in primary care and depressed psychiatric patients. *General Hospital Psychiatry*, 18, 407-415.

Has anyone tried to get to urban primary care patients into a clinical trial providing free psychotherapy? I think it is easier to get volunteers for a free sigmoidoscopy. If what we offer is so wonderful, how come we cannot give it away?

Date: Thu, 28 Jan 1999 14:07:24 -0500
From: "John W. Bush" <jwb@alumni.stanford.org>
To: dhaaga@american.edu
Subject: Re: Evidenced-based practice and psychology - a request

Dave,

That's reassuring (though Luborsky's data aren't).

Date: Thu, 28 Jan 1999 15:10:51 -0500
From: John Hunsley <hunch@uottawa.ca>
To: jcoyne@umich.edu, jwb@alumni.stanford.org, dhaaga@american.edu
Subject: Re: Evidenced-based practice...replies

Jim:

You asked the following:

>Has anyone tried tot get to urban primary care patients into a clinical
>trial provding free psychotherapy? I think it is easier to get
>volunteers for a free sigmoidoscopy. If what we offer is so wonderful,
>how come we cannot give it away?

I am sure you know the following article, but here's the reference anyway.

Munoz, R. F., Ying, Y. W., Bernal, G., Perez-Stable, E. J., Soresen, J. L.,
Hargreaves, W. A., Miranda, J., & Miller, L. S. (1995) Prevention of
depression with primary care patients: A randomized controlled trial.
American Journal of Community Psychology, 23, 199-213.

Regards,
John

John Hunsley, Ph.D., C.Psych.
Professor, School of Psychology
Director, Centre for Psychological Services
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
Phone: (613) 562-5881
Fax: (613) 562-5169
Email: hunch@uottawa.ca

Date: Thu, 28 Jan 1999 20:56:28 +0000
From: Paul Salkovskis <paul.salkovskis@psychiatry.oxford.ac.uk>
To: sechrest@U.Arizona.EDU, sscpnet@listserv.acns.nwu.edu
Subject: Re: Evidenced-based practice and psychology - a request

Boy did I open Pandora's box.

Thanks for all the responses to this, but I really was not saying that we should not use evidence. I was talking about "Evidence based mental health" as an offshoot of "Evidence based medicine", the ideal articulated by David Sackett. I am saying that it is NOT ENOUGH. Oops, probably was shouting, headcold and glum together. The idea that Cochrane Groups have the golden vote sticks in my craw, because systematic reviews and meta analysis are very very bendable, and can be biased. You know the stuff, on average swans are white, therefore there are no black swans.... a black swan should refute the proposition, a replicated black swan clinches it. Meta analyses and systematic reviews are great when there are loads of great big studies, and thats where sackett was right. When there are a handful of studies, and 30percent get disallowed because of xy and z, and there have been recent radical

developments, the results are dull, dull, dull and fail to reflect the state of the art. The effort and money going into basically meaningless systematic reviews appalls me, and I wish that those doing these things would spend the time doing studies, preferably crucial studies. Mental health work is not yet at the stage of needing this approach, and may become distorted as a result (see for example the suicide review done recently). The notion of "the weight of evidence" curdles my blood.....I have heard it seriously suggested at conferences that the numbers of patients treated in controlled trials by psychological means is far less than those in pharmacotherapy studies, suggesting that pharm is better. Hmm, whats wrong with that? Then there's the issue of overvaluing systematic reviews/meta analysis relative to experimental programmes, which is yet another funding issue.

Lee, there are cochrane mental health groups (the schizophrenia one is in our dept). A senior cochrane person referred me a patient for one of my current RCTs. However, he also prescribed the patient fluoxetine at the same time, and was surprised when I pointed out the problem. My point is that this stuff has become a bandwagon and the only show in town, with wonderkind leaping aboard without the first idea of scientific principles in RCTs let alone the bigger, multidimensional picture. We now have "pragmatic trials" taking huge slices of funding. and on and on. Might as well write the paper here and now! I'm arguing for the science of clinical psychology. I suspect you know the work of Monte Shapiro; much of what he said in the 50s and 60s is apposite now. EVTs is one dimension, but beware of assuming its the whole cake, there's more than one baby in that bathwater.
Yours with mixed metaphors....

