

## GUEST EDITORIAL

# Routine clinical outcomes measurement in old age psychiatry

### Introduction

It is puzzling that major public health systems in the developed world do not engage in the routine clinical outcomes measurement (RCOM) of their interventions. For example, the budget of the U.K. National Health Service (NHS) for 2008/9 was £81 billion, with 15 million inpatient episodes, yet a Martian's question "How much better are these patients for all this expense?" cannot be answered for any but a tiny proportion of them, and then only by local enthusiasts. It is true that death rates of patients are now routinely and publicly reported, but as Florence Nightingale – whose reputation was made and then almost destroyed by mortality statistics (Iezzoni, 1996) – herself remarked:

“if the function of a hospital were to kill the sick, statistical comparisons of this nature would be admissible. As, however, its proper function is to restore the sick to health as speedily as possible, the elements which really give information as to whether this is done or not, are those which show the proportion of sick restored to health, and the average time which has been required for this object...” (Nightingale, 1863)

One could speculate about the reasons for a lack of interest by governments in health outcomes. Costs in overcoming technological obstacles, disagreement about desired outcomes and their measurement, fear of appearing inefficient, the possibility that outcomes measurement may constrain finance-driven reorganization (Oxman *et al.*, 2005) or even subtle pressure from industry are factors that may play a part. However, psychiatry appears to be one specialty where, at least in certain countries, RCOM is gaining ground. This is also puzzling: if one wanted to name the specialty in which it would be most difficult to undertake this, it might be psychiatry. Yet in mental health services in Ohio (Ohio State Mental Health Services, 2008), Australasia (Burgess *et al.*, 2006), Utah (Lambert and Burlingame, 2007), Colorado (Huxley and Evans, 2002), Ontario (Smith, 2008) and in some U.K. NHS Mental Health Trusts, RCOM is becoming part of the landscape of clinical activity in mental health services in a way that is not apparent elsewhere in medicine.

In the U.K., government direction and financial support for the development of RCOM in mental health services has been at best ambivalent; motivation has perforce come from clinicians. The third puzzlement is that the only U.K. Mental Health Trust in which RCOM has yet developed to any great extent has been led by old age psychiatry, the very subspecialty where, at first glance, one might expect least rewards from such a process.

To sum up: hardly anyone in the health profession appears interested in RCOM. Of those who are, the specialty with the most difficult measurement issues is leading development worldwide; and within this specialty in the U.K., the leading subspecialty has some of the worst likely clinical outcomes. This can be expressed more positively as follows: if RCOM can be achieved in old age psychiatry, then it can be achieved in all mental health services. If it can be achieved in mental health services, then it can be achieved in all medical specialties. The purpose of this review is to describe this process for old age psychiatry clinicians, especially in countries in which, unlike Australasia, government support is absent or ambivalent.

### Principles of RCOM

Routine clinical outcomes measurement is defined in many ways, but these are typical:

- a measure of the “attributable effect of an intervention or lack of intervention on a previous health state” (National Centre for Health Outcomes Development, 2009);
- a way of measuring “the end results of particular health care practices and interventions” (Foundation for Health Services Research, 1994);
- assessment of “a change in the health of an individual, group of people or population which is attributable to an intervention or series of interventions” (Frommer *et al.*, 1992).

All imply three dimensions of measurement: change in health status, intervention, and context or case mix (since different outcomes are expected in different conditions and circumstances). Although clinical change measurement issues dominate the initial phases of RCOM, case mix/context and intervention measurement are all necessary to understand outcomes. Of these, interventions are proving the most difficult to trap.

## Measurement: change in health status

### Which measure?

Whose outcomes are they anyway? Different stakeholders (patient, clinician, carer, general practitioner and others) have different reasons for interest in outcomes, perceived salience of outcomes, why particular measures would be preferred, and desired outcomes (Long and Jefferson, 1999). There are advantages and disadvantages to using measures devised by clinicians and based on their judgments (e.g. HoNOS65+; Turner, 2004), those devised by clinicians but based on patient judgments (e.g. CORE-OM; Barkham *et al.*, 2005) and those devised largely by patients and based on their judgments (e.g. the Te Pou initiative in New Zealand; Gordon *et al.*, 2004). There are also advantages and disadvantages to joint assessments between clinician and patient, advocated by some in this field. However, there is consensus that clinician-rated measures would provide the most practical starting point of RCOM within a secondary care mental health service (Fonagy *et al.*, 2005).

