South African Journal of Psychology 2014, Vol. 44(4) 381–403 © The Author(s) 2014
Reprints and permissions: sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0081246314533750 sap.sagepub.com
SAGE

PsySSA
Psychological Society of South Africa

State of the Science

Evidence-based psychotherapies I: qualifiers and limitations in what we know

Alan E Kazdin

Abstract

Psychosocial interventions have advanced remarkably for the treatment of a broad range of psychiatric disorders and related sources of impairment among children, adolescents, and adults. By one count, over 320 interventions have been identified as evidence-based in light of rigorously controlled studies and replication of treatment effects. Yet, how evidence-based psychotherapies have been evaluated and reported raises questions about their impact. This article evaluates both methodological and substantive issues that limit what can be stated about evidence-based psychotherapies and their effects. Among the methodological topics are the control conditions to which evidence-based psychotherapies are compared, selective reporting of measures, and the limited evidence that evidence-based psychotherapies have clinically significant impact. Among the substantive issues are the paucity of research on moderators that would help us better direct patients to treatments from which they are likely to profit and our limited understanding of the mechanisms responsible for therapeutic change. The issues discussed are fundamental to what can be stated about the impact of evidence-based psychotherapies and impact and the bases for their effects. It is not clear at present whether concerted efforts are in place to alter research in ways that would redress the issues. There are, however, novel new directions for research that build on the evidence-based psychotherapies, and these are addressed in a companion article.

Keywords

Evidence-based psychotherapy, limitations of Evidence-based treatments, psychotherapy outcome research

The development of psychosocial interventions for the treatment of psychological dysfunction reflects a remarkable contribution that has emerged over a few decades. Currently, many

Department of Psychology, Yale University, USA

Corresponding author:
Alan E Kazdin, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520-8205, USA.
Email: alan.kazdin@yale.edu

interventions have been delineated as evidence based.[1] The criteria used to designate treatment as evidence based vary among multiple disciplines (e.g., psychology, psychiatry, social work), diverse professional organizations within and among countries (e.g., in the Americas, European Union), and private and public agencies within a given country. Although there is no single consensus definition about what it means to be an evidence-based psychotherapy (EBP), for the most part, several of the following criteria are commonly invoked:

- Careful specification of the patient population;
- Random assignment of participants to intervention and comparison or control conditions;
- Use of treatment manuals that document the procedures;
- Multiple outcome measures (raters, if used, are naive to conditions), including, of course, measurement of the problem or disorder targeted in treatment;
- Statistically significant differences at the end of the intervention period between treatment and control or comparison conditions; and
- Replication of outcome effects, ideally by an independent investigator or research team.

Drawing on these criteria, one count has identified over 320 EBPs for mental health and substance abuse disorders (US Department of Health and Human Services [USDHHS], 2014). EBPs encompass interventions for children, adolescents, and adults and have been elaborated and evaluated in books (e.g., Christopherson & Vanscoyoc, 2013; Nathan & Gorman, in press; Weisz & Kazdin, 2010) and journal articles (e.g., Baldwin, Christian, Berkeljon, & Shadish, 2012; Cavanagh, Strauss, Forder, & Jones, 2014; Richards & Richardson, 2012).

The need for EBPs has become salient with parallel advances in psychiatric epidemiology and improved documentation of the high rates and enormous burdens of mental illness. For example, surveys from the World Health Organization (WHO) assessing the global burden of mental illness found the rates of lifetime psychiatric disorder among 17 participating countries to range from 12.0% to 47.4%, with the highest lifetime prevalence estimate in the United States (Kessler et al., 2009; Kessler & Wang, 2008). Apart from prevalence, the personal and social burdens of mental disorders are astounding. Mental disorders are more impairing than common chronic medical disorders (Druss et al., 2009). For example, in 2004, the burden of depressive disorders (e.g., years of good health lost because of disability) was ranked third among the list of mental and physical diseases worldwide (World Federation for Mental Health, 2011). By 2030, depression is projected to be the number one cause of disability, ahead of cardiovascular disease, traffic accidents, chronic pulmonary disease, and HIV/AIDS (WHO, 2008).

With these considerations in mind, the need for effective ways of addressing mental illness is more salient than ever. EBPs are intended to have an impact on clinical dysfunction, both at the level of the individual and, of course, society at large. Addressing mental health needs entails more than developing effective interventions. Once developed, other challenges emerge such as training the workforce to deliver the interventions and ensuring that those in need of services actually receive them. Yet, a prerequisite for having impact is establishing that one has interventions to offer, with evidence on their behalf. The very term "evidence based" suggests that we are very far along in developing interventions that have impact. We have treatments with an evidence base, but what we can say about these treatments and their impact?

The purpose of this article is twofold. First, the article evaluates how EBP research is conducted and the implications of that for what can be said about the impact of the interventions. Second, the article identifies key research topics that would place EBPs on firmer footing both in relation to research and clinical practice. The overall goal is not only to identify fundamental limitations about what we know but also to go beyond to suggest new directions. This article focuses

on the limitations; a companion article (Kazdin, in press) will address current and new directions that can augment the impact of our interventions.

What conclusions may we draw from the evidence?

With an accumulating pile of EBPs, one just assumes that, in fact, we have effective interventions for a variety of problems. And, of course, in some ways we do. Yet, fundamental questions about "effective" treatments are not resolved and in most cases are not really receiving much attention. Consider the evaluation from two perspectives: first methodological and then substantive questions. We begin with the methodological issues and look at three topics: the control and comparisons that are used in EBP studies, the measures that are used and reported to draw conclusions, and the outcomes of treatment and what they show and more importantly do not show.

Control and comparison groups to establish EBPs

The control and comparison groups that are used to evaluate EBPs warrant special scrutiny because they are directly related to what can be said when a treatment is shown to be effective. Typically, one of three comparison groups is used.

No-treatment control group. A standard requirement for identifying an EBP is to show that a treatment is better than no treatment (or a wait-list control condition). At the end of the clinical trial, the treatment group is shown to surpass the no-treatment group in symptom reduction and related measures. The methodological benefits of a no-treatment group are very well known. When participants are assigned randomly to treatment or no treatment, several potential methodological problems (e.g., subject-selection bias, statistical regression, effects of repeated testing, and maturational changes over time) usually become implausible as an explanation of group differences at posttreatment. Yet, there is a problem in developing an effective intervention that relates to the interpretation of any finding.

In any treatment, there are many seeming accouterments or ancillary factors in which treatment is embedded and these may contribute to or be responsible for therapeutic change. Such factors as attending treatment sessions, having personal contact with a therapist, hearing a reasonable rationale that describes the putative origins of one's problem, and undergoing a procedure directed toward ameliorating the problem may exert influence on client performance and generate their own therapeutic effects (e.g., Duncan, Miller, Wampold, & Hubble, 2010). These factors are referred to as *common* or *nonspecific factors* of psychotherapy because they are ingredients in most therapies. When we consider specific therapy techniques (e.g., cognitive behavior therapy, interpersonal psychotherapy), we usually do not know mechanisms of action or processes through which they achieve their effects. It might be specific facets of the procedures that are conceptually driven (e.g., activities and exercises directed toward change), the common or nonspecific factors, or some combinations that are responsible for change. Thus, the comparison of treatment versus no-treatment is useful as a first step but is not sufficient in providing evidence that the treatment or at least any specific facet of the treatment led to change.

The significance of nonspecific effects has emerged from different lines of research. First, we have known for some time, that an intervention, even if "fake" is likely to lead to therapeutic changes (e.g., Paul, 1966 as a landmark study, but also see Lambert & Ogles, 2013). A fake treatment is one in which some procedure is invented to provide a credible activity but is not based on theory, research, or clinical experience as a veridical intervention for the problem to which it is applied.

Second and related, we also have learned that treatment and control conditions vary in the extent to which they mobilize expectations for improvement and change. These expectations are derived from how reasonable and credible the intervention is and how likely it seems it will be as an effective intervention. Although expectations mobilized by treatment are part of the set of common factors, it is worth distinguishing between them. As I elaborate in the next section, expectations for change that are generated by a given treatment or control condition are related to the degree of therapeutic change.

Among the criteria for delineating a treatment as evidence based is establishing that the treatment is better than no treatment. This places the bar very low and virtually any treatment should pass that bar. In light of the strength of common factors and expectancies in leading to change, perhaps the only truly interesting finding would be if treatment did *not* surpass a no-treatment (or wait-list) condition. I believe it is reasonable to assert that the treatment versus no-treatment comparison by itself is not useful to delineate a treatment as evidence based because it leaves uncontrolled too many factors about treatment delivery. We can say the treatment was responsible for change, but we cannot say what it was about the specific treatment that led to change. Endless replications of the treatment effect that still rely on the no-treatment (or wait-list) comparison do not resolve the matter.