Paul Salkovskis

Paul Salkovskis,
University of Oxford Department of Psychiatry,
Warneford Hospital,
Oxford, OX3 7JX, UK

Tel +44 (0)1865 226475
Fax +44 (0)1865 226234

email: paul.salkovskis@psych.ox.ac.uk

Date: Thu, 28 Jan 1999 16:40:33 -0500
From: "James Coyne" <jcoyne@umich.edu>
To: jwb@alumni.stanford.org, dhaaga@american.edu, hunch@uottawa.ca
Cc: SSCPNET@listserv.acns.nwu.edu

I am very familiar with the San Francisco study and by the time they implemented a system insuring that patients could/would make it to sessions, it was no longer recognizably primary care. They got babysitting,

transportation and other assistance that was otherwise missing in their lives and undoubtedly quite an intervention by itself.

John Hunsley, Ph.D., C.Psych.
Professor, School of Psychology
Director, Centre for Psychological Services
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
Phone: (613) 562-5881
Fax: (613) 562-5169
Email: hunch@uottawa.ca

Date: Thu, 28 Jan 1999 23:46:10 -0500
From: "John W. Bush" <jwb@alumni.stanford.org>
To: paul.salkovskis@psychiatry.oxford.ac.uk
Subject: Re: Evidenced-based practice and psychology - a request

Paul,

It's been a while since I looked into evidence-based medicine, and I didn't look very deeply then. But I did carry away the impression that it entails some rigorous reasoning about data sets of a kind that are rarely available in mental health. If so, besides not being enough this model may also be too much. But maybe I've got it wrong....

Meanwhile, what are Cochrane groups? They sound absolutely awful.

--John

John W. Bush
207 Berkeley Place
Brooklyn, NY 11217-3801

Phone: 718 636-5071
Fax: 718 636-5166
Email: jwb@alumni.stanford.org
Web: <http://www.cognitivetherapy.com>

Date: Thu, 28 Jan 1999 11:55:38 -0500
From: Howard Eisman <howeis7@idt.net>
To: SSCP <sscpnet@listserv.acns.nwu.edu>
Subject: Re: Evidence Based Practice and Psychology

Do current research findings allow a clinician to select a form of therapy for a particular "real world" patient with confidence in its chances of helping the patient?

As a researcher (and academncian), I would have to say that, as yet, few forms of psychotherapy have much empirical support when used with patients in hospital and clinics. The really rigorous research in actual clinical settings is only just beginning.

As a clinician and hospital administrator, I am responsible for delivering the best possible services to patients. I can not refuse services to a patient because there is no form of psychotherapy which has been tested with rigorous research methods and shown to be effective for that patient's particular problems. I must make choices based on the quite imperfect (and politically driven) research which now exists. I must make the assumption that the body of current research does give me some information which I can use to make a choice.

The only alternative to this seems to be the old practice of allowing clinicians to do whatever they please based on their "gut" reaction to a patient or to be quacks using the therapy to which they have allegiance for all their patients.

It seems to me that we have to make the most of the flawed research which now exists until good research is done.

Howard D. Eisman, Ph.D

Date: Wed, 27 Jan 1999 16:00:02 -0600
From: Kenneth Howard <k-howard@nwu.edu>
To: sscpnet@listserv.acns.nwu.edu
Subject: Evidence-based practice

It is hard for me to believe that there are professional practices with no evidence base or empirical support. We must be arguing about what constitutes "evidence" and what qualifies as empirical.

Professor Kenneth I. Howard
Department of Psychology
Northwestern University
2029 Sheridan Road
Evanston, IL 60208-2710
Phone: (847) 491-7373
Fax: (847) 491-7859
e-mail: k-howard@nwu.edu

Date: Wed, 27 Jan 1999 16:22:44 -0600 (CST)
From: "Jeffrey M. Lohr" <jlohr@comp.uark.edu>
To: Kenneth Howard <k-howard@nwu.edu>
Subject: Re: Evidence-based practice

Believe it. Look at the classified adds in the APA Monitor and the National Psychologist, along with the yellow pages in your local phone directory.

