A ‘new’ way of reporting outcomes?

Philip Burgess, Meredith Harris & Tim Coombs

6th Australasian Mental Health Outcomes and Information Conference – June 2017
Australian Mental Health Outcomes and Classification Network

The Australian Mental Health Outcomes and Classification Network (AMHOCN) was established by the Australian Government in December 2003 to provide leadership to the mental health sector to support the sustainable implementation of the National Outcomes and Casemix Collection (NOCC) as part of routine clinical practice. AMHOCN manages the NOCC on behalf of the Australian Government.
Abstract: 18 April 2017

• **Background:** One of the early tasks with the implementation of the National Outcomes and Casemix Collection (NOCC) was to determine how to report statistical indicators of “outcomes”. That work considered a number of methodologies, including Effect Size (ES), Reliable Change Index (RCI) and Standard Error of Measurement (SEM) and concluded that the ES method was fit for purpose. Recently, there have been a number of studies reporting statistical outcomes using an alternative methodology, “normalized T-scores”, with claims that it is a more robust approach.

• **Aims:** This study evaluates different methodologies for the statistical reporting of routine outcome measures.

• **Methods:** Outcome indicators were derived for children & adolescents, adults and older persons, using the HoNOSCA, HoNOS & HoNOS65+ respectively, for consumers receiving specialised mental health care in either admitted or ambulatory settings.

• **Results:** Analyses compared different statistical approaches to measuring outcomes and the implications for reporting these at both an individual-consumer level as well as at a service-level.

• **Discussion:** Choosing the “right” metric for reporting outcomes is not straightforward: it is challenging alone to find consensus about what ought to be the guiding principles. Decisions need to be informed by the purpose and the level of reporting. They need to also consider costs and benefits as they vary in terms of how indicators are derived and how they can be understood by a range of different stakeholders.
Confession: 30 June 2017

• **Background:** One of the early tasks with the implementation of the National Outcomes and Casemix Collection (NOCC) was to determine how to report statistical indicators of “outcomes”. That work considered a number of methodologies, including Effect Size (ES), Reliable Change Index (RCI) and Standard Error of Measurement (SEM) and concluded that the ES method was fit for purpose. Recently, there have been a number of studies reporting statistical outcomes using an alternative methodology, “normalized T-scores”, with claims that it is a more robust approach.

• **Aims:** This study evaluates different methodologies for the statistical reporting of routine outcome measures.

• **Methods:** Outcome indicators were derived for children & adolescents, adults and older persons, using the HoNOSCA, HoNOS & HoNOS65+ respectively, for consumers receiving specialised mental health care in either admitted or ambulatory settings.

• **Results:** Analyses compared different statistical approaches to measuring outcomes and the implications for reporting these at both an individual-consumer level as well as at a service-level.

• **Discussion:** Choosing the “right” metric for reporting outcomes is not straightforward: it is challenging alone to find consensus about what ought to be the guiding principles. Decisions need to be informed by the purpose and the level of reporting. They need to also consider costs and benefits as they vary in terms of how indicators are derived and how they can be understood by a range of different stakeholders.
Works in progress...

- The renovations have started – will they ever end?
  - (AMHOCN seems to do a lot of “work in progress”):
    - Systematic review of the HoNOS tribe (Claudia);
    - Effect of consumer-rated measures (Meredith);
    - Developing a new consumer-rated measure (Gavin, Tim);
- Eventually, AMHOCN gets there (or somewhere ...) ...
Overview

1. What’s “old”? 
2. What’s “new”? 
3. What / when next?
The “old” - previous efforts ...(AMHOIC 2008!)

The challenge of demonstrating change...

Philip Burgess
The “old” - previous efforts ...

Australia and New Zealand Health Policy

Research

Do adults in contact with Australia’s public sector mental health services get better?
Philip Burgess*, Jane Pirkis and Tim Coombs

Address: 1School of Population Health, The University of Queensland, Brisbane, Australia, 2School of Population Health, The University of Melbourne, Melbourne, Australia and 3New South Wales Institute of Psychiatry, Sydney, Australia

Email: Philip Burgess* - p.burgess@uq.edu.au; Jane Pirkis - j.pirkis@unimelb.edu.au; Tim Coombs - timcoombs@bigpond.com

* Corresponding author

Published: 30 August 2006
Received: 20 February 2006
Accepted: 30 August 2006

The “old” - previous efforts ...

