The Leeds Reliable Change Indicator

Simple Excel(tm) applications for the analysis of individual patient and group data

Stephen Morley and Clare Dowzer

University of Leeds

Cite as: Morley, S., Dowzer, C.N. (2014) Manual for the Leeds Reliable Change Indicator: Simple Excel(tm) applications for the analysis of individual patient and group data. University of Leeds, Leeds, UK.

Introduction

The Reliable Change Index/Clinically Significant Change (RCI/CSC) method was popularised by Neil Jacobson in articles appearing in the 1980's and early 1990's but the original statistical principles go back 80 or more years. Since Jacobson popularised the methods there have been several proposed alternatives but in our view the computations rarely make a hugely significant difference to the outcome of the analysis. So we have chosen the KISS principle- 'keep it simple, stupid'. The methods are designed to answer two questions

1. Is an individual's change reliable i.e., is the magnitude of the observed change more that can be explained by errors of measurement? The critical idea here is that all measurement is unreliable to some extent and that a difference between two scores from an individual could be due to the measurement error rather than any other reason e.g. Therapy.

2. Has the individual made big enough change during treatment for this to be regarded as important - is it clinically significant? Note that you can also use this method to ask whether a person has deteriorated: are they reliably worse off?

Figure 1 illustrates the possible outcomes for a single patient.

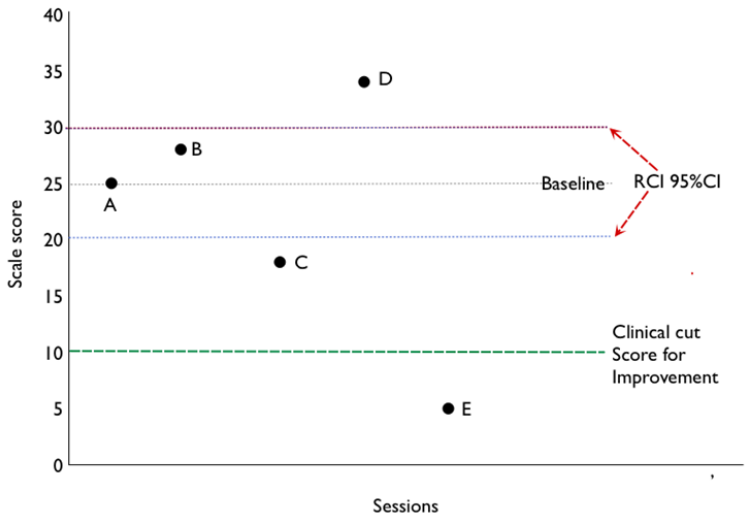


Figure 1: *Possible outcomes from a single patient over time.*

In this figure the Y axis is the score on your outcome measure. On this measure a reduction in the score represents an improvement. The X axis represents various time points. Each of the data points is labelled A to E. The data point A is at pre-treatment. If the person makes no change then at later time points their score would fall on the pale grey line denoting the baseline. The Figure also shows the confidence intervals for the RCI. The upper limit is red and the lower one blue. Data point B indicates that the person has not changed significantly i.e. their data falls within the red and blue lines. At data point C the person has moved

beyond the blue line (lower confidence interval) so they have made a reliable change (improvement). For data point D the person has moved beyond the red line so that is a reliable deterioration. Finally, we have inserted a green line to mark the cut-score representing a clinical improvement. Data point E is therefore not only a reliable improvement but also a clinically significant one.

Determining the criterion to use in your CSC analysis

There are two ways of doing this. You can select an *externally valid criterion* e.g., from a sensitivity/specificity analysis against gold-standard clinical judgment i.e. a diagnostic interview. For the example in this guide we have set a total distress score of below 12 as the cut score. If you don't have an externally determined cut score you can use one based on *statistical criteria.* Jacobson suggests three. They are illustrated in the Figure 2.