Attention placebo and credible control procedures. To address the role of nonspecific-treatment factors and expectancies, therapy research came to use control conditions that provided common factors associated with treatment (e.g., having sessions, meeting with a therapist, engaging in some procedures). This is analogous to providing a placebo pill in a study of the effects of medication. We have learned from many years of medical research that the belief in treatment is important. Placebos, inert substances (e.g., sugar tablets), given under the guise of treatment can alter a variety of disorders ranging in gravity and severity of impairment (e.g., the common cold, migraine headaches, and cancer; Benedetti, 2009; Finniss, Kaptchuk, Miller, & Benedetti, 2010; Kam-Hansen et al., 2014). Even more than that, we know that some placebos (e.g., active placebos that have side effects similar to medication) and some ways of administering placebos (e.g., larger pills vs smaller ones; injections rather than pills) even strengthen placebo effects. Belief of the patient in treatment and perhaps the belief in the physician who administers treatment and similar factors appear to be responsible for change.

The analogy of placebo effects in medicine and nonspecific-treatment effects in psychotherapy breaks down a bit. The reason is that in therapy, unlike medicine, not all parties (therapists, clients, investigators) can be kept "blind" to what the client is receiving. That is, we know who is given a particular treatment condition and that can influence the results, depending on other factors (e.g., measures, who sees the client to obtain these measures at the end of treatment). Also, treatment and control conditions in a psychotherapy study are not identical looking in appearance or procedures (as a medication and placebo usually are) except for the critical ingredient. These differences in psychotherapy and therapy-like control procedures, and the differences in demands they place on the participants can lead to differential expectancies for improvement between treatment and attention-placebo control procedures.

The extent to which treatment leads to change is very much related to how credible the treatment is to clients and the expectances that treatment and control condition mobilize. This means that merely adding a control group that includes some treatment may not be sufficient as a basis for establishing the specific effects of a veridical treatment. It may be the case that the control condition is not as credible and does not generate expectancies for change as well as the veridical treatment group to which it is compared. Indeed, highly credible control conditions often are just as effective (when compared to no treatment) as are treatment conditions (see Lambert & Ogles,

2013). In fact, the more the attention-control condition generates expectancies for improvement that approaches or equals those of the intervention group, the less likely there will be differences between the two conditions (see Baskin, Tierney, Minami, & Wampold, 2003; Boot, Simons, Stothart, & Stutts, 2013; Grissom, 1996).

The potential role of expectancies is routinely neglected in most research on EBPs. That is, expectancies among different conditions are not assessed or taken into account as a possible explanation of group differences. The neglect may be due in part to the "face" validity that many control procedures seem to have. Because control subjects received something and that something is reasonable, it may be assumed that any differences between treatment and control conditions are explained by the specific intervention practices that comprise the EBP. Yet, expectancies for change as an explanation of the group differences has parsimony on its side because group differences among many treatments, many clinical problems, and many studies might be accounted for by differential expectancies.

For EBPs, we need to know whether credibility of the procedures and the expectancies for change that the procedures generated were equal between treatment and control conditions. The matter is not whether the EBP led to change or whether the EBP led to greater change than a control condition. That is handled primarily by randomization, attention to other details (e.g., treatment integrity), and the data analyses. The possible difference in expectancies between conditions is a concern as an explanation for why the EBP worked and why many EBPs work. Now with hundreds of EBPs, it would be useful to identify how many of those and which ones have made implausible expectancies as an explanation for therapeutic change.

My comments do not claim that the effects of EBPs *are* due in whole or in part to expectancies. Rather, the claim is made that expectancies often are very plausible as an explanation of the findings or cannot be ruled out. Research to date tends to support the view that psychotherapy is more effective than nonspecific-treatment control conditions and that nonspecific-treatment control conditions are more effective than no treatment (Lambert & Ogles, 2013). Yet, the demonstrations that psychotherapy is more effective than nonspecific-treatment control conditions rarely show equal credibility and expectations of the conditions that are compared. There are various remedies that have been suggested (e.g., using techniques designed to evaluate demand characteristics in the context of psychotherapy studies, using outcome measures that are not amenable to influence by expectancies; e.g., Boot et al., 2013; Kazdin, in press b). The remedy most often selected is to include as a comparison condition a veridical treatment that is in use already, as discussed next.

Treatment as usual controls. Increasingly, EBP studies have used treatment as usual (TAU) as the comparison condition. TAU refers to the standard, routine, or usual treatment that is provided in a given setting for the same clinical problem or intervention focus that is used to evaluate the new or novel treatment. There are excellent benefits of such a comparison group, beginning with the obvious ethical advantages, that is, treatment is not withheld (no treatment, waiting list) nor is a fake or attention-placebo condition provided. Among the consequences, this means that "real" patients with real and multiple problems can be studied in clinic settings, which is less likely to be the case when control conditions are used that withhold or delay treatment or that consist of providing only a facsimile of a real treatment. In addition, dropping out of the study, which can wreak havoc in the evaluation, is less likely when "control" patients (who do not receive the EBP) are still receiving the usual care that is provided.

A difficulty for treatment evaluation and the accumulation of knowledge is that treatment "as usual" is somewhat of a misnomer. For a given clinical problem, there does not appear to be an "as-usual" psychological care that is standardized among clinics, hospitals, or private practice settings within a given city, state or province, or country. Indeed, within a given clinic, there may not

be a standard as-usual treatment. Therapists routinely individualize treatment based on their training and beliefs about the case.

The diversity of TAU is nicely illustrated in a well-controlled study for the treatment of postnatal depression in women (Mulcahy, Reay, Wilkinson, & Owen, 2010). Mothers with postnatal depression were assigned to receive group interpersonal psychotherapy or TAU. Interpersonal psychotherapy, established as an EBP for depression, focused on dealing with social isolation and feelings of loneliness, receiving sources of social support, and experimenting with new behaviors directed toward these ends. TAU consisted of a variety of treatments that were routinely available in the community, including individual therapy, group therapy, medication, natural remedies, and others. Women assigned to TAU were given written and verbal information about the local services. Assessment revealed that TAU participants accessed a range of services. The EBP was more effective than the varied services provided to the TAU group.

In relation to this article, TAU leaves unaddressed the matter of expectancies and whether they varied between conditions. It is reasonable to consider that receiving a "real" treatment from therapists at a clinic will generate expectancies for change and that treatment will be credible as an intervention. Perhaps, it is safe to say as well that TAU will be better in generating expectancies for change than an attention-placebo control condition. But "reasonable" and "safe to say" are not data based and it would be useful to know. Also, since there is an indefinite number of treatments as usual, one might not make a general comment such as "treatment as usual for anxiety has this or that effect or effect size." The main point is that TAU with its many advantages does not solve the matter of whether treatment expectancies are ruled out as an explanation of any differences between an EBP and a TAU.

There are some scientific dilemmas about TAU: they have shown the full range of outcomes in relation to EBPs; sometimes they are more, sometimes less, and sometimes equally effective as the EBPs to which they are compared (Cahill, Barkham, & Stiles, 2010; Freedland, Mohr, Davidson, & Schwartz, 2011; Wampold et al., 2011; Weisz, Jensen-Doss, & Hawley, 2006). Because TAUs are a moving target, it is difficult to codify which versions are as effective as an EBP, under what conditions, and for what clinical problems. And, that makes replication of a study with TAU difficult, if not impossible, unless conducted at the same setting with the same therapists. Finally, we still want some assurance that credibility and expectancies for change could not explain the group differences favoring one treatment versus another.

General comments. I have highlighted three comparison conditions commonly used to establish the effectiveness of EBPs. There are other comparison conditions such as a direct comparison of two treatments, a comparison of a treatment alone and in combination with another treatment, and a comparison of a treatment in its full version versus that same treatment with one or more components missing. These latter comparison groups contrast different treatments or variations of a given treatment against each other. Although rarely are expectancies for change evaluated and therefore shown to be equal across conditions, it can be assumed that viable treatments, including TAUs, as a general rule are likely to generate expectancies for improvement that are as strong as or closely approximate those provided by the EBP. If so, any treatment differences are less likely to be accounted for by expectancies.

When viable treatments are compared, differences often are difficult to detect. The absence of differences might be explained away by weak statistical power, which is the rule rather than the exception in comparative outcome research (e.g., Kazantzis, 2000; Kazdin & Bass, 1989). Weak power in comparing treatments is not unique at all to psychological interventions and applies to studies where strong control conditions are used. For example, are medications more effective than placebos in treating depression among elderly patients? Perhaps, but a review of the available

clinical trials concluded, "All of the trials were significantly underpowered to detect differences, resulting in inconclusive findings" (Nelson & Devanand, 2011, p. 577).