Australian and New Zealand Journal of Psychiatry

Modelling candidate effectiveness indicators for mental health services
Philip Burgess *, Jane Pirkis ‡, Tim Coombs ¶
* School of Population Health, University of Queensland, Sumner Park, BC, Australia ‡ School of Population Health, University of Melbourne, Melbourne, Victoria, Australia ¶ Australian Mental Health Outcomes and Classification Network, New South Wales Institute of Psychiatry, Sydney, New South Wales, Australia

Online Publication Date: 01 June 2009
The “old” - previous efforts ...

What does ‘clinical significance’ mean in the context of the Health of the Nation Outcome Scales?

Philip Burgess, Tom Trauer, Tim Coombs, Rod McKay and Jane Pirkis

Objective: The aim of this paper was to improve understanding of what ‘clinical significance’ means in relation to the Health of the Nation Outcome Scales (HoNOS) and its older persons and child/adolescent equivalents (the HoNOS65+ and HoNOSCA).

Method: An anonymous, web-based survey was completed by 94 outcome measurement experts, most of whom had clinical responsibilities. Respondents were asked to indicate for acute inpatient and ambulatory settings: the rating on each item which represented a clinically significant problem; the relative importance of each item in determining overall clinical severity; and the items which would not be expected to improve between admission and review, admission and discharge, review and review, and review and discharge.

Results: A score of 2 (‘mild problem but definitely present’) on each HoNOS/HoNOS65+/HoNOSCA item resonates with experts as being evidence of a clinically significant problem that requires active monitoring or intervention. In the main, all items on these instruments are viewed as equally important in making an overall judgement of clinical severity. The items making up the impairment and, to a lesser extent, social subscales are least likely to demonstrate change during the course of an episode of care, according to expert opinion. Generally, these findings apply across instruments and service settings.

Conclusions: Overall, the findings provide support for the content validity and clinical utility of the HoNOS/HoNOS65+/HoNOSCA. Further exploration of the question of clinical significance as reflected in these instruments could take a number of forms.

Key words: clinical significance, Health of the Nation Outcome Scales, outcome measurement.
Effect size distribution

The variance of $g$ is $2(1 - \rho)/n$ times the variance of a non-central $t$, which is defined by Johnson and Kotz (1970) as

$$\sigma^2(t) = \left( \frac{df}{df - 2} \right) (1 + \phi^2) - \frac{\phi^2}{[c(df)]^2}.$$  \hspace{1cm} (6)

Substituting $df = n - 1$ and the non-centrality parameter specified in (3), the variance of $g$ is

$$\sigma^2(g) = \left( \frac{2(1 - \rho)}{n} \right) \left[ \frac{df}{df - 2} \left( 1 + \frac{n}{2(1 - \rho)} \delta^2 \right) - \frac{n\delta^2/[2(1 - \rho)]}{[c(df)]^2} \right]$$

$$= \frac{2(1 - \rho)}{n} \left( \frac{n - 1}{n - 3} \right) \left( 1 + \frac{n}{2(1 - \rho)} \delta^2 \right) - \frac{\delta^2}{[c(n - 1)]^2}. \hspace{1cm} (7)$$
The Italians had a solution – Parabiaghi et al
The Australians thought about that ...
... and we thought about it some more ...
Records identified through database searching (n = 2,590)

Records after duplicates removed (n = 1,210)

Records screened using title and abstract (n = 1,210)

Records excluded (n = 307)

Full-text articles assessed for eligibility (n = 903)

Reported data on the measurement properties of HoNOS family measures (n = 146)

Reported data on the inter-rater reliability of the HoNOS or HoNOS-65+ (n = 26)

Full-text articles excluded, with reasons (n = 757):

- Not published in a full article (n = 93)
- Other original study involving HoNOS (n=555)
- HoNOS measure not clinician-rated (n=3)
- Does not use HoNOS measure(s) (n = 86)
- Duplicate results (n = 1)
- Full text not (yet) available (n = 19)
Which metrics are commonly used?