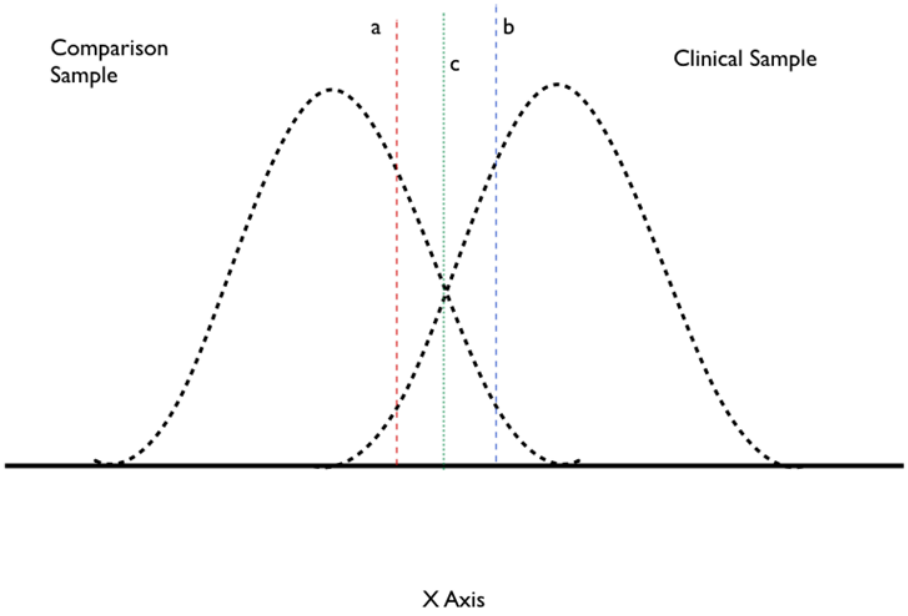


Figure 2: *A schematic representation of Jacobson's three statistical criteria for defining clinically significant change.*

In Figure 2 the distributions of scores of clinical and comparison (non-clinical) samples are shown. It is assumed that they are normally distributed and that they do not overlap completely. In this figure the comparison sample has a lower mean (the peak of the distribution) score than the clinical one, so improvement is a *reduction* or *decrease* in the score (as in the worked example given in the Excel™ spreadsheet). Jacobson suggests that there are three possible criteria (a, b, and c) shown by the lines representing the cut scores.

a. Redline The level of functioning after therapy should fall outside the range of the clinical population (more than 1.96 standard deviations, in the direction of the comparison reference group.

b. Blue line The level of functioning should fall within the range of the comparison non-clinical group, within 1.96 standard deviation of the mean of the comparison group.
c. Green line The level of functioning should place the patient closer to the mean of the comparison group than the mean of the clinical group.

Which criterion should you use? Jacobson suggests that when norms are available for a comparison group criteria **b** or **c** are preferable. It the scores from the groups overlap, as they do in Figure 2, then **c** is preferable. When they do not then **b** should be adopted. If norms for a comparison non-clinical group are not available then **a** is the only criterion available.

To compute these you need information on the distribution of scores (means and SDs) in the clinical population (criterion **a**) and the clinical and non-clinical populations for criteria **b** and **c**. You are asked to provide these data in the Excel™ workbook.

Please note that it's possible that unless you have used the cut score to select a person for treatment it is possible that they may score below the cut score at pre-treatment, so we ought to exclude them from the analysis. There is a bit of a counterbalance to this in that only patients who make a reliable change (improvement) can make a clinically significant change and in clinical experience it is relatively infrequent that those below the cut score at pre-treatment make a further improvement. We can also imagine that some people are below the cut score at pre-treatment but deteriorated during treatment, and we need to count them.

Doing the RCI/CSC analysis

To run the analysis you need to determine RCI/CSC you need the following data.

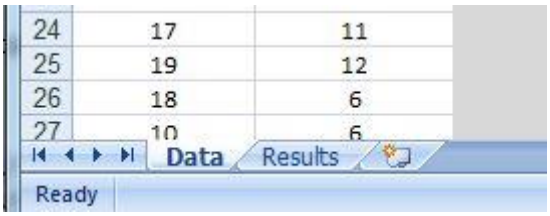
1. The pre-treatment and post treatment scores: If you use the group version it is important that there are ***no*** missing data.
2. Information in the reliability of the measure. You can obtain this from the test manual, or, if the measure is taken from the literature you will need to look at the articles and find details of the reliability. There is a statistic called Cronbach's alpha and this is the preferred measure. For many measures you will find various estimates of alpha in different sources. We suggest you compute the median value or take the estimate that was computed from a dataset as similar to yours as possible. There is a way of combining these data statistically and we are developing a separate guide and Excel™ worksheet for doing that.
3. A cut-score that you want to use to determine the criterion for clinically significant change. There are several options here and they are discussed in more detail in the section on clinically significant change.

