Indices of Change:
Meaningful analysis of outcome measures

CORC Members' Forum - 20th November 2014

Robbie Newman

Adapted from: **Comparison of indices of clinically meaningful change in Child and Adolescent Mental Health Services: difference scores, reliable change, crossing clinical thresholds and 'added value'; an exploration using parent rated scores on the SDQ**

Miranda Wolpert, Anke Görzig, Jessica Deighton, Andrew JB Fugard, Robbie Newman & Tamsin Ford

Accepted for publication in Child & Adolescent Mental Health

ebpu Evidence Based Practice Unit

Part of Anna Freud Centre

UCL

CORC Child Outcomes Research Consortium

Background and aims of the session

- Meaningful analysis of outcome measures
- Why do you collect outcome measures?
- What are people hoping to gain from today's session?

Scenario 1 - Sally

Sally is a 14 year old girl who was referred to Erinsborough CAMHS because of her increasingly high levels of anxiety. Sally's Mum made an appointment with her GP when Sally started avoiding going to school and stopped playing in her football team. After waiting for 3 months to start treatment, Sally has had 12 sessions of CBT with her psychologist, Andrew. Sally is feeling much less anxious than she was when she was first referred and has been able to go back to school and see her friends more often.

Parent SDQ Time 1 - **21**	Parent SDQ Time 2 - **15**
Child SDQ Time 1 - **19**	Child SDQ Time 2 - **missing**
CGAS Time 1 - **63**	CGAS Time 2 – **81**

What would you do with this information?
Specifically, how would you analyse Sally's scores on the Parent SDQ?

Crossing Clinical Threshold

Classifies cases according to the clinical cut-off point (PSDQ ≥ 17)
Cases can be classified as:

- ***Recovered*** (move from clinical to non-clinical)
- ***No change*** (remain clinical or remain non-clinical)
- ***Deteriorated*** (move from non-clinical to clinical)

Parent SDQ Time 1 – **21 (clinical)** Parent SDQ Time 2 – **15 (non-clinical)**
In Sally's case this method would classify her as 'recovered'

What about children that:
a) move from 17 to 16?
b) move from 25 to 17?

- Attempts to determine a score that distinguishes between a clinical and a functional population

- Does not differentiate between smaller and larger changes
- May be difficult to determine clinical cut-off

Reliable Change Index

$$d = (time\ 1\ - time\ 2)/\text{SEdiff}$$

Used to calculate if a change is statistically significant, given the reliability of the measure.

Parent SDQ Time 1 – **21** Parent SDQ Time 2 – **15**

In Sally's case this method would not report a reliable improvement

Given that the treatment has helped Sally go back to school and she and her Mum report that she is feeling much better, does this measure reflect the change seen in therapy?

- Attempts to measure statistically reliable change

- Low sensitivity to small but clinically significant change
- Does not necessarily indicate clinically significant change

Scenario 2 - Andrew

Andrew is a clinical psychologist who has been working at Erinsborough CAMHS for 2 years. He currently has a case load of 6 children and young people (including Sally).

Parent SDQ Time 1	Parent SDQ Time 2
21	15
13	14
20	11
18	16
18	12
25	19

Mean Parent SDQ Time 1 – **19.2**

Mean Parent SDQ Time 2 – **15.5**

What methods would you use to explore the outcomes of Andrew's case-load?

Scenario 2 - Andrew

Parent SDQ Time 1	Parent SDQ Time 2	Difference Score	Crossing the clinical threshold	Reliable Change Index
21	15	6	Recovered	No change
13	14	-1	No change	No change
20	11	9	Recovered	Reliable Improvement
18	16	2	Recovered	No change
18	12	6	Recovered	No change
25	19	6	No change	No change

Mean Parent SDQ Time 1 – **19.2** Mean Parent SDQ Time 2 – **15.5**

CCT: 4 out of 6 young people showed were classified as 'recovered'

RCI: 1 out of 6 young people showed an improvement that was statistically significant

What conclusions would you draw from this data?

Difference Score

$$d = (time\ 1\ - time\ 2)/SD$$

Mean change score between time 1 and time 2. An effect size can be calculated by dividing by the standard deviation at baseline.

Mean Parent SDQ Time 1 – **19.2** Mean Parent SDQ Time 2 – **15.5**
For Andrew's case-load the average difference score would be 4.7. The standard deviation is 4.0, which means the effect size for the difference score is 1.2

Why would calculating the difference score be useful for Andrew?
What are the limitations of using the difference score?

- Simple to understand and calculate.
- Can be standardised by calculating an effect size in groups.

- May not indicate clinical significance
- Difficult to compare if no control group

Added Value Score

$$AVS = 2.3 + 0.8 * T1Total + 0.2 * T1Impact - 0.3T1Emotion - T2Total$$

An algorithm was developed with the aim of removing the influence of random fluctuation, regression to the mean and spontaneous improvement on the change scores of Parent SDQs.

A mean added value score of zero indicates that the population shows no change over that expected in an untreated sample, a negative score indicates that the change in scores is worse than predicted, and a positive score suggests the change in scores is better than predicted

The effect size of the Added Value Score for Andrew's case-load is 0.69

How would you interpret this score?
What are the limitations of using the added value score?
How does the added value score compare to the difference score?

- Attempts to take into account other changes to determine what change has taken place due to the intervention

- Is bound to the population and measure for which the algorithm was developed
- Does not necessarily indicate clinical significance.

Scenario 3 – Your Service!

Which statistical methods (if any) would you want to use in the following contexts within your service:

a) Clinically with a young person or their parents?
b) As a clinician reviewing your case-load for an upcoming supervision?
c) Comparing teams within a service?
d) Comparing services nationally?

What other methods of reporting would you want including in the report?

- Triangulation of data
- Experience of service questionnaires
- Everyday functioning
- Therapeutic alliance
- Anything else?

Conclusions & Feedback

- There is more than one way of analysing outcome measures!
- Each index may be appropriate for different contexts: CCT and RCI may be best suited to use for individual case review; whereas DS and AVS may be more appropriate for case-mix adjusted national reporting.
- Try using the indices locally in your service or speak to the CORC team about how alternative indices can be incorporated into CORC reporting

Was the session today useful?

Was the information presented clear?

Will you try to use any of the alternative indices in your service?

Any other feedback? Anything which could be done differently?