- 2017 HoNOS review of family of measures (Claudia’s presso yesterday – thanks Claudia!):
  - identified 701 original studies using a HoNOS family measure
  - of these 701, 317 examined HoNOS change as an outcome

<table>
<thead>
<tr>
<th>Metric</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any metric below</td>
<td>80</td>
<td>25%</td>
</tr>
<tr>
<td>Effect size (ES)</td>
<td>46</td>
<td>58%</td>
</tr>
<tr>
<td>Classify and Count (CC)</td>
<td>24</td>
<td>30%</td>
</tr>
<tr>
<td>Reliable Change Index (RCI)</td>
<td>13</td>
<td>16%</td>
</tr>
<tr>
<td>Reliable and Clinically Significant Change (RCSC)</td>
<td>13</td>
<td>16%</td>
</tr>
<tr>
<td>Standard Error of Measurement (SEM)</td>
<td>3</td>
<td>4%</td>
</tr>
<tr>
<td>Clinically Significant Change (CSC)</td>
<td>1</td>
<td>~1%</td>
</tr>
</tbody>
</table>
~2006: AUS adopted ES as the preferred metric –

Oh dear ... is it our fault?

• The ratio of the difference between baseline and follow-up scores to the standard deviation of the baseline score;
• Mainly used for aggregate data – comparing groups – but some applications with individual cases
• ES independent of sample size
• Rules of thumb for small, medium and large effects
The variance of $g$ is $2(1 - \rho)/n$ times the variance of a non-central $t$, which is defined by Johnson and Kotz (1970) as

$$\sigma^2(t) = \left(\frac{df}{df - 2}\right)(1 + \phi^2) - \frac{\phi^2}{[c(df)]^2}. \quad (6)$$

Substituting $df = n - 1$ and the non-centrality parameter specified in (3), the variance of $g$ is

$$\sigma^2(g) = \left(\frac{2(1 - \rho)}{n}\right) \left[\frac{df}{df - 2} \left(1 + \frac{n}{2(1 - \rho)} \delta^2\right) - \frac{n\delta^2/[2(1 - \rho)]}{[c(df)]^2}\right]$$

$$= \frac{2(1 - \rho)}{n} \left(\frac{n - 1}{n - 3}\right) \left(1 + \frac{n}{2(1 - \rho)} \delta^2\right) - \frac{\delta^2}{[c(n - 1)]^2}. \quad (7)$$
The less complicated answer ...

4

A positive change score of 4 or more = “Significant Improvement” = “better”

A negative change score of 4 or less = “Significant Deterioration” = “worse”

A change score in between = “No change” = “same”
Reporting Effect Size: “Groups”

Figure 1 - Clinical outcomes of people receiving various types of mental health care, 2010-11

<table>
<thead>
<tr>
<th>State and Territory Public Mental Health Services</th>
<th>Private Hospital Psychiatric Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>People in ongoing community care</td>
<td>People discharged from hospital</td>
</tr>
<tr>
<td>People discharged from community care</td>
<td>People discharged from hospital</td>
</tr>
<tr>
<td>People discharged from hospital</td>
<td>People discharged from hospital</td>
</tr>
</tbody>
</table>

- **A**: Based on difference between first and last clinical ratings made in the year for people in longer term, ongoing community care.
- **B**: Based on difference in clinical ratings at admission and discharge from hospital or community care.

Legend:
- **Significant improvement**
- **No significant change**
- **Significant deterioration**

Understanding Effect Size: “Individuals”

AMHOCN - Web Decision Support Tool

**Options**
- Jurisdiction: National
- Age Group: Adult
- Measure: HOMOS
- View: Total Score
- Level of Analysis: Collection Occasion
- Occasion: Admission
- Individual Comparison: 11
- Status Score: Ambulatory

**Statistics**

<table>
<thead>
<tr>
<th>Summary</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>125392</td>
<td>11.6</td>
<td>6.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.0</td>
<td>7.0</td>
<td>11.0</td>
<td>15.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Chart