Please note that the formulae in the Excel™ workbook have been designed so that in the results an improvement following treatment is always shown as a positive number. This

applies to the change score used to compute the RCI and CSC and also to the effect size used to benchmark your programme.

Data entry in the Excel workbook

The workbook has two sheets 'Data' and 'Results' that you can access on the tabs at the bottom.



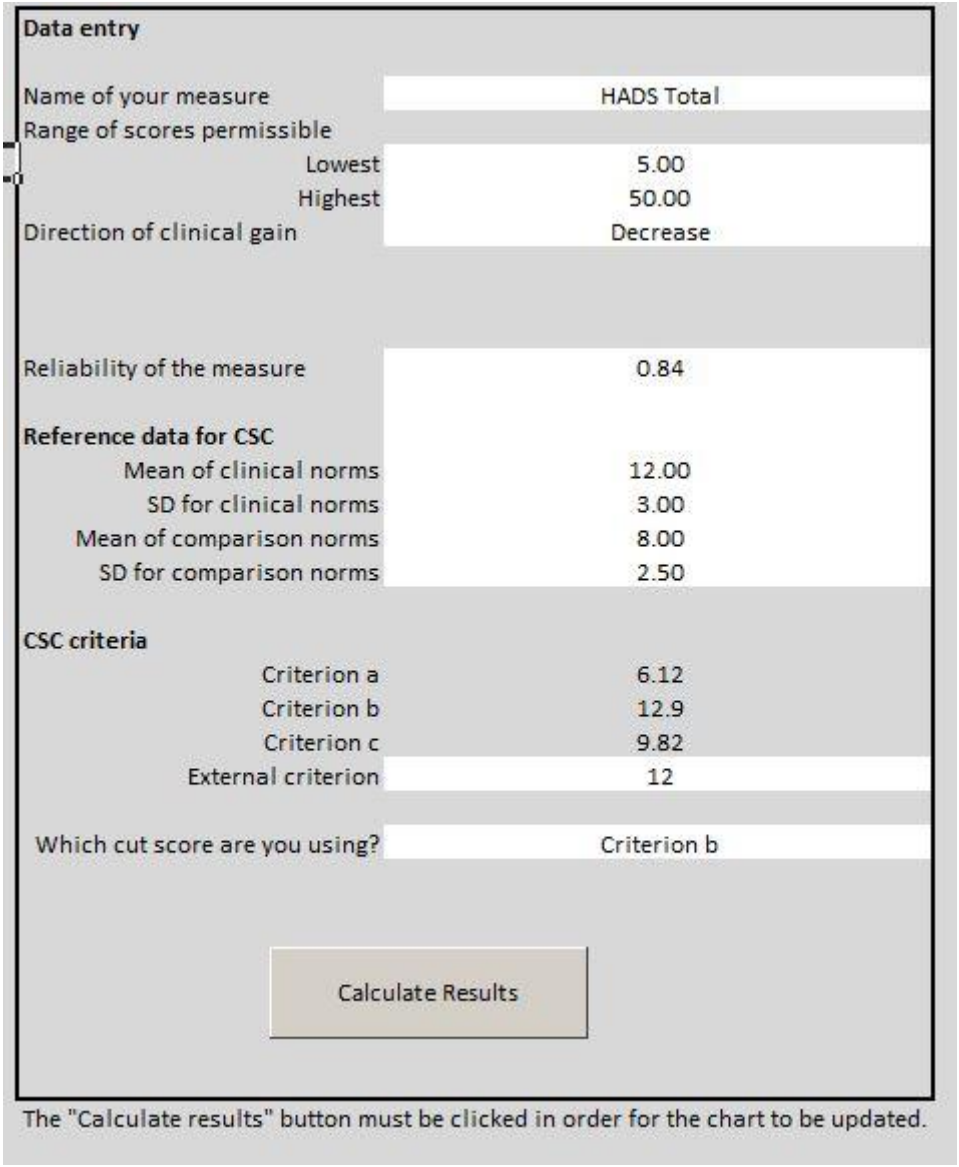
There are two sets of data you need to enter.