Although measuring change retrospectively might be tempting – for example, using a clinical global impression of change (NIMH Early Clinical Drug Evaluation, 1976) – the benefits of prospective measurement are that they avoid bias if clinician memory fails or if patients have a series of clinicians attending to them. The 12 Health of the Nation Outcome Scales (HoNOS; Wing *et al.*, 1998) were developed for this purpose, are feasible, and are sensitive to change (Gigantesco *et al.*, 2007). Each scale is scored 0 (no problem), 1 (non-clinical problem), 2 (mild), 3 (moderate) or 4 (severe problem). Ratings are based on all information available to the clinician from any source and on the worst that the patient has been in each domain in a given period (usually two weeks). For undisclosed reasons, old age psychiatrists in the U.K. decided that they must have a different glossary, namely HoNOS65+ (Burns *et al.*, 1999), which at publication had already been superseded (Macdonald, 1999) and was unworkable. Its scoresheet forced a choice between acute and chronic cognitive impairment, and there were other problems (Allen *et al.*, 1999). The HoNOS65+ glossary had already been further developed in the U.K. by the only NHS Old Age psychiatry service implementing RCOM at the time, and had been translated into Spanish, Dutch and Korean (Royal College of Psychiatrists Research Unit, 2008). However, the glossary used in Australia and most of New Zealand is more or less that originally published. No-one knows whether different glossaries make any difference to

the way HoNOS65+ ratings are derived in everyday practice, nor, by the same token, whether there are any differences between ratings made using HoNOS65+ and HoNOS.

### Who measures?

Bilsker and Goldner (2002) predicted types of bias that would be introduced when staff rate their patients' health before and after treatment that they themselves have managed. These included gaming (tendency to adjust severity scores so that they are higher at referral and lower at discharge than justified), selection and attrition bias (putting more effort into obtaining baseline and/or follow-up data for patients with better likely outcomes than others), and detection bias (assigning unrealistically positive rating change to patients who have completed treatment). They also predicted that measures feasible enough for use in everyday work would have too low validity, inadequate sensitivity or too low inter-rater reliability for meaningful use. Most of these predictions are now being tested (and found to be unjustified) using data from a RCOM program (Macdonald and Trauer, 2009), but the wide publicity given to them has inhibited the development of programs which were not already under way, perhaps contributing in no small way to "outcomes torpor" in the U.K.

Although current U.K. guidance suggests beginning with staff-based ratings, in Australia and New Zealand patient (consumer) ratings have been gathered from the outset. There have been difficulties with low rates of completion (Trauer, 2004), and in comparing the results with HoNOS-based staff measures. It is possible that measures that chime best with patients' experiences, such as those being developed by Te Pou (Gordon *et al.*, 2004), will help with the former, but may aggravate the latter, which is important both for aggregate analysis and facilitating meaningful discussions between individual patients and staff about perceptions of progress. To rectify the latter, self- and carer-rated versions of HoNOS have been developed (Stewart, 2009), and further work in this direction is proceeding in the U.K.

In Ontario, Canada, outcomes measurement uses the Camberwell Assessment of Need (Phelan *et al.*, 1995), which comprises all three perspectives (Smith, 2008); a version for older people also exists (Reynolds *et al.*, 2000).

### When do they measure?

It is obvious that measurement must be repeated if outcomes are to be assessed. The bare minimum number of collection occasions is two: at assessment and discharge from a team or ward. For very

short episodes, such as in some liaison settings, outcomes measurement is not possible, since the period covered by the two assessments overlap; nevertheless, gathering data at inception is advised because (a) it is not always possible to know in advance how long the contact will be, and (b) baseline data are useful anyway. For long-term treatment, for instance of patients with chronic psychotic states or dementia with persistent behavioral or comorbid problems, measures need repeating at regular review (e.g. every six months). In long-term care, for instance in care homes for people with dementia, the rating period covered by HoNOS65+ may need modification.

### Measurement: case mix and context

It is crucial that at the same time as data on clinical change are gathered, so data on variables that affect outcomes are also captured. Without these, understanding differences in outcomes would be impossible. Minimally, there needs to be some categorization of patient problem; but for many non-medical clinicians (and even some consultant psychiatrists) difficulties arise in using systems like ICD-10 or DSM-IV, based on a misunderstanding of their purpose, and this leads to the absence of data on which outcomes analyses depend. Objections include:

- my patient's problem cannot be summarized by the classification;
- entering a diagnosis will prejudice my patient's future, especially if it is wrong;
- diagnosis is a male-dominated oppressive act, particularly against women (Caplan and Cosgrove, 2004).

