The challenge of demonstrating change in routine clinical practice

Philip Burgess, Jane Pirkis & Tim Coombs

Milestones

- 1992 – At the very outset, the National Mental Health Strategy emphasised that in order to improve mental health service outcomes, there is a need for consumer outcomes to be reviewed on a regular basis.
Milestones

1993 – 1998 (First plan)
- Andrews et al’s review of Adult measures
- Stedman et al’s review of Child & Adolescent measures
- Mental Health Classification and Services Costs Project
- Initial field trials …

1999 – 2003 (Second plan)
- Australian Government provides funds for States & Territories to develop necessary infrastructure
- National Outcomes and Casemix Collection agreed by all levels of Government
- Significant investment in the training of the mental health workforce
Milestones

- 2003 and beyond (Third plan – pre COAG)
  - Australian Mental Health Outcomes and Classification Network established
  - First National aggregate level reports published early 2005
  - Key performance indicators for Australia’s public mental health services published June 2005 (Phase1)

Recent developments

- 2006
  - ($ 4 billion over 5 years!)
  - States & Territories to provide regular reports along 12 Progress Measures
Recent developments

- Phase 2 Key Performance Indicator development commences, focussing specifically on ‘effectiveness’:

  “care, intervention or action achieves desired outcome”

- COAG Action Plan Accountability Requirements # 6 mandate the following Progress Measure:

  *Mental health outcomes of people who receive treatment from State and Territory services and the private hospital system*

Collection Occasion Rate per 10,000 Population per Quarter in Financial Year per Jurisdiction

![Chart showing the collection occasion rate per 10,000 population per quarter in financial year per jurisdiction from 2004 to 2007.](chart.jpg)

Health Outcomes Conference 2008: The challenge of demonstrating change in routine clinical practice
### Standardised Collection Occasion Ratios

![Graph showing standardised collection occasion ratios](image)

### Completion Rates of NOCC Clinical Measures – Adult Services

<table>
<thead>
<tr>
<th>Collection Occasion Age Group</th>
<th>Mental Health Service Setting</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inpatient</td>
<td>Community Residential</td>
<td>Ambulatory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoNOS</td>
<td>89, 83, 84</td>
<td>90, 95, 70</td>
<td>87, 91, 59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSP-16</td>
<td>x, x, x</td>
<td>25, 89, 40</td>
<td>x, 82, 43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer self-report</td>
<td>x, x, x</td>
<td>42, 56, 21</td>
<td>25, 23, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnoses</td>
<td>x, 90, 96</td>
<td>x, 98, 99</td>
<td>x, 98, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH Diagnoses</td>
<td>x, 89, 89</td>
<td>x, 97, 82</td>
<td>x, 97, 90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus of Care</td>
<td>x, x, x</td>
<td>x, x, x</td>
<td>x, 92, 79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH Legal Status</td>
<td>x, 89, 87</td>
<td>x, 98, 86</td>
<td>x, 98, 88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Health Outcomes Conference 2008: The challenge of demonstrating change in routine clinical practice
Outstanding issues

- Is this valid – are we really measuring effectiveness?
- Is this reliable – are we going to get the same result if we, or someone else, repeated it?
- Is this sensitive – how good at measuring change when there has been real change?
- Is this specific – how good at telling us when there has been no real change?

“Significance” – who decides

- Statistical significance? – large numbers of observations can make ‘small’ differences ‘significant’
- How much change? The ‘size of the effect’?
- How significant? Traditionally 0.05 (meaning occurs less than 1 in 20). Do we have to be that ‘certain’?
“Significance” – who decides

- How reliable? That is, the difference is not caused by ‘errors’ in our measurement?
- Clinical significance? Change in relation to what – the ‘anchor’ points in mental health are very unclear
- A ‘good’ outcome in say an inpatient setting may not be a ‘good’ outcome in ambulatory settings

Some more difficult questions

- What kind of change do we expect given the different patterns of care, different levels of acuity, different service settings?
- In some instances, ‘no change’ may be a reasonable outcome, over a given period of time, for a consumer at a particular stage of illness?
- Absolute change or relative change?
Answer #1: Effect Size?