Individual data On the Data tab the white areas indicate where you can enter data. The grey areas in the sheet are locked to protect the formulae. The data we have entered are a random sample of 100 individuals drawn from a larger data set. These data are the pre-treatment and post treatment total scores on the Hospital Anxiety and Depression Scale (HADS): this is the sum of the anxiety and depression sub scales and it is a pretty good measure of overall distress. You can simply delete these data and insert your own.

	B	C	D	E	F
1	PreTreatment	PostTreatment	Change Score	Reliable Change	CSC
2	15	10	5	Nochange	
3	19	8	11	Improve	CSC
4	8	1	7	Nochange	
5	25	10	15	Improve	CSC
6	20	2	18	Improve	CSC
7	19	4	15	Improve	CSC
8	17	4	13	Improve	CSC

You will see the grey area includes three more columns headed 'Change Score', 'Reliable Change' and 'CSC' (clinically significant change). You cannot put data in these cells but they will be populated automatically as you enter the data and Excel™ computes the results. You will also note that the data in the Reliable Change and CSC columns are colour coded just to help you see the status of individuals.

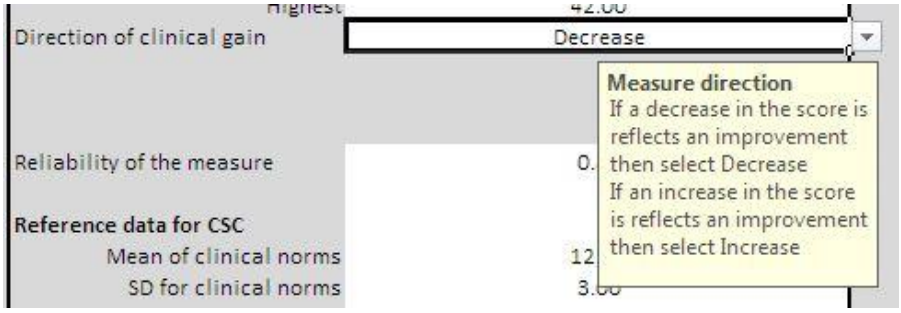
Data about the measure You will also find a box on the Data sheet where you must enter data about the measure you have used.



Name of measure: Enter the name of your measure. This is used on the Results sheet to label the analysis.

Range of scores permissible: Enter the lowest (usually 0) and highest values obtainable on your chosen measure.

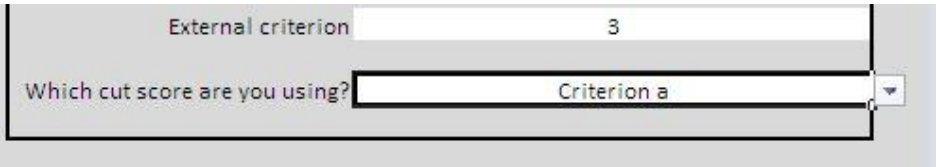
Direction of clinical gain: You have two options 'Decrease' and 'Increase' and when you put the cursor on the cell you will activate a drop down box. Enter Decrease if an improvement is shown by a decrease in the score and enter Increase if an improvement is indicated by an increase in the score



Reliability of the measure: Enter the reliability of the measure. It must lie between 0 and 1.0.

The next section is **Reference Data for CSC**. Here you need to enter available data from reference norms from both a clinical population and a comparison population. It asks for the mean and SD values for reference samples and uses these to generate three clinical cut points (a, b and c) – these were explained in the section Determining the criterion to use in your SCS analysis – and you can see these values in the greyed in cells below. If you do not have data from these populations then leave the cells blank. It just means that you will not be able to get analyses for various cut scores. You may also enter your own value for a cut point in the box labelled External criterion.