The first is amenable to some influence. Diagnosis is a shorthand description of the working hypothesis on which intervention is predicated. Any clinician involved in any intervention will have a working hypothesis (unless the intervention is chosen at random) so the issue becomes whether their hypothesis maps well enough onto the classification to be acceptable, if not perfect. We have found that even psychoanalytic psychotherapists are able to categorize their patients using ICD-10 (including the magnificent Z-codes). The second objection assumes a diagnosis is right or wrong; it is merely a hypothesis. The third objection is unanswerable; service purchaser insistence is the only remedy.

Other context variables include age, ethnicity, gender and social circumstances; all allow a richer analysis of outcomes than would be possible without them.

### Measurement: intervention

This is the most vexing area of RCOM. Although data on frequency and duration of staff contacts with community patients can readily be captured, there are no extant mental health information systems that adequately capture the intervention content of the contact, save when it is for a defined activity such as electroconvulsive therapy or a formal psychological therapy. In the U.K. most pharmacological interventions for community mental health patients are executed by general practitioners; data are not yet synchronized with secondary mental health care systems so outcomes cannot be systematically related to medication. Inpatient electronic prescribing is almost here, and when it arrives such analysis will be possible, but only for a minority of patients and episodes. Overall length of contact can be seen both as an intervention and as an outcome, and needs to be incorporated into the analysis of outcomes with this in mind.

### Categorical outcomes

Events themselves can be outcomes, particularly but not exclusively for old age psychiatry patients. Physical problems such as falls, acute medical and surgical problems, and admission to a care home or a general hospital and, of course, death should be recorded in a way that allows analysis using context and intervention data. For instance, outcomes for inpatients who are discharged to their source domicile, even if it was a care home, are likely to be better than those discharged for the first time to a care home; simply recording discharge would not discriminate between these groups. With larger and larger datasets of old age psychiatry patients, as yet unknown risk factors for physical health problems could emerge, especially if medication was recorded, while known risk factors could fade in significance.

### Implementation of RCOM

From the discussion above and from experience it would seem that in order to set up a RCOM program with a reasonable chance of integration into everyday practice the following are required:

- a culture of enquiry and of willingness to take risks
- an electronic patient record system with easy extraction from data warehouse
- resources and staff time set aside for training and receiving feedback
- management willing to consider impact of outcomes data on their decisions

- resources and personnel to extract, analyze and proactively present outcomes, case mix and, where available, intervention data to clinical teams
- regular reports on data quality as part of a performance management process by senior managers
- clear management of the process.

All active RCOM programs have many of these requirements – few have all. Trying to implement RCOM without many of them will be problematic at best. In particular, a “top-down” drive in which performance management is not complemented by active feedback may be temporarily successful but not sustainable.

### Objections to RCOM

Apart from the theoretical objections to completely staff-based measurement (dealt with above), there are some which are particularly relevant to old age psychiatry services.

There is a widely-held fear that measuring outcomes might suggest that individual clinicians or services are ineffective (Meehan *et al.*, 2006). This is especially true for old age psychiatry services which include substantial numbers of people with dementia. In the U.K. and Australia this fear has not been borne out (Macdonald, 2002; Spear *et al.*, 2002). Patients with dementia do get better overall, and if not in cognitive and physical health then in behavior and depression. Within our U.K. service, differences in apparent effectiveness between teams, controlling for case-mix and severity, has led to discussions about organization and resources, but no staff member has been yet singled out for particular attention, nor has any service been closed because of apparent ineffectiveness. There is anecdotal evidence that, compared with others, ward closure and services under threat are related to having no or low quantities of outcomes data rather than data apparently showing relative ineffectiveness. In Australia, providers tendering for competitive contracts to provide services are increasingly finding that having an outcomes system in place gives them a competitive edge over those that do not. It is felt that having a relatively objective system that tracks the progress of patients reflects a commendably evidence-based attitude to the core business (T. Trauer, personal communication, 2008).

The only published trial of RCOM found no clinical benefits of feedback of data from three measures to patient and staff members, but this was conducted for only seven months in what appeared to be a relatively stable patient group (Slade *et al.*, 2006). Although cost-benefit was shown in terms of reduced admissions, the choice of measures and

postal feedback to a staff group not necessarily “signed-up” to RCOM may have been critical; some of the key aspects of a successful RCOM system were absent. Implementing RCOM is a complex intervention not amenable to simple evaluation of this sort.

HoNOS65+ ratings take about 5 minutes to carry out; when fed back to teams along with context data there follow thoughtful discussions about the organization and delivery of treatment. Training in HoNOS65+ with other disciplines from other teams is enjoyable and enhances service cohesion.