- 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

Answer #2: Reliable Change?

- Whether the magnitude of observed change is not due to ‘noise’ in the measurement system – the instrument and/or the rater
- RCI = Change score divided by the standard error of the differences
- If this difference is +/- 1.96, then conclude its statistically reliable at 95% level
Answer #3: Standard Error of Measurement?

- The standard deviation of an individual score estimated as the SD for a sample adjusted for its statistical (un)reliability
- Rule of thumb that 1 SEM corresponds to a minimal clinically important change for an individual

And there are still others

- So far, each has pros & cons
- ES not easily grasped. Few applications about individual level change;
- RCI extended to test for ‘clinically significant change’ – requires normative data on a ‘well’ population
- SEM relatively simple but not widespread. Has a much lower threshold for detecting change
Mathematically …

- Minimum change required for clinically meaningful improvement depends on
  - how much ‘noise’ (variation) there is in the change scores
  - how reliable the measures are
- Note, all three are distribution based methods
  - if change scores are not normally distributed, then the change statistics are more likely to be invalid
Health Outcomes Conference 2008: The challenge of demonstrating change in routine clinical practice

Histogram

Age Group: Adult (18-64), Current Mental Health Service Setting: Psychiatric Inpatient Service

Mean = 7.29
Std. Dev. = 7.505
N = 24,506

Histogram

Age Group: Adult (18-64), Current Mental Health Service Setting: Ambulatory mental health service

Mean = 1.81
Std. Dev. = 6.332
N = 24,150

Health Outcomes Conference 2008: The challenge of demonstrating change in routine clinical practice
Answer #4: Count & classify - A new look at the HoNOS family?

- **Step 1:** A classification of severity (Lelliott)
  a. Very Severe = ‘3+’ on at least 2 items;
  b. Moderately Severe = ‘3+’ on 1 item only;
  c. Mild = ‘2’ on at least 1 item;
  d. Sub-clinical = No more than 1 on all items

- **Step 2:** Calculate change as movement from one class to another to indicate:
  a. Got better - Improved;
  b. No change - Stable (?)
  c. Got worse - Deteriorated

### % Severity Status – Adults

<table>
<thead>
<tr>
<th>Severity Status</th>
<th>Inpatient</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Admission</td>
<td>Discharge</td>
</tr>
<tr>
<td>Sub clinical</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>Mild</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>Moderate</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Very Severe</td>
<td>65</td>
<td>18</td>
</tr>
</tbody>
</table>
% Outcome Status – Adults

<table>
<thead>
<tr>
<th>Outcome Status</th>
<th>Inpatient</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>70</td>
<td>46</td>
</tr>
<tr>
<td>No Change</td>
<td>24</td>
<td>39</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>6</td>
<td>15</td>
</tr>
</tbody>
</table>

Count & classify

- **Disadvantages**
  - May be not all HoNOS scales equally important
  - Floor & ceiling effects

- **Advantages**
  - Doesn't depend on normal distribution assumptions
  - Simple to understand;
  - Difficult to game
Issues to be resolved:

- It’s too simple – just a number?
- It’s too complicated
  - Who gets counted?
  - What’s a ‘good’ outcome?
- What about:
  - If we don’t have a matched pair?
  - If the clinician’s view is different from the consumer’s view?

Issues to be resolved:

- Why do some consumers with ‘sub clinical’ or ‘mild’ severity get admitted to inpatient services?
- Some consumers’ community episodes ‘end’ following clinical assessment that an admission is the most appropriate treatment option – how do we measure that?
- How do we account for case complexity?
A few more issues ...

- The process of detailed clinical assessment could well lead to revised clinical ratings that don’t really reflect ‘deterioration’
- Change means different things to different stakeholders – eg, cost for marginal improvement vs greater coverage
- Need to include consumers’ perspectives on outcomes – what does a null return mean?

Some useful references