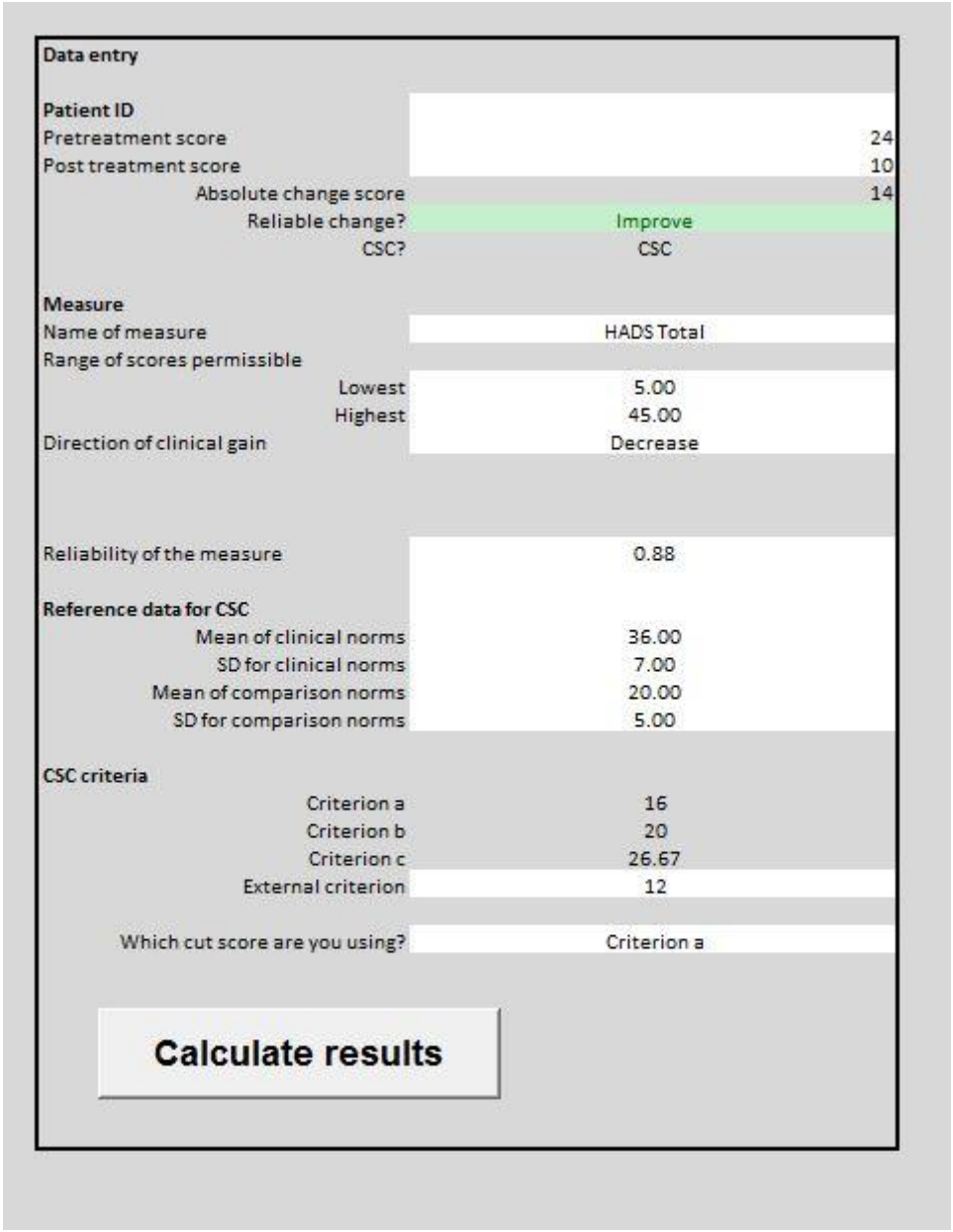
Finally, you are asked to specify which cut score you want to use for the analysis. Once again there is a drop down box and you can select criterion a, b, c or the External criterion option.



You can run and rerun the analyses using different criteria if you wish by simply changing the cut score. However it is good practice to decide beforehand which criterion you are going to use as your primary outcome.