All data reported from 1st of July, 2012 to 30th of June, 2015.
Understanding Effect Size: “Individuals”

AMHOCN - Web Decision Support Tool

Options
- Jurisdiction: National
- Age Group: Adult
- Measure: HoNOS
- View: Total Score
- Level of Analysis: Episode Transition
- Transition: Admission/Discharge
- Individual Comparison: 4
- Change Score: Ambulatory

Statistics

<table>
<thead>
<tr>
<th>Summary</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>38730</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Percentiles

<table>
<thead>
<tr>
<th>Percentile</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-4.0</td>
<td>0.0</td>
<td>3.0</td>
<td>7.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Chart

Significant Improvement: Green
Improvement: Yellow
No Change: Orange
Deterioration: Red
Significant Deterioration: Black

All data reported from 1st of July, 2012 to 30th of June, 2015.

Comparison

Score: 4
Percentile: 61.8

61.8% of assessments for consumers with the episode transition 'Admission/Discharge' while in Ambulatory care have a change score of 4 or less on the HoNOS Total Score.

38.2% of assessments for consumers like this have a change score of more than 4 on the HoNOS Total Score.
Understanding Effect Size: “Individuals”

AMHOCN - Web Decision Support Tool

Options

- Jurisdiction: National
- Age Group: Adult
- Measure: HoNOS
- View: Total Score
- Level of Analysis: Collection Occasion
- Occasion: Admission
- Individual Comparison: 7
- Status Score: Ambulatory
- Service Setting: Ambulatory

Statistics

Summary

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>125392</td>
<td>11.6</td>
</tr>
<tr>
<td>6.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percentiles

<table>
<thead>
<tr>
<th>Score</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>7.0</td>
<td>11.0</td>
<td>15.0</td>
<td>20.0</td>
<td></td>
</tr>
</tbody>
</table>

All data reported from 1st of July, 2012 to 30th of June, 2015.

Comparison

- Score: 7
- Percentile: 27.4

27.4% of assessments for consumers at Admission to Ambulatory care score 7 or less on the HoNOS Total Score: 72.6% of assessments for consumers like this score than 7 on the HoNOS Total Score.
So, what’s the problem?

- The metric: the most commonly reported metrics (Effect Size, Reliable Change, Standard Error of Measurement) are **distribution based** methods:
  - *if change scores are not normally distributed, then the change statistics are less likely to be valid...*

- **4 points** means different things:
  - Admission = 11 & Discharge = 7: **Significant Improvement**;
  - Admission = 40 & Discharge = 36: **Significant Improvement?**

  99.9% of assessments for consumers at Admission to Ambulatory care score 36 or less on the HoNOS Total Score;

  0.1% of assessments for consumers like this score more than 36 on the HoNOS Total Score
Moderate Effect Size

<table>
<thead>
<tr>
<th>Significant Deterioration</th>
<th>No Change</th>
<th>No Change</th>
<th>Significant Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= - 4</td>
<td></td>
<td></td>
<td>&gt;= 4</td>
</tr>
</tbody>
</table>

Change
"New" methods? deBeurs et al 2015

Comparing Methods to Denote Treatment Outcome in Clinical Research and Benchmarking Mental Health Care

Edwin de Beurs, Marko Barendregt, Arco de Heer, Erik van Duijn, Bob Goeree, Margot Kloos, Kees Kooiman, Helen Lionarons and Andre Merks

1 SBG, Bilthoven, the Netherlands
2 Clinical Psychology, Leiden University, Leiden, the Netherlands
3 GGZ-Delfland, Delft, the Netherlands
4 Symaedia, Leeuwarden, the Netherlands
5 Propersona, Renkum, the Netherlands
6 Riagg Rijnmond, Vlaardingen, the Netherlands
7 Lionarons-GGZ, Heerlen, the Netherlands
8 Emergis, Goes, the Netherlands