There is much more non-empirically based practice than the legitimate alternative.

Jeffrey M. Lohr, Ph.D.
Department of Psychology
216 Memorial Hall
University of Arkansas
Fayetteville, Arkansas 72701
Phone (501) 575-5813
Fax (501) 575-3219

Date: Fri, 29 Jan 1999 08:53:31 -0800
From: jmcquaid@ucsd.edu (John McQuaid)
To: Ricardo_F._Munoz@sfgh.org
Subject: Re: Evidenced-based practice...replies (2)

Hi Ricardo-

Your prevention study came up on the SSCPNET (I don't know if your a member). I was about to respond to some of Dr. Coyne's comments which I have pasted in below, but I thought I would forward this to you for your consideration first. I pulled out your article and didn't find a description of assistance or treatment refusal, and I have lent my copy of your book to a student who is developing a tertiary prevention project with me. Anyway, I thought that Coyne's pessimism on recruitment for treatment trials is interesting.

Congratulations on the paper! It is very nice to have it out and I'm really glad we tagged at AABT. On a related note, I received the Depression Risk Checklist data from Charlene (I hope she's still working out well for you folks), and I'm starting to put together a paper based on your data and our data down here. If I remember correctly, we had hoped to coauthor a paper describing the utility of the instrument in predicting onset of depressive episodes. Are you still interested? I'm getting faster at turning papers around myself, and I'm getting more assertive about pursuing co-authors dragging their feet, so I think it could be done in a reasonable amount of time. I think the thing I'm learning most on this job is to always put writing first on the list

of things to do, because everything else just forces it's way into being done. Since I've taken that tack, things are moving a little better.

Enough rambling. I would be happy if your interested in pursuing this paper, but certainly understand if you have a full plate. On a final note, Sarah and I will be in the Bay Area the evening of 3/12 through Sunday 3/14. My father's 60th birthday is April 19th, but they are having a St. Patrick's day/surprise birthday for him on 3/13. The early date is reflective of my family's morbid sense of humor; no McQuaid male has made it past 60 in 3 (maybe 4) generations, and my father's younger brothers are going to announce to him that they scheduled an early birthday to make sure that he sets a precedent for them.

We would like to stop by and say Hi if you have time on Sunday. let us know if that would work. Otherwise, I am planning to be up there a couple times this summer as well, so we may stop by then.

Take it easy!

John

John R. McQuaid, Ph.D.
Psychology Services (116-B)
San Diego VA Medical Center
3350 La Jolla Village Drive
San Diego, CA 92161
office: (619) 552-8585, x3693
fax: (619) 552-7414
email: jmcquaid@ucsd.edu

Date: Fri, 29 Jan 1999 08:57:17 -0800
From: jmcquaid@ucsd.edu (John McQuaid)
To: SSCPNET@listserv.acns.nwu.edu
Subject: Re: Evidenced-based practice...replies (2)

Dear SSCPNETers-

My apologies for accidentally sending a clearly personal reply to the net.

John R. McQuaid, Ph.D.
Psychology Services (116-B)
San Diego VA Medical Center
3350 La Jolla Village Drive
San Diego, CA 92161
office: (619) 552-8585, x3693
fax: (619) 552-7414
email: jmcquaid@ucsd.edu

Date: Fri, 29 Jan 1999 08:18:05 -0900
From: "Gregory J. Meyer" <afgjm@UAA.ALASKA.EDU>
To: "John W. Bush" <jwb@alumni.stanford.org>
Subject: Re: Evidenced-based practice and psychology - a request

fyi: As Paul well-knows, starting last year, the BMJ Publishing Group began the journal "Evidence-Based Mental Health: Linking Research to Practice." It follows the same general format as Evidence-Based Medicine and the ACP Journal Club, with abstracts of articles meeting criteria for relevance followed by brief commentary.

Greg

Gregory J. Meyer
Psychology, U of Alaska Anchorage
3211 Providence Dr, Anchorage, AK 99508
907-786-1741 (ph) 907-786-4898 (fax)
afgjm@uaa.alaska.edu