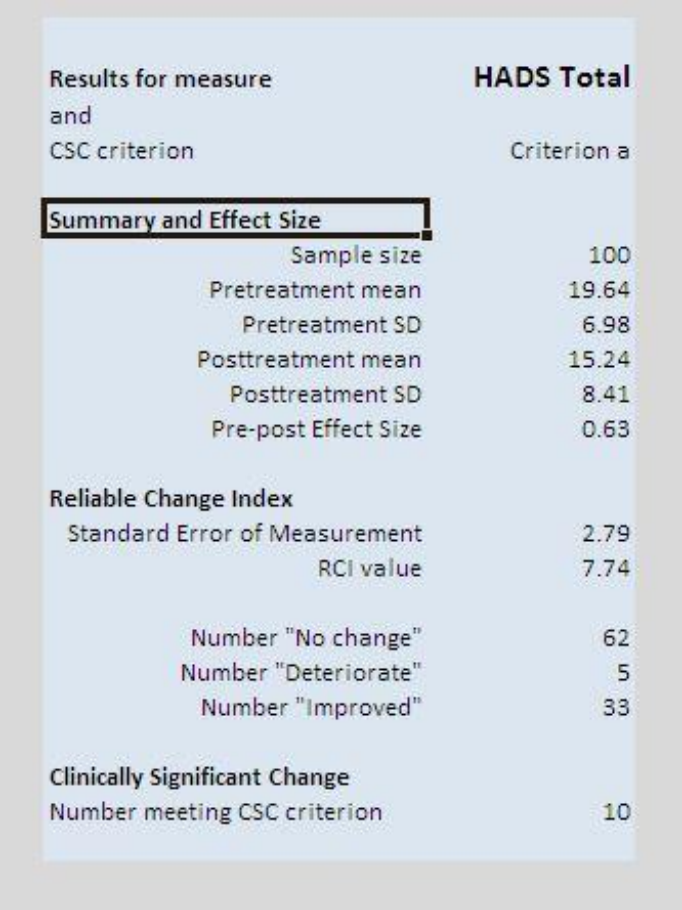
The single case version

In the single case version the data entry is all in one place. You can enter the patient's ID and post treatment score. One you have entered details about the measure you will get a colour coded output telling you if they have made a reliable change.



Analysis and Results

On the Results tab you will find a box giving the complete set of results for the measure and the criterion you have entered for the clinically significant change analysis.



Results for measure and CSC criterion	HADS Total Criterion a
Summary and Effect Size	
Sample size	100
Pretreatment mean	19.64
Pretreatment SD	6.98
Posttreatment mean	15.24
Posttreatment SD	8.41
Pre-post Effect Size	0.63
Reliable Change Index	
Standard Error of Measurement	2.79
RCI value	7.74
Number "No change"	62
Number "Deteriorate"	5
Number "Improved"	33
Clinically Significant Change	
Number meeting CSC criterion	10

Summary and Effect Size

The first set of results is the **Summary and Effect Size**. This includes a summary of your measure. It includes the total number (sample size) in your data set, the pre and post-treatment means and pre-treatment standard deviations and it also reports the Pre-post effect size which you can use if you are going to benchmark your service data. The effect size is not useful for an individual.

RCI analysis

For an individual to have made a reliable change (better or worse) then their change score must be larger than the RCI value. The box on the Results sheet shows the Standard Error of Measurement for your measure and the RCI value. The Data sheet shows the results for each individual in column E where they are colour coded: green = reliable change, yellow = no change and red = reliable deterioration. These data are aggregated in the Results sheet and if you look in the summary box you will see that 62 people made no change, 5 deteriorated and 34 made a reliable change i.e., around about 1/3rd – not bad.

Clinically Significant Change (CSC) analysis

Finally we need to work out whether the person has made a clinically significant change. In this example we have used criterion 'a' for the CSC analysis. This means that their post-treatment score must be less than 6.12. The analysis compares each individual change score

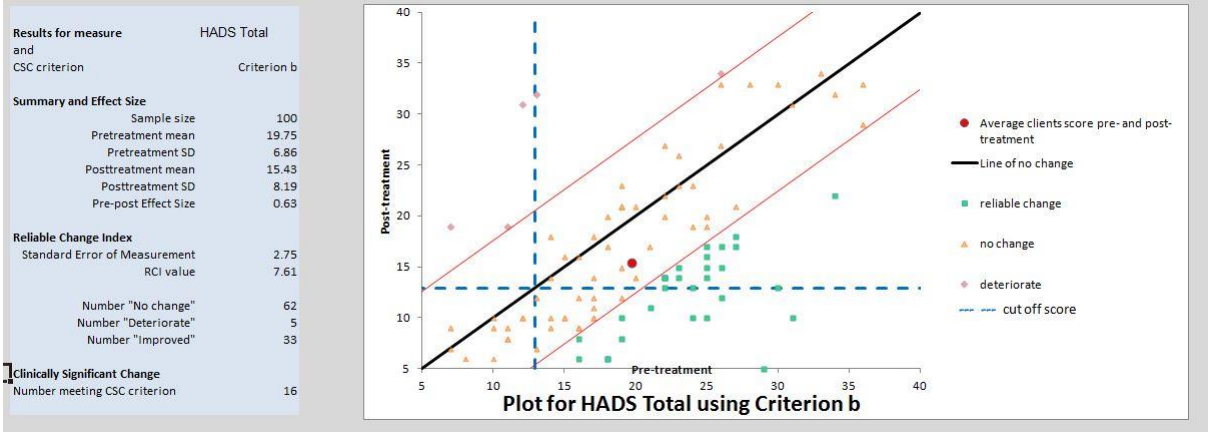
with the CSC criterion *but* importantly it only counts it as valid if the person has also made a reliable change. You have to make a reliable change before you can make a clinically significant one. Once again we've used conditional formatting to colour the CSC green in the Data sheet so that you can look at individuals. We have also summed the total number with a CSC in the Results sheet. In the present example 10 people made a clinically significant change i.e., 10%. That's about one third of those who made a reliable change.