Approaches based on continuous indicators (the size of the pre-to-post-test change; effect size or ΔT) and on categorical indicators (Percentage Improvement and the Jacobson–Truax approach to Clinical Significance) are evaluated to determine which has the best methodological and statistical characteristics, and optimal performance, in comparing outcomes of treatment providers. Performance is compared in two datasets from providers using the Brief Symptom Inventory or the Outcome Questionnaire. Concordance of methods and their suitability to rank providers is assessed. Outcome indicators tend to converge and lead to a similar ranking of institutes within each dataset. Statistically and conceptually, continuous outcome indicators are superior to categorical outcomes as change scores have more statistical power and allow for a ranking of providers at first glance. However, the Jacobson–Truax approach can complement the change score approach as it presents outcome information in a clinically meaningful manner. Copyright © 2015 John Wiley & Sons, Ltd.
Denoting treatment outcome in child and adolescent psychiatry: a comparison of continuous and categorical outcomes

Edwin de Beurs · Marko Barendregt · Bente Rogmans · Sylvana Robbers · Marieke van Geffen · Marleen van Aggelen-Gerrits · Huub Houben

Received: 29 April 2014/Accepted: 23 August 2014/Published online: 3 September 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Various approaches have been proposed to denote treatment outcome, such as the effect size of the pre-to-posttest change, percentage improvement, statistically reliable change, and clinical significant change. The aim of the study is to compare these approaches and evaluate their aptitude to differentiate among child and adolescent mental healthcare providers regarding their treatment outcome. Comparing outcomes according to continuous and categorical outcome indicators using real-life data of seven mental healthcare providers, three using the Child Behavior Checklist and four using the Strengths and Difficulties Questionnaire as primary outcome measure. Within each dataset consistent differences were found between providers and the various methods led to comparable rankings of providers.

Statistical considerations designate continuous outcomes as the optimal choice. Change scores have more statistical power and allow for a ranking of providers at first glance. Expressing providers’ performance in proportions of recovered, changed, unchanged, or deteriorated patients has supplementary value, as it denotes outcome in a manner more easily interpreted and appreciated by clinicians, managerial staff, and, last but not least, by patients or their parents.

Keywords Treatment outcome research · Effect size (ES) · Reliable change index (RCI) · Percentage improvement (PI) · Benchmarking
Normalized T-scores

• 2 studies from The Netherlands, proposing an alternative metric for reporting change:
  – These studies were designed primarily thinking about reporting change for **benchmarking purposes**, not “individual” level change;
  – Their rationale was that **raw change scores do not reflect like-with like differences** at an interval level;
  – The studies examined the CBCL, SDQ, BSI & & OQ-45 – measures that have “population” norms;
    • **The HoNOS does not have “population” norms** ....
NORMALIZED T SCORES
The transformation of raw test scores to standard scores is a common means of obtaining score comparability. Unfortunately, unless the distributions from which the scores are drawn have the same shape, standard scores will not be comparable across distributions. This problem will be most pronounced if distribution A is positively skewed and distribution B is negatively skewed. In this situation a Z score of 2.00 will represent a very different centile equivalent in the two distributions.

McCall [1] has suggested a procedure that normalizes a distribution and uses a standard score based on a conversion of Z scores to mean \( M = 50 \), standard deviation \( \sigma = 10 \). The result is called a McCall T score or a normalized T score.

The procedure requires finding the centile equivalent of each score, converting that centile to a Z score from an appropriate table of the unit normal distribution, and then converting the Z score to a standard score with \( M = 50 \), \( \sigma = 10 \). The resulting scores, if plotted against frequency, will be distributed normally regardless of the shape of the original distribution.

REFERENCE
CHAPTER X

SCALING THE TEST. T SCALE—AGE VARIABILITY UNIT

I. THE METHOD OF SCALE CONSTRUCTION

Preparation of Test Material.—Recently I undertook, at the suggestion of Thorndike, the task of constructing a much-needed series of reading scales. A careful study of previous methods of scale construction led to the conviction that there was great need for revision. Perhaps the best way to show the method evolved and why it was evolved to correct some of the defects of existing methods is to describe in detail just how one of these reading scales was constructed. The steps in the process, with some alterations, were as follows:
The following drama will illustrate the need for a commonly understood reference point:

TRAGI-COMEDY OF ERRORS

ACT FIRST

Railroad Station, Richmond, Va.