There are well-controlled and well-powered illustrations that also show the challenge of demonstrating differences between viable psychosocial treatments. For example, a recent multisite randomized controlled trial (RCT; $N = 341$) compared the effectiveness of cognitive therapy with psychodynamic–supportive therapy for the treatment of major depression. The treatments were equally effective in patient recovery from depression (Driessen et al., 2013). It might well be that the two treatments reflect different ways of altering the clinical problem but the absence of differences might also mean that they both were highly credible treatments from the perspective of the clients and generated strong expectancies for change. Expectancies for change is a parsimonious if not more parsimonious an explanation by suggesting the two treatments produced similar outcomes through a similar process rather than through two different treatment-specific processes.

The overall conclusion one can reach is that, by and large, the role of expectancies as an explanation for the effectiveness of EBPs is not easily ruled out. This is important because we believe we have a few hundred EBPs as separate techniques or variations. Do these interventions operate through multiple processes in special ways or are they multiple ways of mobilizing one or two processes (e.g., expectancies for change)? This is critical to know if we tell clinicians to administer these components of treatment in quite specific ways that may or may not be necessary. Perhaps, we can bolster the efficacy of TAUs, which often compete with EBPs in outcomes they achieve, by mobilizing credibility and expectancies for change at the beginning of treatment.

Measurement selection and interpretation

Another set of methodological issues that influence the conclusions about EBPs pertains to the outcome measures and how they are reported. Two assessment issues introduce ambiguities in what can be said about EBPs and their impact.

Selective analyses and reporting of measures. In any investigation and its write-up for publication, multiple decisions are made as to how the data will be analyzed and reported. These include when to stop running subjects if early peeks at the data are made; what measures to emphasize or even include in the write-up; what subjects, if any, are labeled as outliers and deleted; which of multiple data analyses to include; and many others (Kazdin, in press b). There are no hard and fast rules in reaching the decisions, but it is clear that the decisions and their reporting can give a slanted, incomplete, and distorted view of findings (John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). These concerns have emerged in scientific research more generally. Yet, one practice in particular has emerged in the context of EBPs and pertains to the measures that are used and reported.

Typically, EBP studies use multiple outcome measures. The measures may focus on symptoms, stress, adaptive functioning depending on the clinical problem of course, and include multiple methods of assessment (e.g., self- or other-report questionnaires, interviews, psychophysiological measures) as relevant to the clinical domain. On completion of the study, it would be useful to see the results for all of the measures, examine whether the treatment and control conditions were different on each of these outcomes, and then reach conclusions based on this overall picture. Yet, in reporting the results, the conclusions are not uniformly based on the full set of measures. In fact, there appear to be inconsistencies in reporting of measures. These inconsistencies favor and are in the direction of reaching the conclusion that the EBP was more effective than the comparison or control condition.

First, there is a problem in reporting within individual studies. Thus, when multiple measures are used, not all of the measures may show differences favoring the intervention versus the control or comparison group. The inconsistencies occur with measures that might be delineated as the primary focus (e.g., target symptoms or disorder) as well as among measures that are more ancillary (e.g., other symptoms, functioning in other domains; see, for example, De Los Reyes & Kazdin, 2006). Those measures that show significant differences are used as the basis to draw conclusions that the treatment is effective. Yet, in a given study, the conclusions about treatment effectiveness can be shown to be different if all of the measures were used, analyzed, and reported (De Los Reyes & Kazdin, 2008).

Second, and related, the conclusions about the evidence base for treatment are compounded when one examines different studies of the same or closely related variations of a treatment as applied to a given clinical problem. That is, the measures that are selected to support the treatment may vary across studies. This means that replications of treatment are not quite replications if the reporting is selective and varied outcomes are counted among different studies and interpreted as if they were consistent. We know generally that there is a publication bias in research, namely, that results showing group differences are more likely to be published than results showing no group differences (e.g., Dwan et al., 2008; Joober, Schmitz, Annable, & Boksa, 2012; Spielmans & Kirsch, 2014). Understandably, studies of EBPs underscore and emphasize the differences that have emerged and favor the conclusion that treatment is effective. But it is often the case that no effect or mixed effects could be drawn from the same study based on what was emphasized and reported.

The problem of basing conclusions on the measures that show an effect while ignoring, minimizing, or not reporting other measures that do not has a partial remedy. When clinical trials compare multiple interventions or an intervention against a control group, some funding agencies (e.g., National Institutes of Health [NIH]), some organizations (e.g., World Health Organization), and a large consortium of journal editors (the International Committee of Medical Journal Editors) require individuals to register their clinical trials in advance of conducting the study. Investigators complete information to convey exactly what the outcome measures are, what the primary measures will be, and how the measures will be examined (see De Angelis et al., 2005; Laine, Goodman, Griswold, & Sox, 2007). For example, the web site, ClinicalTrials.gov as part of NIH in the United States is the largest clinical trials database and as at this writing over 160,000 studies have been registered; they encompass all 50 states in the United States and 185 countries (http://clinicaltrials.gov/). This material is available in the public domain and allows all to see what was specified as the main measures from which conclusions will be drawn. In principle, this is excellent as a strategy to avoid reporting results on measures based on whether they showed the expected results. Also, this fosters dialogue when investigators challenge why outcomes that are primary are different in the published report from what was noted in advance when the study was registered (e.g., Grant & Beck, 2013). Yet, it is important to bear in mind that the vast majority of studies of EBPs probably are not funded; of those that are funded, they may not be funded by agencies that require specification of primary, secondary, and all outcome measures in advance of analyzing the data. Also, the development of EBPs has proceeded long before the reporting and registering requirements were put into place.

In short, the evidence on behalf of a given technique is often drawn from selective reporting of measures within a study and then different measures or constellations of measures for a given technique as applied to a particular clinical domain across several studies. The bias favors saying that treatment is effective and better than some control condition. At this point, there is no agreed-on solution that goes beyond the remedies noted above. The assessment challenge is clear; not all measures are equally important, psychometrically sound, or sensitive to change. Moreover, we can

expect many of the measures not to yield the same verdict at a given point in time (posttreatment) or from one point to another (posttreatment and follow-up) even among measures that are psychometrically sound and sensitive to change (Achenbach, Krukowski, Dumenci, & Ivanova, 2005; De Los Reyes, Thomas, Goodman, & Kundey, 2013). At present, there are no guidelines for selecting among measures and analyses, nor are there any consistent reporting standards. In relation to EBPs, the implication is that for some unspecifiable proportion of studies, the effectiveness of treatment would not be evident if all the measures were used or if those that did not show change were given equal attention as the attention accorded to the measures that did.

Arbitrary metrics. An assessment issue infrequently raised is the notion of *arbitrary metric* (Blanton & Jaccard, 2006). A measure is an arbitrary metric if we have no firm basis in knowing how that measure represents the construct of interest and where a person stands on that construct outside of the score on the measure. That is, there is an ambiguous and unclear referent. For example, consider the Beck Depression Inventory and the Hamilton Rating Scale for Depression—two well-studied measures of depression. Much is known about samples and how they score on these measures and various correlates associated with performance. Moreover, psychometric properties have been evaluated with many samples so we have manifold demonstrations of the various types of reliability and validity. The notion of an arbitrary metric does not challenge psychometric properties obtained with these measures. Yet, the problem of arbitrary metrics is not resolved by those properties.

In the case of treatment research, the concern is that we do not know how the individual is functioning (e.g., in relation to depression and impairment) in everyday life based on scores on the scales designed to measure these constructs. The questionnaire or interview may show change and individuals may attain a specific criterion score. Yet, we do not know precisely how the change or score translates to functioning in everyday life. The matter is more easily conveyed by looking at outcome measures that are not arbitrary metrics, as is often the case, for example, in studies of diabetes control, substance use, or obesity. Each of these has measures (e.g., blood level of index, weight) that reflect or map on to functioning and the target construct of interest. Of course, we could have a self-report or clinical interview asking about diabetes control, substance use, or weight and these report measures might well be highly correlated with the direct indices, but that is not the same as knowing precisely how scores on the questionnaires map on to actual functioning.

In research on EBPs, often we do not know that change has occurred beyond that measured by the arbitrary outcome metrics or whether individuals who show change on the measures have changed in everyday life or have changed in a way that makes a difference. Most EBPs rely heavily on arbitrary metrics. Consequently, it is important to keep in mind that it is possible that patients as a group or individually may not have changed on the underlying dimension in an important way or at least we do not know the extent of change beyond scores on the measure.

