The Challenge of Demonstrating Change in Routine Clinical Practice

At the very outset, in 1992 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.

Over the course of the First National Mental Health Plan (1993 – 1998), there were several notable achievements. Andrews et al’s (1994) review of Adult measures; Stedman et al’s (1997) review of Child & Adolescent measures, the Mental Health Classification and Services Costs Project (Buckingham et al 1998) as well as the initial field trial reported by Trauer in 2003.

Under the Second National Mental Health Plan (1999 – 2003), the Australian Government provided funds for States & Territories to develop necessary infrastructure. A National Outcomes and Casemix Collection was agreed by all levels of Government and there was also significant investment in the training of the mental health workforce.

Under the Third National Mental Health Plan (2003 – 2008), the Australian Mental Health Outcomes and Classification Network was established. The first National aggregate level reports published early 2005 and also reported by Burgess et al (2006).

In parallel, Key performance indicators for Australia’s public mental health services was published June 2005 (Phase1). More recently, in 2006 the Council of Australian Governments (COAG) released a National Action Plan on Mental Health 2006 – 2011, with a funding commitment of $4 billion over 5 years. Under the National Action Plan, the States & Territories agreed to provide regular reports along 12 Progress Measures.

These initiatives have major implications for the reporting of outcomes. Phase 2 of the national key performance indicator development work is focused specifically on the domain of ‘effectiveness’ defined as whether the “care, intervention or action achieves desired outcome”.

The COAG National Action Plan Accountability Requirements mandate the following Progress Measure: Mental health outcomes of people who receive treatment from State and Territory services and the private hospital system.

Slides 8 and 9 show that routine collection of outcomes measures is still evolving with significant variation among jurisdictions in terms of the volume of measures reported. Notwithstanding, it is agreed by most key stakeholders that these data should be used with respect to national reporting of outcomes from public sector mental health services.

There remain some concerns, however, regarding the validity, reliability, sensitivity and specificity of routinely collected mental health outcome measures. Some of these concerns relate
to the statistical aspects of reported measures, other concerns relate to the distinction between ‘statistical’ and ‘clinical’ significance.

It is often noted that the concept of change has different meanings to different stakeholders – clinicians, consumers, carers and administrators, planners and managers. A ‘good’ outcome in an inpatient setting may not be a ‘good’ outcome in ambulatory settings. There are a range of other complex issues. For example, 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.

AMHOCN has been investigating several options for the reporting of these nationally agreed mental health outcome measures. Three main approaches to key performance indicators have been explored: (i) effect size - ES; (ii) reliable change - RC; and (iii) standard error of measurement - SEM.

Effect Size is calculated as the ratio of the difference between baseline and follow-up scores to the standard deviation of the baseline score. It has been mainly used for the reporting of aggregate data – comparing groups – but some applications with individual cases (Eisen et al, 2007). ES is independent of sample size and there are ‘rules of thumb’ for small, medium and large effects.

Reliable Change is estimated by correcting the magnitude of observed change for the inherent ‘unreliability’ of the measurement tool. In practice, the change score is divided by the standard error of the differences. If this difference is +/- 1.96, then it is concluded that the difference is statistically reliable at 95% level.

Standard Error of Measurement is the standard deviation of an individual score estimated as the SD for a sample adjusted for its statistical (un)reliability. Again, there are rules of thumb – a 1 SEM corresponds to a minimal clinically important change for an individual.

Each approach has various pros & cons. ES is not easily grasped and there are few applications about individual level change. RC is attractive to some practitioners with backgrounds in social science methodologies but other finds it limiting to simply note that the change is ‘reliable’. SEM also corrects for the measure’s ‘unreliability’ and has a much lower threshold for detecting change.

The challenge involves determining what level of minimum change is required for clinically meaningful improvement. This depends on at least two factors: (i) how much ‘noise’ (variation) there is in the change scores; and (ii) how reliable the measures are. All three approaches are distribution based methods. If change scores are not normally distributed, then the change statistics are more likely to be invalid.

Slides 23 & 24 show Effect Size indicators for completed adult inpatient episodes in 2006-2007. Slide 23, based on a ‘medium’ effect size of 0.5 shows, approximately 50% of episodes should ‘significant’ improvement but 48% show no change. The ‘no’ change classification can be further disaggregated by reporting ‘trend’ data – thus, 36% show some improvement whereas 9% show some deterioration.

An alternative method is what we term ‘count and classify’. This approach is based on the work of Lelliott reported by Charlwood et al (1999) and adapted by Parabiaghi et al (2005). First, a
classification of ‘clinical’ severity is determined for each baseline Health of the Nation Outcomes Scale (HoNOS) measurement and this level of severity is compared to that measured at follow-up.

The ‘count & classify’ approach has greater appeal to clinicians who can relate a client’s mental health status in terms of this severity classification and, importantly, it is intuitive and simple to understand. There are difficulties, however, with this approach, particularly its susceptibility to ‘floor’ and ‘ceiling’ effects where some consumers’ mental health status are at the extremes where there is either no opportunity for improvement or no opportunity for deterioration.

Some of these issues are difficult to resolve. For some stakeholders, the simplicity of effectiveness indicators masks the complexity of mental illness; for others, there is suspicion that the complexity of the statistics undermines the communication of reporting outcomes. Other issues go beyond the science of the approach and involve vexed matters such as defining what is a good outcome for consumers of public sector mental health services.

There is further work to be done that needs to address the concepts of case complexity as it impacts on service outcomes and to also consider how consumer’s own views can be accounted for in the analysis and reporting of outcome measures.

References