Enter Traveler, Native of Baltimore, Native of Savannah, Bostonian and Author of "How to Measure in Education."

Traveler: Is New York City farther than Philadelphia?

Author: Define your point of reference. (Exit Author.)

Native of Baltimore: Yes.


292

How to Measure in Education

Native of Savannah: Yes.

Bostonian: No! (Exit Bostonian.)

Traveler: How much farther is New York City than Philadelphia?

Native of Baltimore: About twice.

Native of Savannah: About one-tenth!

The End
Normalized T-scores: How to do it?

• 3 initial steps:
  1. Find centile equivalent of each “raw” score;
  2. Convert that centile to a Z score from a table of the unit normal distribution; and then
  3. Convert the Z score to a standard score with a mean of 50 & a SD of 10
... a simple look-up table ...

<table>
<thead>
<tr>
<th>HoNOS Total Score</th>
<th>Normalized T-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.8</td>
</tr>
<tr>
<td>1</td>
<td>4.3</td>
</tr>
<tr>
<td>2</td>
<td>6.3</td>
</tr>
<tr>
<td>3</td>
<td>9.1</td>
</tr>
<tr>
<td>4</td>
<td>12.6</td>
</tr>
<tr>
<td>5</td>
<td>17.0</td>
</tr>
<tr>
<td>6</td>
<td>22.2</td>
</tr>
<tr>
<td>7</td>
<td>28.3</td>
</tr>
<tr>
<td>8</td>
<td>35.0</td>
</tr>
<tr>
<td>9</td>
<td>42.3</td>
</tr>
<tr>
<td>10</td>
<td>49.8</td>
</tr>
<tr>
<td>11</td>
<td>57.4</td>
</tr>
<tr>
<td>12</td>
<td>64.7</td>
</tr>
<tr>
<td>13</td>
<td>71.5</td>
</tr>
<tr>
<td>14</td>
<td>77.6</td>
</tr>
</tbody>
</table>
Normalized T-scores: Classifying change

• Several different options proposed;
• Some involving judgements about “clinically significant” change; and
• the “delta” threshold = “5”
  – (familiar... 5 is half a standardised T-score SD; 4 is half a raw score SD) ....
Comparing ES & delta classifications

- Modelled with the Adult Ambulatory Completed Episode cohort ...
- Both methods result in 3 groups: improved, deteriorated & no change:
  - Kappa = 0.84;
  - Spearman’s rho = 0.90
Comparing metrics …

<table>
<thead>
<tr>
<th>Raw Effect Size</th>
<th>Normalized T-score delta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Improved</td>
<td>92%</td>
</tr>
<tr>
<td>No change</td>
<td>7%</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>-</td>
</tr>
</tbody>
</table>
Some odd findings...

• While strong concordance overall, instances where:
  – ES = Improved & delta = No change...
  – ES = No change & delta = Improved;
  – ES = No change & delta = Deteriorated;
  – ES = Deteriorated & delta = No change

• delta can classify episodes as Deteriorated when in fact the raw score change is “positive”
Some odd findings – but can be resolved!

• delta references the difference between the baseline and follow-up normalized T-scores.....

• seems more logical to reference the change score to the baseline normalized T-score ....
End thoughts & opinion ...

• Normalized T-scores a bit fiddly but easy to publish a ready-reckoner look-up table ...
  – Noting not much material difference <> Normalized T using option presented today;

• Normalized T seems more meaningful for understanding & communicating progress for individual consumers, BUT
  – focus should be on change relative to baseline
• Pro’s & Con’s – as usual ...

• Note too other (related) alternatives (percentile change) are gaining traction ...

• There are really difficult issues with the HoNOS:
  – What is the target level of change?
  – How does it apply to real-world episodes of mental health care?
• This methodology is really exciting when thinking about the “new” consumer-rated measure ...
  – Based on the K10/K6/K5 and:
  – reference “recovery” to the general, community population
    • NSMHWB 1997 & 2007
    • NATSISS/NATSIHS 2004-05, 2008 & 2014-15

• And kids: SDQ
  • Young Minds Matter (YMM) 2015

• And primary mental health care (PHN NMDS)....