It may be useful to consider broadening the types of measures we use to better try to link therapeutic changes to real-world referents. For example, qualitative research is designed to study individuals intensively and to obtain descriptions of experience that are detailed, rich, and "thick," as they are sometimes called (Miles, Huberman, & Saldaña, 2014). Such measures by design often focus on the lives of individuals and in concrete as well as thematic ways. Qualitative measures may assess the real-world referents and changes and how any benefits of treatment are reflected in everyday experiences and individual functioning. Such measures already have been advocated in relation to evaluation of psychotherapy outcome (e.g., Hill, Chui, & Baumann, 2013; Lutz & Hill, 2009). In the very few studies that have used qualitative assessment, it is clear that dimensions that clients report as important at the end of treatment (e.g., gaining more control in their lives, being

able to care for others more, being more open emotionally) often depart from the usual outcomes of symptom change (e.g., Morris, 2005; Nilsson, Svensson, Sandell, & Clinton, 2007).

In the context of treatment evaluation, knowing whether clients genuinely improved beyond what was shown by the arbitrary metrics would be extremely valuable. Qualitative measures are not necessarily *the* solution or the only solution. I mention this as one avenue, but the overall problem is the point to note. Often, it is difficult to discern from standard measurement procedures how or whether EBPs have had palpable impact on daily life.

General comments. As researchers and consumers of research, perhaps we make the following measurement assumptions: (1) All measures that are used in a given study will be included in the write-up; (2) once measures are included in the write-up, they will be integrated into the conclusions about the impact of treatment; and (3) change on measures of interest map on to changes in functioning of individuals in daily life. The first two concerns can lead to biased reporting of findings where "statistically significant" differences are more likely to be published and may promote neglect of other measures within a study either by not including them or by ignoring their implications for the conclusions.

There are no formal guidelines on how to address biases in reporting or how to integrate and prioritize measures. Multiple solutions can address the matter. First, one solution is to compute effect sizes across studies as in meta-analyses to permit combining measures within and across studies. Yet, the computations cannot be done if the selected measures are not reported in the original investigations. Also, combining measures may not give weight to primary outcomes or dilute or inflate effects by summing across measures that were not otherwise meant to be combined. Second and already mentioned, some funding agencies require investigators to identify publicly primary measures in advance of running the study. Conclusions then could be based on those measures. Yet studies can have multiple primary measures and that raises the challenge of how to draw conclusions when some of those measures change and others do not. Third, novel ways of combining measures and having preset decision rules have been proposed (e.g., De Los Reyes & Kazdin, 2006). This would begin by recognizing the multidimensional nature of therapeutic change and integrating all of the measures to decide the extent to which a treatment is evidence based or meets criteria in varying degrees and strength of the evidence. Yet, multidimensional conclusions are not likely to give clear verdicts about a given treatment, which can be easily communicated. Finally, some journals are changing in ways that may reduce selective reporting of data and analyses (e.g., *Archives of Scientific Psychology*, Cooper & VandenBos, 2013; *Psychological Science*, Eich, 2014). Authors are required to provide more explicit information about the design of the study and then submit data on which the findings were based. In the treatment literature, at present, there is no resolution of the matter in drawing conclusions reached about a treatment in a given study or across studies.

Data evaluation and criteria

Conclusions about whether a treatment is effective are based on three broad criteria: statistical significance, magnitude of the relation (e.g., effect size), and clinical significance. Statistical significance as a way of drawing conclusions is very familiar and there is no unique problem that is associated with drawing inferences about EBPs that is not applicable to null hypothesis statistical significance testing more generally. Two other indices deserve more attention in part because they are confused and because even when distinguished they have important implications regarding what can be said about a treatment designated as evidence based.

Magnitude of the strength of relation. There are two broad indices that measure the strength of the relation of the independent (treatment) and dependent (outcome) variables. First, are measures of effect size (ES) that focus on the mean differences between groups and these are represented by Cohen's *d*. Second, there are measures that focus on shared variance or overlap of the variables and *r* and r^2 represent these.[2] I will use ES as the generic term and Cohen's *d* here because it is the most widely used index of ES, although my comments about treatment do not depend on which of the many options is used. In the case of Cohen's *d*, mean differences between groups is divided by the pooled standard deviation. This reflects the amount of spread or separation of the means of the groups (e.g., treatment vs control) in standard deviation units. To aid the interpretation of research, Cohen (1988) gave as a guideline for *d* that .2, .5, and .8 are small, medium, and large ESs, respectively.

ES is a statistical concept and derivation and has no necessary relation to the extent to which treatment has impact in ways that actually have any practical value to the clients. Yet, magnitude of effect often is confused with practical or clinical significance (e.g., Rutledge & Loh, 2004; Sun, Pan, & Wang, 2010; Weisz et al., 2013). A large ES does not translate to clinically significant change or impact. Consider briefly an example in which a large ES may have no practical or clinical value and a small ES which in fact does.

For example, consider the results of a hypothetical study for the treatment of obesity, as I have noted elsewhere (Kazdin, 2013). We might recruit individuals 100 pounds (~45 kg) overweight and twice or more their ideal weight (criteria sometimes used to define "morbid obesity") and randomly assign them to treatment or no-treatment conditions. All cases are followed up for 1 year after treatment ends. Let us assume that everyone in the intervention group loses 2 pounds (.91 kilograms) and everyone in the control group gains 2 pounds. At the end of the study, all participants may still be very obese. ES for this result might be very large in part depending on the consistencies of the changes and variability (pooled variance) of the groups. Yet ES does not convey whether the weight and health status have actually improved for anyone in palpable ways. We do not know from the ES, but a 2-pound loss is not very likely to qualify because the risk for serious health outcomes (e.g., diabetes, osteoarthritis, heart disease, or cancer) is unlikely to be affected by such a small change in weight. In other words, magnitude of effect as a statistical derivation has no necessary connection to clinical impact or practical significance.

The previous example was one in which ES could be large but the practical significance or clinical utility might be trivial. The "other way" shows the problem too, that is, a small correlation as a measure of a relation can reflect an important practical effect. For example, individuals who experience clinical depression after a heart attack are four times more likely to die in the following 6 months than those without depression (see Rutledge & Loh, 2004). Yet, using ES equivalent yields a small correlation between depression ($r = .22$) and heart attack and only a small amount of shared variance ($r^2 = .048$ or 4.8%). And the utility of aspirin in preventing a second heart attack has an important practical effect. Indeed, one clinical trial comparing aspirin and placebo was stopped early because the results were so clear in favor of aspirin, but the strength of the relationship was small ($r = .03$; see Rutledge & Loh, 2004).

As the examples convey, drawing conclusions about the therapeutic value of an EBP on client functioning does not follow from measures of the magnitude of the relation and treatment outcome. This does not gainsay the importance of ES and related measures as supplements to statistical significance, which is a broader methodological issue. Yet, ES does not translate to practical impact, that is, how clients are doing in everyday life with respect to domains focused on during treatment (e.g., symptoms, impairment, or adaptive functioning). Also, when ESs are larger for one type of treatment rather than another, it also follows that the difference in ES does not necessarily translate to a practical difference. For example, a recent meta-analysis of 52 studies comparing an EBP with

TAU found that EBPs surpassed TAU with a mean ES difference of .29 (Weisz et al., 2013). The difference in ES does not necessarily translate into ways that make any practical difference; it could, but ES is not the way to find out.

Clinical significance. Indices of clinical significance of therapeutic change have been developed to redress limits of statistical significance and ES in relation to the importance of the change. Clinical significance refers to the practical value or importance of the effect of an intervention, that is, whether it makes any "real" difference to the clients or to others in their functioning and everyday life.[3] When clinical significance is evaluated, researchers have relied primarily on one of three indices, noted only briefly here (see Kazdin, in press b).

First, clinical significance has been defined by showing that patient functioning (e.g., symptoms) *falls within normative levels of functioning* once treatment is completed. The criterion begins by showing that the patients were well outside of a normative range of functioning within the symptom domain (e.g., depression, anxiety) before treatment began. Yet, after treatment, the patients' symptoms fell within the range of a similar (e.g., age, sex) community sample functioning well in everyday life. This criterion requires measures for which normative data and ideally information about the range or cutoff points that discriminate clinical and nonclinic samples in relation to some other criterion. Alternatively, the range can be defined in standard deviation units (e.g., ±1 *SD* above the mean of the normative sample). There are nuances and research issues related to the normative levels of functioning (e.g., what is the appropriate peer group for comparison; improvement can be important without approaching any normative levels, as might be the case for patients with autism spectrum disorder or schizophrenia).