Graphical output

The graphical output is shown below. It plots the pre and post treatment data points, the line of no change, the RCI (red, parallel lines) and cut scores. Individual data points are colour coded and the average of all the data is also shown.

The single case version just plots one data point.

You can copy and paste the data into Word or just take a screen shot using the snipping tool utility in Windows or the equivalent in OSX.



Benchmarking

Benchmarking allows you to compare the average outcome of your service with the average outcome of the aggregated data from randomised controlled trials. Developing benchmarks for a service requires some time and a small amount of statistical experience. In the clinic it is not usually possible to obtain an untreated control group and in effect you only have the equivalent of a treatment arm from a trial. The method by Minami and his colleagues (2007 and 2008) uses relevant randomised controlled trials. Once these are identified the pre-treatment and post treatment values of the outcome measures are extracted. These data are then meta analysed to produce estimates of the change between pre and post treatment time points. The estimates are in the form of an average effect size. The raw effect size for each measure is defined as:

$$ES(d) = \frac{Pretreatment\ Mean - Posttreatment\ Mean}{Pretreatment\ Standard\ Deviation}$$

and these values are aggregated in the meta-analysis.

In some fields of study the outcomes are very well defined and researchers and clinicians agree on which measures to use. For example, in many depression studies the Beck Depression Inventory is the standard measure. Fenton and Morley (2013) developed this method for use in CBT treatment for chronic pain. Figure 3 plots the effect size against the sample size (N) of the routine clinical treatment. It illustrates the essential features referred to in this account.