### Threats to RCOM

There are two major threats to RCOM. First, if the necessary resources mentioned above are not diverted to RCOM it will not survive. In particular, if there is no provision or capacity for feedback to clinical teams and data are merely spirited away for use elsewhere, compliance will drop, even in a strong performance management culture. Second, even when implemented well, the data may be used inappropriately or simplistically by commissioners and purchasers primarily for purposes other than outcomes measurement, such as those of the misnamed “payment by results” process in the U.K. (Fairbairn, 2007) in which services will be paid according to the complexity and severity of their patients at intake. In this case, gaming (Bilsker and Goldner, 2002) – which is not yet evident – may grow to damage severely the capacity of RCOM to deliver the benefits it promises.

### Conclusions

The potential benefits of RCOM are vast. Validated outcome measures mean that scarce resources can be targeted to areas in which services have most impact. Since RCOM measures like HoNOS65+ cover the most relevant domains, team meetings structured around significant (2+) scores are punchier and, best of all, shorter (Stewart, 2008). Feedback of patient, clinician and carer outcomes data will enhance individual patient care and, when aggregated, team effectiveness. RCOM threatens to revolutionize the goals of healthcare and eliminate arbitrary and politically influenced target-setting based on dubious proxies for outcomes, like readmission rates. Importantly, when pre- and post-outcomes data are available, whimsical reorganization (Oxman *et al.*, 2005) will be exposed as pointless or destructive. Finally, RCOM



can produce large datasets in which the real-world effectiveness and hazards of interventions can be identified in a way that randomized controlled trials, alone or amalgamated, can never do.

### Conflict of interest declaration

The author is remunerated for training in HoNOS65+.

### Acknowledgments

Thanks to Patricia and Douglas Macdonald for comments on earlier drafts, Jonathon Artingstall, all clinical staff and administrators involved in setting up and managing the RCOM programme from 1997, Richard Carthew and Matthew Broadbent, Alice Mills and successive clinical and managerial members, over the years, of the Mental Health in Older Adults Clinical Outcomes Group.

ALASTAIR MACDONALD

South London and Maudsley NHS Foundation Trust, and Institute of Psychiatry, London, U.K.

