Reliable and clinically significant change

Mounted by Chris Evans in 1998, last updated 25.v.05

Reliable change

Reliable change was a concept introduced by
Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984). "Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance." Behavior Therapy **15**: 336-352.

and modified after a correction first published by:
Christensen, L. & Mendoza, J. L. (1986). "A method of assessing change in a single subject: an alteration of the RC index." Behavior Therapy **17**: 305-308.

The best early summary of the method is:
Jacobson, N. S. & Truax, P. (1991). "Clinical significance: a statistical approach to defining meaningful change in psychotherapy research." Journal of Consulting and Clinical Psychology **59**(1): 12-19 and our own paper:
Evans, C., Margison, F. & Barkham, M. (1998) The contribution of reliable and clinically significant change methods to evidence-based mental health Evidence Based Mental Health **1**:70-72 is another readable introduction.

Reliable Change (RC) is about whether people changed sufficiently that the change is unlikely to be due to simple measurement unreliability. You determine who has changed reliably (i.e. more than the unreliability of the measure would suggest might happen for 95% of subjects) by seeing if the difference between the follow-up and initial scores is more than a certain level. That level is a function of the initial standard deviation of the measure and its reliability. If you only have a few observations it will be best to find some typical data reported for the same measure in a service as similar as possible to yours. The reliability parameter to use is up to you. Using Cronbach's alpha or another parameter of internal consistency is probably the most theoretically consistent approach since the theory behind this is classical reliability theory. By contrast a test-retest reliability measure always includes not only simple unreliability of the measure but also any real changes in whatever is being measured. This means that internal reliability is almost always higher than test-retest and will generally result in more people being seen to have changed reliably.

Thus using a test-retest reliability correlation introduces a sort of historical control, i.e. the number showing reliable change can be compared with 5% that would have been expected to show that much change over the retest interval *if there had been no intervention*.

I recommend using coefficient alpha determined in your own data but if you can't get that then I'd use published coefficient alpha values for the measure, preferably from a similar population.

The formula for the standard error of change is:

```
SD1*sqrt(2)*sqrt(1-rel)
```

where SD1 is the initial standard deviation
sqrt indicates the sqare root
rel indicates the reliability

The formula for criterion level, based on change that would happen less than 5% of the time by unreliability of measurement alone, is:

```
    1.96*SD1*sqrt(2)*sqrt(1-rel)
```

I've written a little Perl program to calculate this for you:

- the HTML form to use the program
- the Perl program itself if you want/need a copy

Clinically significant change

Clinically significant change was introduced in the same 1984 paper by Jacobson, Follette & Revenstorf. However, this is different, not about whether the change is greater than might be expected by simple measurement unreliability and solely about the state the person achieves. Clinically significant change is is change that has taken the person from a score typical of a problematic, dysfunctional, patient, client or user group to a score typical of the "normal" population. Jacobson, Follette & Revenstorf (1984) offer three different ways of working this out.

- Their method (A): has the person moved more than 2 SD from the mean for the "problem" group?
  i.e. crit_a = mean(patients) + 2*stdev(patients) (if the measure is a "health" measure i.e. higher scores, better state; crit_a = mean(patients) - 2*stdev(patients) (if the measure is a "dysfunction" or problem measure).
- Their method (B): has the person moved to within 2 SD of the mean for the "normal" population? i.e. crit_b = mean(normative data) - 2*stdev(normative data) (if the measure is a "health" measure i.e. higher scores, better state; crit_b = mean(normative data) + 2*stdev(normative data) (if the measure is a "dysfunction" or problem measure).
- Their method (C): has the person moved to the "normal" side of the point halfway between the above? i.e. crit_c = (crit_a + crit_b)/2

Their methods (A) and (B) are straightforward though there are questions about what referential data to use for the "normal" mean and s.d. and there is a question whether you should use your own data for the "problem" group (I believe you should, with an only mildly disturbed group this can make it difficult to show clinically significant change).

However, there is a final twist on their method (C) which is what you do if the s.d.s for the "problem" and the "normal" groups are not equal. They suggest you take the distance of the criterion from the "problem" and "normal" means in terms of the pertinent s.d.s, i.e.:
(crit_c - mean(patients))/stdev(patients) = (mean(normative data) - crit_c)/stdev(normative data) (if the measure is a "health" measure i.e. higher scores, better state)
this gives:
crit_c = (stdev(normative data)*mean(patients) + stdev(patients)*mean(normative data))/(stdev(normative data) + stdev(patients))
(this is the same whether the measure is positively, i.e. health, tuned, or negatively, i.e. problem, tuned).

This arithmetic is really trivial but I hate arithmetic so I've written a little Perl program to calculate this for you:

- the HTML form to use the program
- the Perl program itself if you want/need a copy

More interestingly, at least for those who want to see the picture of the cutting points, I've written an

R program that plots the two distributions and three cutting points:

- the HTML form to use the program
- the Perl program itself if you want/need a copy

Putting them together

When summarising results you are clearly particularly interested in any people who got reliably worse: all good services recognise they don't always succeed and this is a good criterion on which to select out cases for a some clinical review. Then you are interested in people who got reliably better but not clinically significantly so. This may be because movement into the "normal" range is unrealistic or because your clinic sees people who are not so different from "normal" that that change is easily achieved. Then you are similarly interested in those who got clinically significantly, but not reliably, better. This suggests they were near enough to the boundary between "problem" and "normal" groups to start with that the clinically significant improvement is unreliable (which may mean it likely to relapse). Finally, the people you are interested in most are those who showed both reliable *and*clinically significant improvement. Those who changed most are clearly the ones you might select for positive clinical case review.

Final caveat

Always remember that measures only measure part of the human condition and don't always do that well. Such methods should always be used in parallel with other ways of reviewing clinical work.

- other statistical info. at this site
- reliable and clinical change info.
- statistics for visits to these pages
- tools and measures page
- grids and PCP pages
- main PSYCTC.org site entry page\n