Second, *magnitude of change the clients make from pre- to posttreatment* often defines whether a change is clinically significant and can be used whether or not there is a normative group for comparison. This method, referred to as the Reliable Change Index, is calculated separately for each individual (e.g., Jacobson, Roberts, Berns, & McGlinchey, 1999). The individual's posttreatment score on a measure is subtracted from the pretreatment score. The goal of this subtraction is to measure improvement; hence, which score is subtracted from which is based on the direction of scoring. The difference score is divided by the standard error term based on the sample in the study. An improvement greater than 1.96 is considered to be clinically significant. This number (1.96) is in standard deviation units and is the commonly used criterion when a statistical test (*t* test) compares two groups. There are many variations of the change index that differ in the amount of change required and how the change is computed (e.g., Ogles, 2013; Speer & Greenbaum, 1995). In each case, the change measure is reasonable but there is no clear basis for identifying whether the change in fact translates to functioning in everyday life. A large change in standard deviation units, especially on arbitrary metric measures, does not necessarily map on to clear referents in the clients' everyday functioning.

Third, no longer *meeting psychiatric diagnostic criteria for a disorder* is another index used to measure of clinical significance. In many treatment studies, individuals are recruited and screened on the basis of whether they meet criteria for a psychiatric diagnosis (e.g., major depression, posttraumatic stress disorder). Those with a diagnosis are included in the study and assigned to various treatment and control conditions. A measure of clinical significance is to determine at the end of treatment whether individuals continue to meet criteria for the original (or other) diagnoses. Presumably, if treatment has achieved a sufficient change, the individual no longer meets criteria for the diagnosis.

As more disorders are recognized to be on a spectrum, it is clear that the cutoff for meeting and not meeting the criteria for a disorder is arbitrary. Patients who met criteria for the disorder at pretreatment and then who miss the cutoff at posttreatment cannot be inferred to be that much better

off. One can no longer meet diagnostic criteria on the basis of a small change on one or more indices (e.g., one symptom below severity, duration of the symptom just below in the duration criterion) and still be impaired. Even so, at first blush, this criterion has intuitive appeal, especially when phrased as follows: a person had a condition before treatment but no longer has that condition after treatment. But few psychiatric diagnoses may be categorical, and taking a multidimensional condition and making it categorical does not aid in evaluating the clinical impact of treatment.

Other indices. Although these are the main criteria used to evaluate clinical significance, there are many others that might be used including indices of impairment, complete elimination of maladaptive symptoms (e.g., panic attacks, tics), subjective evaluation of happiness, and quality of life (see Kazdin, in press b). One effort that has been applied to both physical and mental health is the notion of recovery where the outcomes are not so clear such as chronic medical disorders or outcomes of psychotherapy (e.g., for depression, anxiety).

Recovery has been defined in the context of mental disorders and addictions as "A process of change through which individuals improve their health and wellness, live a self-directed life, and strive to reach their full potential." (Substance Abuse and Mental Health Services Administration [SAMHSA], 2011, quote from web page, see full reference). The definition encompasses functioning in different spheres including health (living in a healthy way or managing one's condition), home (having a stable and safe place to live), purpose (e.g., having a purpose and being involved in meaningful activities), and community (e.g., having relationships with others and social networks with support, friendships, love, and hope). There are multiple facets of recovery as one can see and with that recovery is clearly dimensional rather than all or none and can vary by individual dimension (e.g., purpose, home; SAMHSA, 2011; Whitley & Drake, 2010).

The concept of recovery and its evaluation would be a useful addition for evaluating EBPs and might be applicable across many treatments and clinical problems. There might be adaptations needed as a function of age (children, elderly) and even a modular measure that could have sections or domains of recovery that are especially applicable. Recovery might be useful as well in moving to outcomes that can be said to be "clinical significant" because they address real-world daily functioning, depending on how recovery is measured. Of course, we would need to know that these measures not only assess the key concepts of recovery but also map on to performance in daily life, that is, they are not arbitrary metrics.

General comments. EBPs by definition are superior to controls, including no-treatment or other control groups. "Superior" is the key concern here. Statistical significance and ES are the usual indices. These are important scientifically of course, but as a rule tell us little about the impact of treatment on people and their everyday lives. Sometimes change on measures might be large and qualitative (e.g., no further nonsuicidal self-injury behavior, no panic attacks). Also, sometimes changes are evident on outcomes that reflect daily performance (e.g., placement of children with autism spectrum disorder into regular grade level classrooms, reduced rates of rehospitalization for patients with a diagnosis of schizophrenia, reduced recidivism for sexual offenders). Thus, my comments have important exceptions. Even so, the net effect is that for most EBPs and most outcomes (e.g., treatment of anxiety, depression), we do not really know the benefits and how they translate into patient experience and functioning.

In light of these considerations, it is quite possible to have EBPs that provide impressive outcome results in terms of statistical significance and ES, but are unclear on whether they have impact on client functioning in practical ways. As an example, in a study of depression I mentioned previously, two treatments (cognitive therapy with psychodynamic–supportive therapy) were

effective in terms of statistical significance (Driessen et al., 2013). A measure of clinical significance provides a perspective on interpreting outcome. Although the two treatments were equally effective, arguably they were not very effective in terms of clinical significance. Only 21% to 24% of the patients met criteria for recovery, indicating that more, different, or better treatments for depression are still needed. The study conveys that treatment can be effective (statistical change, ES) without having much impact on criteria related to functioning in everyday life for most people who received the treatment.

How do EBPs lead to change and for whom?

I mentioned the evaluation of EBPs would be from two perspectives: methodological and substantive. The distinction is useful but blended. The methodological issues discussed previously (e.g., control groups) very much address substantive questions about EBPs, what they entail, what their effects are, and so on. Yet, limited progress on core substantive questions that guide treatment research further hampers the impact of treatment and the benefits EPBs could have on clinical care.

Moderators

Moderator refers to some characteristic that influences the direction or magnitude of the relation between the intervention and outcome. If treatment outcome varies as a function of characteristics of the client or therapist (e.g., sex, culture, ethnicity, biomarkers) or treatment delivery (e.g., individual vs group treatment; brief vs extended), these characteristics are moderators. Historically, the dominant question guiding psychotherapy has been about moderators, as illustrated by "What treatment, by whom, is most effective for this individual with that specific problem, under which set of circumstances?" (Paul, 1967, p. 111). The question continues to dominate the treatment research agenda (e.g., DeRubeis et al., 2014; Kaplan, 2008; Roth & Fonagy, 2005).

After decades of study, psychotherapy literature is saturated with research findings on moderators that encompass characteristics of the clients, the therapists, treatment, and processes that emerge during treatment (Lambert, 2013). Even so, two critical issues remain. First, we do not know how to translate research on moderators into clinically useful information. This means that we cannot usually use the information to direct patients to treatments from which they are likely to profit and away from those from which they are less likely to profit. Among the issues, when a variable serves as a moderator, does that mean the variable is a moderator for the one treatment in which it was studied or for most or all treatments? This means we cannot use the information to guide decision making regarding treatment options. Also, just because a variable moderates treatment outcome, that does not mean those with poor standing on the moderator will not profit from treatment. Patients who have the untoward standing on a moderator can still respond well (e.g., statistical significance, ES) even though less well than those with favorable standing on that moderator (e.g., Kazdin & Whitley, 2006). In general, how to translate moderator findings into useful information remains to be exploited.

Second, we rarely understand why and how a given variable serves as a moderator. For example, socioeconomic disadvantage, stress, and severity of conduct problems moderate the effects of treatment for children with a diagnosis of conduct disorder (e.g., Kazdin, 2010; Leijten, Raaijmakers, de Castro, & Matthys, 2013). Individuals with greater socioeconomic disadvantage, high family stress, and severity of dysfunction respond less well to parent management training, a well-established EBP. However, the treatment–moderator findings typically do not help in relation to theory, research, or practice. Without knowing why the moderator has an effect there is not much we can do to controvert the process or means through which the moderator exerts its influence. In the

absence of further explanation, moderator research to date has not advanced our understanding of the underpinnings of treatment effects, nor has it been especially useful for decision making when selecting among treatments for clients.

Two lines of moderator research may point the way to future advances. First, efforts to understand moderators and how they relate to fundamental features of information processing may hold special promise. For example, biological features of how patients process information (e.g., neuromarkers) and psychological characteristics with which they are associated (e.g., processing social cues, working memory, emotion regulation) serve as moderators of treatment response (psychotherapy, medication; for example, Doehrmann et al., 2013; Furey et al., 2013). These moderators can relate to fundamental biological and psychological features of disorders and perhaps elucidate targets for intervention. Such work also has the potential of elaborating unifying processes across diverse treatments (Harvey et al., 2014).