Email: Alastair.Macdonald@iop.kcl.ac.uk

### References

- Allen, L. et al.** (1999). Experience and application of HoNOS65+. *Psychiatric Bulletin*, 23, 206.
- Barkham, M., Culverwell, A., Spindler, K. and Twigg, E.** (2005). The CORE-OM in an older adult population: psychometric status, acceptability, and feasibility. *Aging and Mental Health*, 9, 235–245.
- Bilsker, D. and Goldner, E. M.** (2002). Routine outcome measurement by mental health-care providers: Is it worth doing? *Lancet*, 360, 1689–1690.
- Burgess, P., Pirkis, J. and Coombs, T.** (2006). Do adults in contact with Australia's public sector mental health services get better? *Australian and New Zealand Health Policy*, 3, 9.
- Burns, A., et al.** (1999). Health of the Nation Outcome Scales for elderly people (HoNOS 65+). Glossary for HoNOS 65+ score sheet. *British Journal of Psychiatry*, 174, 435–438.
- Caplan, P. J. and Cosgrove, L.** (2004). *Bias in Psychiatric Diagnosis*. Lanham, MD: Rowman and Littlefield.
- Fairbairn, A.** (2007). Payment by results in mental health: the current state of play in England. *Advances in Psychiatric Treatment*, 13, 6.
- Fonagy, P., Matthews, R. and Pilling, S.** (2005). *Outcomes Measures Implementation: Best Practice Guidance*. Adapted from the Report from the Chair of the Outcomes Reference Group. Leeds: National Institute for Mental Health in England.
- Foundation for Health Services Research** (1994). *Health Outcomes Research: A Primer*. Washington, DC: Foundation for Health Services Research.
- Frommer, M., Rubin, G. and Lyle, D.** (1992). The NSW Health Outcomes Program. *NSW Public Health Bulletin*, 3, 135–137.
- Gigantesco, A., Picardi, A., de Girolamo, G. and Morosini, P.** (2007). Discriminant ability and criterion validity of the HoNOS in Italian psychiatric residential facilities. *Psychopathology*, 40, 111–115.
- Gordon, S., Ellis, P., Haggerty, C., Pere, L., Platz, G. and McLaren, K.** (2004). *Preliminary Work Towards the Development of a Self-Assessed Measure of Consumer Outcome*. Auckland: Health Research Council of New Zealand.
- Huxley, P. and Evans, S.** (2002). Quality of life routine outcomes measurement: lessons from experience in the USA and the UK. *Epidemiologia e Psichiatria Sociale*, 11, 192–197.
- Iezzoni, L. I.** (1996). 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. *Annals of Internal Medicine*, 124, 1079–1085.
- Lambert, M. J. and Burlingame, G. M.** (2007). Uniting practice-based evidence with evidence-based practice. Utah has brought all stakeholders together in a consumer-focused outcomes measurement system. *Behavioral Healthcare*, 27, 16–20.
- Long, A. and Jefferson, J.** (1999). The significance of outcomes within European health sector reforms: towards the development of an outcomes culture. *International Journal of Public Administration*, 22, 385–424.
- Macdonald, A. J.** (1999). HoNOS 65+ glossary. *British Journal of Psychiatry*, 175, 192.
- Macdonald, A. J.** (2002). The usefulness of aggregate routine clinical outcomes data: the example of HoNOS65+. *Journal of Mental Health*, 11, 645–656.
- Macdonald, A. J. and Trauer, T.** (2009). Objections to routine clinical outcomes measurement in mental health services: any evidence so far? Unpublished paper.
- Meehan, T., McCombes, S., Hatzipetrou, L. and Catchpoole, R.** (2006). Introduction of routine outcome measures: staff reactions and issues for consideration. *Journal of Psychiatric Mental Health Nursing*, 13, 581–587.
- National Centre for Health Outcomes Development** (2009). *Concepts and Frameworks: Definitions*. Available at: <http://www.nchod.nhs.uk>.
- Nightingale, F.** (1863). *Notes on Hospitals*, 3rd edn. London: Longman, Green, Longman, Roberts, and Green.
- NIMH Early Clinical Drug Evaluation** (1976). Clinical global impressions. In W. Guy (ed.), *ECDEU Assessment Manual for Psychopharmacology*, revised edition (pp. 217–222). Washington, D.C: U.S. Department of Health and Human Services Public Health Service, Alcohol Drug Abuse and Mental Health Administration, NIMH Psychopharmacology Research Branch.
- Ohio State Mental Health Services** (2008). *Ohio Mental Health Datamart*. Available at: <http://outcomesdatamart.mh.state.oh.us/Screen1/odmhFirstScreen.jsp>.
- Oxman, A. D., Sackett, D. L., Chalmers, I. and Prescott, T. E.** (2005). A surrealistic mega-analysis of redisorganization theories. *Journal of the Royal Society of Medicine*, 98, 563–568.
- Phelan, M. et al.** (1995). The Camberwell Assessment of Need: the validity and reliability of an instrument to assess

the needs of people with severe mental illness. *British Journal of Psychiatry*, 167, 589–595.

- Reynolds, T. et al.** (2000). Camberwell Assessment of Need for the Elderly (CANE): development, validity and reliability. *British Journal of Psychiatry*, 176, 444–453.
- Royal College of Psychiatrists Research Unit** (2008). *HoNOS65+ Glossary*. Available at: <http://www.rcpsych.ac.uk/clinicalservicestandards/honos/olderadults.aspx>.
- Slade, M. et al.** (2006). Use of standardised outcome measures in adult mental health services: randomised controlled trial. *British Journal of Psychiatry*, 189, 330–336.
- Smith, D.** (2008). Creating buy-in into initiatives in outcome measurement: an implementation methodology. *Australian and New Zealand Journal of Psychiatry*, 42 (Suppl. 4), A9.
- Spear, J., Chawla, S., O'Reilly, M. and Rock, D.** (2002). Does the HoNOS 65+ meet the criteria for a clinical outcome indicator for mental health services for older people? *International Journal of Geriatric Psychiatry*, 17, 226–230.
- Stewart, M.** (2008). Making the HoNOS(CA) clinically useful: a strategy for making the HoNOS, HoNOSCA, and HoNOS65+ useful to the clinical team. In *Proceedings of the 2nd Australasian Mental Health Outcomes Conference*. Available at: <http://amhoc2008.com.au/program.php>.
- Stewart, M.** (2009). Service user and significant other versions of the Health of the Nation Outcome Scales. *Australasian Psychiatry*, 17, 156–163.
- Trauer, T.** (2004). Consumer and service determinants of completion of a consumer self-rating outcome measure. *Australasian Psychiatry*, 12, 48–54.
- Turner, S.** (2004). Are the Health of the Nation Outcome Scales (HoNOS) useful for measuring outcomes in older people's mental health services? *Aging and Mental Health*, 8, 387–396.
- Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B., Hadden, S. and Burns, A.** (1998). Health of the Nation Outcome Scales (HoNOS): research and development. *British Journal of Psychiatry*, 172, 11–18.