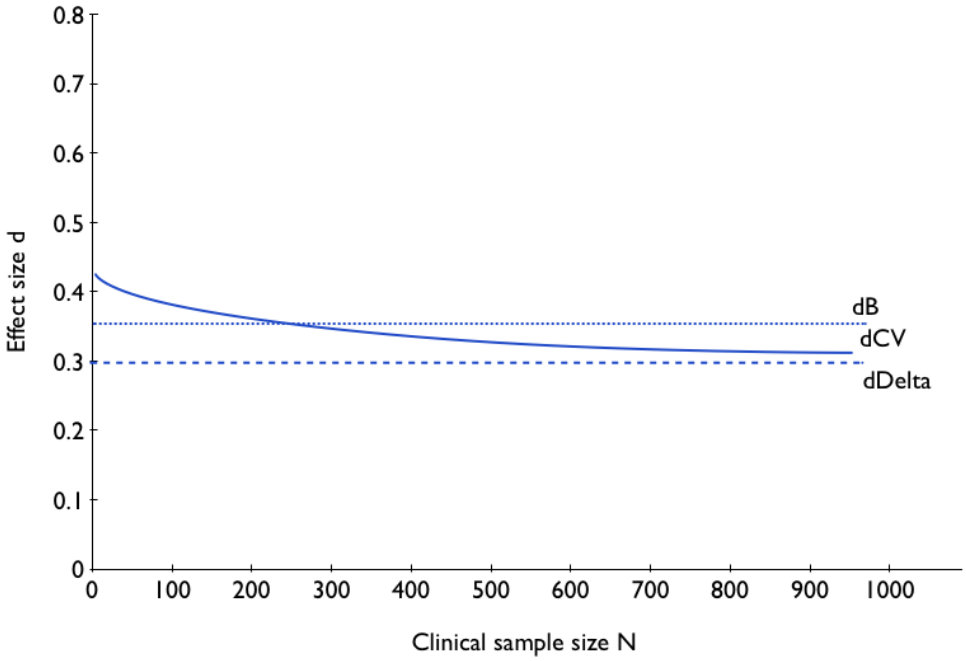


Figure 3. *Sketch of the essential features in applying meta-analytically*

A benchmark from the meta-analysis of the RCTs is computed, d_B. It represents the point estimate of the treatment arms of the RCTs. In the example in Figure 3 it is 0.35. The question is, 'what value of d ($d_{Observed}$) obtained from an evaluation of a routine clinical treatment (your data) is acceptable?' Will 0.29 suffice? Or what about 0.45? Minami et al. approach this from the perspective of setting a critical value ($d$cv) that must be exceeded for one to have reasonable confidence that $d_{Observed}$ is a good enough equivalent to d_B. The first step is to determine the lower bound of the range of d that one would be prepared to accept as being equivalent to d_B. In their studies of psychotherapy effectiveness Minami et al. determined that the lowest level, known as d_{Delta} should be $d_B - 0.2$. The justification for using the value 0.2 came from the common consensus that an ES = 0.2 is trivial. In addition there is evidence from meta-analyses that the difference between effective, *bona fide*, treatments in psychotherapy is in the range ± 0.2. To be blunt, the effect sizes in the pain RCTS are smaller than those found in the psychotherapy literature and if we adopted Minami's suggestion we'd have a massive overlap between the effects of treatment and non-treatment. So we suggest that we adopt a d_{Delta} at 20% lower than the benchmark value of 0.35. In Figure 3 we have simply sketched an arbitrary value of d_{Delta} at around 0.30 for

illustrative purposes. The critical value that the $d_{Observed}$ value must obtain in order to be sure at the 95% confidence level is d_{CV} (where CV stands for critical value). This value is determined by the sample size (N) available in your PMP and by something called the non-centrality parameter λ, of the *t* distribution[1]. The d_{CV} is plotted in Figure 3. If you look at this you will see that when there are relatively small samples in the routine clinical treatment one must obtain a $d_{Observed}$ value that is greater than the benchmark value in order to have confidence that the routine treatment is within the acceptable range.

If the routine clinical treatment meets the benchmark criterion you can probably stop the analysis there; but what if the obtained value of ($d_{Observed}$) is notably smaller? We might then need to test this against the benchmark for the control treatments. In this case we would use the value of d for the control treatment and set a value for d_{CV} that is at the *upper* end of the range i.e. Delta would take a positive value and the plots of d_{CV} and d_{Delta} would lie above the value of the benchmark. The test is then conducted to determine if $d_{Observed}$ exceeds the critical value. If it does then we might conclude that our treatment, while not as good as the RCT treatment benchmark, is better than the untreated control benchmark.

Computing the critical value in Excel™ is not (to my knowledge) possible at the moment using the available functions in the software. You can compute the relevant parameters in R.

[1]. The non-centrality parameter, $\lambda = \sqrt{N} \times (dB - Delta)$, where Delta = the difference in the effect size you are willing to accept as the lowest bound; 0.2 in Minami's case.

References

Fenton G, Morley S. A tale of two RCTs: Using Randomized Controlled Trials to Benchmark Routine Clinical (Psychological) Treatments for Chronic Pain. Pain 2013. doi: dx.doi.org/10.1016/j.pain.2013.06.033.

Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. J Consult Clin Psychol 1999;67(3):300-307.

Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19.

Minami T, Serlin RC, Wampold BE, Kircher JC, Brown G. Using clinical trials to benchmark effects produced in clinical practice. Quality & Quantity: International Journal of Methodology 2008;42(4):513-525. doi: DOI 10.1007/s11135-006-9057-z.

Minami T, Wampold BE, Serlin RC, Hamilton EG, Brown GS, Kircher JC. Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. J Consult Clin Psychol 2008;76(1):116-124. doi: 10.1037/0022-006X.76.1.116.

Minami T, Wampold BE, Serlin RC, Kircher JC, Brown GS. Benchmarks for psychotherapy efficacy in adult major depression. J Consult Clin Psychol 2007;75(2):232-243. doi: 10.1037/0022-006X.75.2.232.