Second, the ways in which moderators are studied are likely to change. Moderators (e.g., severity of dysfunction, comorbidity, cultural identity) are usually studied one at a time in any individual study. Individual moderators tend to be weak in how they predict outcome (e.g., ESs) and may not emerge as statistically significant. Recently, methods for integrating and combining multiple moderators have been elaborated (Kraemer, 2013). When multiple moderators are combined, moderator × treatment interactions emerge that otherwise would not be evident (e.g., Frank et al., 2011; Wallace, Frank, & Kraemer, 2013).

Improvements in moderator research are greatly needed to advance our understanding of EBPs as well as to improve decision making regarding the appropriate care for individual patients. While an EBP might be the first line of attack for a given clinical problem, we ought to be able to use moderator research to identify for whom a given treatment is likely or not likely to work. Also, I mentioned already that TAU occasionally surpasses or is equal to the impact of EBPs. Moderator research might well identify who needs something special beyond the usual care—apparently not everyone, but we do not have a way of identifying those individuals.

Mediators and mechanisms of change

Notwithstanding the rich outcome literature, we have very little idea about how treatments "work" or what leads to change. Without understanding how therapeutic change is achieved, we are at a disadvantage in knowing how we might optimize change through activation of specific components or processes. Research on mediators and mechanisms underlying therapeutic change aims to elaborate how treatment works.

Mediator is a construct that shows a statistical relation between an intervention and outcome. This is an intervening construct that suggests processes about why change occurs or on which change depends. Mediation is evident when several conditions are met: (1) the intervention leads to change in outcome measures, (2) the intervention alters the proposed mediator, (3) the mediator is related to outcome, and (4) outcome effects are not evident or significantly (statistically) less evident if the mediator did not change. These relations convey that change was mediated (e.g., correlated with, depended on) by some construct. Even when these conditions are met, ambiguity can remain about the precise role of the mediator (Bullock, Green, & Ha, 2010; Kazdin, 2014). Among the sources of ambiguity is that the mediator may serve as a proxy for some other variable that was not assessed. Also, a mediator may not emerge because the timeline for changes in the mediator may vary among participants and not be detected on the one occasion in which the mediator is assessed. In general, a mediator may not, and usually is not intended to, explain precisely how the change comes about. For example, cognitions may be shown to mediate change in therapy. However, this does not explain precisely how the change came about (i.e., what are the intervening

steps between cognitive change and reduced stress or anxiety). *Mechanism* refers to a greater level of explanatory specificity than mediator and reflects the steps or processes through which therapy (or some independent variable) actually unfolds and produces the change. Mechanism explains how the intervention translates into events that lead to the outcome or precisely what was altered that led to symptom change.

Psychotherapy research has benefitted greatly from advances in cognitive, affective, and social neuroscience, as well as the development of noninvasive neuroimaging and other biological assays to examine the structure, function, and activities of the brain that are associated with clinical dysfunction (especially anxiety and depression) and changes over the course of psychotherapy (e.g., Frewen, Dozois, & Lanius, 2008; Porto et al., 2009; Quidé, Witteveen, El-Hage, Veltman, & Olff, 2012; Roffman, Marci, Glick, Dougherty, & Rauch, 2005). Much of this work, but certainly not all, is correlational when characterizing disorders or change processes. Yet, the strength of the research stems from converging lines of evidence that move closer to identifying mechanisms of action. For example, from the reviews cited previously, research has

- Identified neurological characteristics associated with specific disorders and subtypes;
- Evaluated change in neural processes in "regions of interest" in light of characteristics associated with specific disorders;
- Induced or provoked symptoms (e.g., sadness manipulations in healthy samples; trauma stimuli among patients with posttraumatic stress disorder) to demonstrate brain areas implicated in dysfunction;
- Demonstrated "normalization" of neurological structures, function, and activity after therapy is completed;
- Elucidated similarities and differences in specific brain processes altered by different interventions (e.g., medication, psychotherapy) for a given disorder (e.g., major depression); and
- Documented some similarities in what brain processes are altered by the same intervention (e.g., cognitive behavior therapy) as applied to different disorders (e.g., obsessive compulsive disorders, depression).

Neuroimaging studies capitalize on a set of tools to elaborate processes involved in therapeutic change. Additional methods have been used to study brain receptors, intracellular signaling, and target genes that reflect dysfunction and change over the course of treatment (e.g., Duman, Schlesinger, Russell, & Duman, 2008; Sen, Duman, & Sanacora, 2008). Capitalizing on novel methods of assessment may help identify processes that change over the course of effective intervention as well as suggest biomarkers that moderate treatment effects.

A clear research direction and priority will be elaboration of biological changes that underlie and are associated with changes in symptoms, behaviors, and subjective experience. There is a concern that drawing on advances in neural assessment methods unwittingly will contribute to the "biologizing" of psychotherapy, that is, reverting to biological and "reductionist" explanations. However, core psychological and biological processes increasingly are studied together (e.g., cognitive, social, and cultural neuroscience) to elaborate the connections, reciprocal relations, and conversions of experience and brain processes. Also, many memory and cognitive processes are likely to be involved both in clinical dysfunction and therapeutic change (Harvey et al., 2014).

General comments

Highlighting the key points, it appears reasonable to say or at least not too extreme to say that as a general rule,

- EBPs, by definition, are statistically superior to various control and comparison conditions and often they produce strong effects (statistically), but we are a little unsure whether these effects translate to genuine clinical benefits that affect the daily lives of people who are treated;
- We are not sure that many EBPs are superior to the control or comparison condition because, we know that on some occasions conclusions only draw on measures within the study that have yielded the predicted effects;
- We do not know for whom a particular EBP is suitable or unsuitable (in varying degrees) and in ways that will help decision making about treatment (moderator); and
- We do not know how and why most EBPs work (e.g., mechanisms of action).

There are critical qualifications to my comments in evaluating the EBP literature. First, there is no EBP Platonic Form, that is, some single construct, concept, or ideal that actually reflects all of the individual EBPs. I have tried to characterize the literature as a whole, but in the process there may be many exceptions where treatments are not likely to be encompassed by key points. For example, I mentioned that expectancies and credibility of treatment might well explain the effects of many EBPs. But this is not very plausible in many circumstances. For example, in one randomized controlled study, children with autism spectrum disorder received an EBP treatment for 2 years and improved (e.g., cognitive, motoric, and daily living skills) and no longer met diagnostic criteria for the disorder (Dawson et al., 2010). TAU controls actually declined. Expectancy for change and credibility of treatment and control conditions do not seem plausible. The plausibility is further decreased by the large bodies of background research on behavior change (e.g., human and nonhuman animal research on experimental and applied behavior analysis) that underlie many of the treatment procedures. But the point is well taken; not all of the arguments I have noted apply to all EBPs as applied to diverse clinical problems.

Second, while I believe it is defensible to note that we have little idea how the vast majority of EBPs effect change, here too glossing over exceptions would overstate the case. Basic nonhuman animal laboratory research on extinction and the biological and neurological underpinnings of fear, fear enhancement, and reduction has very much guided treatments based on graduated exposure for the treatment of anxiety (e.g., Davis, 2011). Also, in the case of depression, nonhuman animal models have elaborated key processes that underlie change and again serve as a likely basis or target for treatment (e.g., Duman et al., 2008). It would be too strong to state we are completely in the dark as to why and how EPBs lead to change, particularly for exposure-based treatment for anxiety. Yet, for many, and perhaps most of the 300+ treatments, the specific mechanisms through which they achieve change are not at all clear.

Overall, discussing EBPs as a broad class has the disadvantage of glossing over important exceptions. Even so, it is meaningful to look at broad class and common characteristics in relation to methodological practices (e.g., reliance on statistical significance and ES with little attention to clinical impact). Also, it is useful to reiterate that we do not understand for whom treatment is effective and why, and how treatment actually leads to change. Add to this the evidence that EBPs are not always more effective than TAUs and we have no idea under what circumstances they are. These points do not seem to be minor or nuanced limitations to what we are entitled to say about EBPs and their effects.

Discussion and conclusion

Identifying psychosocial treatments with evidence represents an enormous research advance. We know that for a treatment to be evidence based, it should surpass the impact of and be more

effective than various control and comparison conditions (e.g., no treatment, attention-placebo control, TAU). That is not trivial. We have had hundreds and hundreds of psychosocial interventions for decades if one begins with the formal development of psychotherapy and not revert back through the ages where arguably psychotherapeutic interventions (e.g., talk, persuasion, belief, meditative practices) have been used to alleviate mental and physical health problems. We are now at the first point in history where behavioral and social sciences have established a large set of treatments (a few hundred) with rigorous scientific evidence on their behalf. This accomplishment has to be savored as an evolutionary leap that allows us to consider what is needed for the next breakthrough.

This article began by considering what it means when we say we have an EBP or that a given treatment is more effective than some control or comparison condition. The research encompasses features that qualify or limit what we can say based on the comparison groups to which EBPs are compared (no-treatment but also TAU), criteria to establish treatments as effective (statistical significance and ES), issues in reporting and measurement (selective reporting of measures, arbitrary metrics), and very little evidence about how treatments actually affect functioning of patients in daily life (clinical significance). When we say that we have an EBP for this or that clinical problem, that statement may be true. My comments reflect an effort to clarify what that statement does and does not mean.

Also, when we note that we have EBPs, it important to convey as well that these treatments are not invariably better than TAU, as provided in a given clinical setting. They usually are more effective, at least in the published literature, but all permutations seem to be evident. That is, EBPs have been more or less effective than or equally effective as a TAU. This is disturbing to say the least because TAU is a moving target, and we have no finite set of specifiable treatments (as usual) that we can study or study as easily as we might study cognitive behavior therapy or graduated exposure. In addition, we cannot yet say that TAUs work equally well (or better) for this or that particular clinical problem or population but not in other contexts.

Critical research questions remain to further establish EBPs and augment their clinical utility. Two areas were highlighted. First, we still do not have a set of moderators that will help us direct patients to treatments from which they are likely to profit and away from others for which they are not. This is critical because one way of increasing the effectiveness of treatment in patient care is better triage, that is, placing clients in the treatments to which they are likely to respond. Moderators have been studied extensively, but as a rule we do not know how to translate that information for patient care and whether a given moderator is unique to a specific treatment or will affect responsiveness to many, most, or all treatments.

Second, the mechanisms through which therapies achieve change warrant much more attention. We can generate more EBPs and that is critical for many problems for which there are no such treatments. In addition, a priority is to explain how the effects are achieved among those EBPs we have. Understanding mechanisms and how they are modified might greatly expand the range of effective interventions and also unify scores and scores of treatments that are conceptualized as working in different ways. More work is needed on establishing the mechanisms of action of EPBs. What is the path and what are the steps from talk, cognitive exercises, practice, or other processes in the sessions to changes in symptoms?

It is appropriate, if not urgent, to go beyond the usual claim that "more research is needed." Arguably, very different research is needed that goes beyond piling up more treatments with clinical impact that is not quite clear. We would want to be assured that any given treatment or clinical application makes a palpable difference from treatments usually provided.

Novel directions in treatment research are already emerging and these build on but also depart considerably from current research. Two major directions include research on transdiagnosis and

transtreatment and on models of delivering treatment in ways that depart from the usual model (one-on-one individual therapy with a mental health professional) on which the vast majority of EBPs have been based. These directions are discussed in a companion article (Kazdin, in press a) and move in directions that have the potential of reducing the burden of mental illness.

Funding

This project received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. For present purposes, psychosocial interventions refer to procedures based on psychological methods that alter functioning by focusing on affect, cognition, and behavior. These are distinguished here from more biologically based interventions (e.g., diet, medication, brain stimulation). The interventions focus on the full range of psychiatric disorders but also on multiple domains that may not focus on a diagnosis (e.g., stress, bereavement) where there is impairment in daily functioning. Psychosocial interventions with evidence in their behalf have been designated many different terms including empirically supported treatments, empirically validated treatments, evidence-based treatments, evidence-based psychotherapies, evidence-based practice, and others (e.g., American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006; Chambless & Ollendick, 2001). For this article, I will use the term evidence-based psychotherapies to cover the full range of interventions that draw on psychological methods to effect change.
2. There are many other measures of ES (e.g., Hedges's *g*, Glass's Δ), strength of the relation as represented by (R and R^2) in regression analyses, and others such as omega^2 (ω^2), eta squared (η^2), epsilon^2 (ε^2), and phi (φ) that cover slightly different circumstances, and not all such measures easily fit in one of the two categories (e.g., Grissom & Kim, 2011).
3. Clinical significance is the term used to reflect such changes in the context of treatments for psychological dysfunction. Applied or practical significance are other terms and can be applied more broadly (e.g., education, counseling, industry, organizations) beyond psychotherapeutic interventions.

References

Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin*, *131*, 361–382.

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271–285.

Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy*, *38*, 281–304.

Baskin, T. W., Tierney, S. C., Minami, T., & Wampold, B. E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology*, *71*, 973–979.

Benedetti, F. (2009). *Placebo effects: Understanding the mechanisms in health and disease*. New York, NY: Oxford University Press.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*, 445–454.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanisms? Don't expect an easy answer. *Journal of Personality and Social Psychology*, *98*, 550–558.

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology*, *49*, 421–453.

Cavanagh, K., Strauss, C., Forder, L., & Jones, F. (2014). Can mindfulness and acceptance be learnt by self-help? A systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. *Clinical Psychology Review*, *34*, 118–129.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716.

Christopherson, E. R., & Vanscoyoc, S. (2013). *Treatments that work with children: Empirically supported strategies for managing childhood problems* (2nd ed.). Washington, DC: American Psychological Association.

Cohen, J. S. (1988). *Statistical power analysis in the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H., & VandenBos, G. R. (2013). Archives of Scientific Psychology: A new journal for a new era. *Archives of Scientific Psychology*, *1*(1), 1–6.

Davis, M. (2011). NMDA receptors and fear extinction: Implications for cognitive behavioral therapy. *Dialogues in Clinical Neuroscience*, *13*, 463–474.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., . . .Varley, J. (2010). Randomized controlled trial of an intervention for toddlers with Autism: The Early Start Denver Model. *Pediatrics*, *125*, e17–e23.

De Angelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., & Weyden, M. B. V. D. (2005). Is this clinical trial fully registered?—A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, *352*, 2436–2438.

De Los Reyes, A., & Kazdin, A. E. (2006). Conceptualizing changes in behavior in intervention research: The Range of Possible Changes Model. *Psychological Review*, *113*, 554–583.

De Los Reyes, A., & Kazdin, A. E. (2008). When the evidence says, "yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science*, *17*, 47–51.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–149.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations: A demonstration. *PloS One*, *9*(1), e83875.

Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., . . .Gabrieli, J. D. (2013). Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry*, *70*, 87–97.

Driessen, E., Van, H. L., Don, F. J., Peen, J., Kook, S., Westra, D., . . .Dekker, J. J. M. (2013). The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: A randomized clinical trial. *American Journal of Psychiatry*, *170*, 1041–1050.

Druss, B. G., Hwang, I., Petukhova, M., Sampson, N. A., Wang, P. S., & Kessler, R. C. (2009). Impairment in role functioning in mental and chronic medical disorders in the US: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry*, *14*, 728–737.

Duman, C. H., Schlesinger, L., Russell, D. S., & Duman, R. S. (2008). Voluntary exercise produces antidepressant and anxiolytic behavioral effects in mice. *Brain Research*, *1199*, 148–158.

Duncan, B. L., Miller, S. D., Wampold, B. E., & Hubble, M. A. (Eds.). (2010). *The heart and soul of change: Delivering what works in therapy* (2nd ed.). Washington, DC: American Psychological Association.

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., . . .Williamson, P. R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS One*, *3*, e3081.

Eich, E. P. R. (2014). Business not as usual. *Psychological Science*, *25*, 3–6.

Finniss, D. G., Kaptchuk, T. J., Miller, F., & Benedetti, F. (2010). Biological, clinical, and ethical advances of placebo effects. *Lancet*, *375*, 686–695.

Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A., . . .Forgione, R. N. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine*, *41*, 151–162.

Freedland, K. E., Mohr, D. C., Davidson, K. W., & Schwartz, J. E. (2011). Usual and unusual care: Existing practice control groups in randomized controlled trials of behavioral interventions. *Psychosomatic Medicine*, *73*, 323–335.

Frewen, P. A., Dozois, D. J. A., & Lanius, R. A. (2008). Neuroimaging studies of psychological interventions for mood and anxiety disorders: Empirical and methodological review. *Clinical Psychology Review*, *28*, 228–246.

Furey, M. L., Drevets, W. C., Hoffman, E. M., Frankel, E., Speer, A. M., & Zarate, C. A. (2013). Potential of pretreatment neural activity in the visual cortex during emotional processing to predict treatment response to scopolamine in major depressive disorder. *JAMA Psychiatry*, *70*, 280–290.

Grant, P. M., & Beck, A. T. (2013). Cognitive therapy in patients with schizophrenia—Reply. *JAMA Psychiatry*, *70*, 543–544.

Grissom, R. J. (1996). The magical number .7 +/− .2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology*, *64*, 973–982.

Grissom, R. J., & Kim, J. J. (2011). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York, NY: Taylor & Francis.

Harvey, A. G., Lee, J., Williams, J., Hollon, S. D., Walker, M. P., Thompson, M. A., & Smith, R. (2014). Improving outcome of psychosocial treatments by enhancing memory and learning. *Perspectives on Psychological Science*, *9*, 161–179.

Hill, C. E., Chui, H., & Baumann, E. (2013). Revisiting and re envisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. *Psychotherapy*, *50*, 68–76.

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. (1999). Methods for defining and determining the clinical significance of treatment effects in mental health research: Current status, new applications, and future directions. *Journal of Consulting and Clinical Psychology*, *67*, 300–307.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry & Neuroscience*, *37*, 149–152.

Kam-Hansen, S., Jakubowski, M., Kelley, J. M., Kirsch, I., Hoaglin, D. C., Kaptchuk, T. J., & Burstein, R. (2014). Altered placebo and drug labeling changes the outcome of episodic migraine attacks. *Science Translational Medicine*, *6*, 218ra5.

Kaplan, R. M. (2008). Special issue: Mediation and moderation. *Health Psychology*, *27*(2, Suppl.), S99–S100.

Kazantzis, N. (2000). Power to detect homework effects in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *68*, 166–170.

Kazdin, A. E. (2010). Problem-solving skills training and parent management training for Oppositional Defiant Disorder and Conduct Disorder. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 211–226). New York, NY: Guilford Press.

Kazdin, A. E. (2013). Evidence-based treatment and usual care: Cautions and qualifications. *JAMA Psychiatry*, *70*, 666–667.

Kazdin, A. E. (2014). Moderators, mediators, and mechanisms of change in psychotherapy. In W. Lutz & S. Knox (Eds.), *Quantitative and qualitative methods in psychotherapy* (pp. 87–101). New York, NY: Taylor & Francis.

Kazdin, A. E. (in press a) Evidence-based psychotherapies II: New directions to extend impact of treatment. *South African Journal of Psychology*.

Kazdin, A. E. (in press b) *Research design in clinical psychology* (5th ed.). Boston, MA: Pearson.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *57*, 138–147.

Kazdin, A. E., & Whitley, M. K. (2006). Comorbidity, case complexity, and effects of evidence-based treatment for children referred for disruptive behavior. *Journal of Consulting and Clinical Psychology*, *74*, 455–467.

Kessler, R., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., … & Wang, P. (2009). The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Surveys. *Epidemiologia e Psichiatria Sociale*, *18*, 23–33.

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., . . .Zaslavsky, A. M. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine*, *352*, 2515–2523.

Kessler, R. C., & Wang, P. S. (2008). The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annual Review of Public Health*, *29*, 115–129.

Kraemer, H. C. (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: A parametric approach. *Statistics in Medicine*, *32*, 1964–1973.

Laine, C., Goodman, S. N., Griswold, M. E., & Sox, H. C. (2007). Reproducible research: Moving toward research the public can really trust. *Annals of Internal Medicine*, *146*, 450–453.

Lambert, M. J. (Ed.). (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). New York, NY: John Wiley & Sons.

Lambert, M. J., & Ogles, B. M. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York, NY: John Wiley & Sons.

Leijten, P., Raaijmakers, M. A., de Castro, B. O., & Matthys, W. (2013). Does socioeconomic status matter? A meta-analysis on parent training effectiveness for disruptive child behavior. *Journal of Clinical Child and Adolescent Psychology*, *42*, 384–392.

Lutz, W., & Hill, C. E. (2009). Quantitative and qualitative methods for psychotherapy research: Introduction to special section. *Psychotherapy Research*, *19*, 369–373.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: SAGE.

Morris, B. (2005). *Discovering bits and pieces of me: Research exploring women's experiences of psychoanalytic psychotherapy*. London, England: Women's Therapy Centre.

Mulcahy, R., Reay, R. E., Wilkinson, R. B., & Owen, C. (2010). A randomised control trial for the effectiveness of group interpersonal psychotherapy for postnatal depression. *Archives of Women's Mental Health*, *13*, 125–139.

Nathan, P. E., & Gorman, J. M. (Eds.). (in press). *Treatments that work* (4th ed.). New York, NY: Oxford University Press.

Nelson, J. C., & Devanand, D. P. (2011). A systematic review and meta-analysis of placebo-controlled antidepressant studies in people with depression and dementia. *Journal of the American Geriatrics Society*, *59*, 577–585.

Nilsson, T., Svensson, M., Sandell, R., & Clinton, D. (2007). Patients' experiences of change in cognitive-behavioral therapy and psychodynamic therapy: A qualitative comparative study. *Psychotherapy Research*, *17*, 553–566.

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 134–167). Hoboken, NJ: John Wiley & Sons.

Paul, G. L. (1966). *Insight versus desensitization in psychotherapy: An experiment in anxiety reduction*. Stanford, CA: Stanford University Press.

Paul, G. L. (1967). Outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*, 109–118.

Porto, P. R., Oliveira, L., Mari, J., Volchan, E., Figueira, I., & Ventura, P. (2009). Does cognitive behavioral therapy change the brain? A systematic review of neuroimaging in anxiety disorders. *Journal of Neuropsychiatry and Clinical Neurosciences*, *21*, 114–125.

Quidé, Y., Witteveen, A. B., El-Hage, W., Veltman, D. J., & Olff, M. (2012). Differences between effects of psychological versus pharmacological treatments on functional and morphological brain alterations in anxiety disorders and major depressive disorder: A systematic review. *Neuroscience and Biobehavioral Reviews*, *36*, 626–644.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*, 329–342.

Roffman, J. L., Marci, C. M., Glick, D. M., Dougherty, D. D., & Rauch, S. L. (2005). Neuroimaging and the functional neuroanatomy of psychotherapy. *Psychological Medicine*, *35*, 1385–1398.

Roth, A., & Fonagy, P. (2005). *What works for whom: A critical review of psychotherapy research* (2nd ed.). New York, NY: Guilford Press.

Rutledge, T., & Loh, C. (2004). Effect sizes and statistical testing in the determination of clinical significance in behavioral medicine research. *Annals of Behavioral Medicine*, *27*, 138–145.

Sen, S., Duman, R., & Sanacora, G. (2008). Serum brain-derived neurotrophic factor, depression, and antidepressant medications: Meta-analyses and implications. *Biological Psychiatry*, *64*, 527–532.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, *63*, 1044–1048.

Spielmans, G. I., & Kirsch, I. (2014). Drug approval and drug effectiveness. *Annual Review of Clinical Psychology*, *10*, 741–766.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2011). *SAMHSA announces a working definition of "recovery" from mental disorders and substance abuse disorders*. Retrieved from www.samhsa.gov/newsroom/advisories/1112223420.aspx

Sun, S., Pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology*, *102*, 989–1004.

US Department of Health and Human Services (USDHHS). (2014). *Substance Abuse and Mental Health Services Administration (SAMHSA), National Registry of Evidence-based Programs and Practices (NREPP)*. Retrieved from http://www.nrepp.samhsa.gov/AboutNREPP.aspx

Wallace, M. L., Frank, E. J., & Kraemer, H. C. (2013). A novel approach for developing and interpreting moderator profiles in randomized controlled trials. *JAMA Psychiatry*, *70*, 1241–1247.

Wampold, B. E., Budge, S. L., Laska, K. M., Del Re, A. C., Baardseth, T. P., & Gunn, W. (2011). Evidence-based interventions for depression and anxiety versus intervention-as usual: A meta-analysis of direct comparisons. *Clinical Psychology Review*, *31*, 1304–1312.

Weisz, J. R., Jensen-Doss, A. J., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*, *61*, 671–689.

Weisz, J. R., & Kazdin, A. E. (Eds.). (2010). *Evidence-based psychotherapies for children and adolescents* (2nd ed.). New York, NY: Guilford Press.

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. J. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry*, *70*, 750–761.

Whitley, R., & Drake, R. (2010). Recovery: A dimensional approach. *Psychiatric Services*, *61*, 1248–1250.

World Federation for Mental Health. (2011). *The great push: Investing in mental health*. Occoquan, VA: Author.

World Health Organization (WHO). (2008). *The global burden of disease: 2004 update*. Geneva, Switzerland: Author.