

The effect of guidance in the use of routine outcome measures in clinical meetings

Tom Trauer, Glenda Pedwell and Lisa Gill

Abstract

Routine outcome measures (ROM) have now been introduced into all Australian public mental health services, but experience suggests that many clinical staff lack expertise in using them. In 2005 under the Quality Through Outcomes Network (QUATRO), the Victorian Department of Human Services set up three teams aimed at consolidating the use of ROM and furthering sustainability through a variety of peer-support activities. We report on an initiative undertaken by one of these teams. QUATRO team members attended team meetings of four adult community teams (three metropolitan and one rural) fortnightly over about 3 months. QUATRO staff contributed to discussion of outcome measures during routine clinical review, using local and national outcome measurement data and tools, and their own expertise. Attitudes of clinicians toward ROM in general and the specific instruments were assessed at the beginning and end of the period, and again after about 5 months. Qualitative findings consist of observations of factors that assist and hinder use of ROM. The initiative identified steps that staff can take to make ROM more useful in their work.

Aust Health Rev 2009; 33(1): 144–151

Tom Trauer, BA(Hons), PhD, ABPsS, MAPsS, Professor Department of Psychiatry, University of Melbourne, Melbourne, VIC.

Glenda Pedwell, BachOT, GradDipPubPol, Formerly Project Manager, Quatro North East Victoria Department of Human Services, Melbourne, VIC.

Lisa Gill, BAppSci(OT), GradCertHealth ServMgmt, Manager, Clinical Innovation and Interdisciplinary Projects Caulfield Hospital, Division of Rehabilitation, Aged and Community Care, Alfred Health, Melbourne, VIC.

Correspondence: Professor Tom Trauer, Department of Psychiatry, University of Melbourne, St Vincent's Mental Health Service, St Vincent's Hospital, 41 Victoria Parade, Fitzroy, Melbourne, VIC 3065.

Tom.Trauer@med.monash.edu.au

What is known about the topic?

Routine outcome measurement (ROM) is mandated in public mental health services in Australia, but early experience shows that attitudes are mixed. This may be due in part to clinicians being unfamiliar with how to use and interpret ROM data.

What does this paper add?

This project, aimed specifically at improving clinicians' understanding and confidence in using ROM in their regular weekly clinical meetings, found significant positive change in attitudes toward outcome measures. Most staff preferred information presented in graphical format over text or numbers and most respondents considered the consumer self-rating measure to be more useful than the two clinician-completed measures.

What are the implications?

Services should invest in suitable technology to make ROM easier for clinicians. Senior clinical and management staff can promote ROM through personal example. Training in ROM needs to pay as much attention to practical utility in the workplace as to theory and principles.

ROUTINE OUTCOME MEASUREMENT (ROM), comprising both provider- and consumer-completed instruments, was implemented in all public mental health services in Australia in 2003. In adult services the instruments used are the Health of the Nation Outcome Scales (HoNOS),¹ the sixteen-item version of the Life Skills Profile (LSP-16),² and in Victoria, the Behaviour and Symptom Identification Scale (BASIS-32).³ (In some other states and territories the Kessler 10 or the Mental Health Inventory were chosen to be the adult consumer-rated measure). The first two of these are completed by clinicians and the third is intended to be offered to consumers for voluntary, but encouraged, completion. In preparation for implementation, the majority of mental health clinicians in Victoria were trained in rating and collecting the suite of outcome measures. Signifi-

cant changes to service systems and processes were also implemented to enable completion of the measures and entry of the data collected, into a statewide database, whence they are aggregated at a national level.

In early 2005, the Australian Mental Health Outcomes and Classification Network (AMHOCN) produced its first round of tabular reports, presenting aggregate scores on measures across the country. An online utility, the Decision Support Tool, that enabled comparison of individual consumer ratings with national averages (for the HoNOS family of measures initially) became available in 2006. Clinicians typically were unaware of the existence of, and how to access, this resource.

Since implementation, outcome measures have been collected but with varying frequency and varying compliance with the national collection protocol. A survey of leaders within services 2 years after implementation in Victoria showed that use of outcome measure data for any purpose was extremely rare despite the resources committed to collection.⁴ Lack of access to suitable reports was thought to be a contributing factor. As well, while clinicians had been trained in collection of the measures, there had been no training in how to understand reports and results. Previous research found that provision of technical support enhanced the use of data by clinicians.⁵ The QUATRO North East project team wanted to learn more about whether the provision of further guidance to mental health clinicians and team leaders improved their understanding of what ROM results mean and of how these results could be used to inform clinical practice. Of particular interest would be the information pertaining to changes in a consumer's health over time and at different collection occasions.

The potential uses for outcomes assessment data occur at two levels, the aggregate data level and the individual clinical utility level;⁶ the latter of these is the primary focus in this study. The project team set out to trial new forms of feedback within clinical review meetings to determine whether there was any effect on staff perception

and clinical use of outcome measurement data. The intent was to integrate discussion of outcome measures into existing team meetings.

Methods

Expressions of interest were sought from mental health service teams in the North East of Victoria, resulting in four adult community case management continuing care teams (one rural and three metropolitan) being chosen. Staff members had various levels of training in the collection of outcome measures. Some of the teams had been involved in pilot implementation of outcome measures in 2000⁷ and others began collection with the national roll-out of ROM in 2003.

Initial meetings were held between the project team and the four team managers and consultant psychiatrists (where they were available). The project team then attended routine clinical review meetings in the four continuing care teams on a fortnightly basis from April to July 2006. A minimum of two members of the project team, at least one of whom was an experienced clinician, attended each meeting. In most meetings, standard ROM reports of consumers being reviewed were projected onto a wall using a data projector. Where possible, the same results were entered into the Decision Support Tool and similarly displayed. The actions of the project team related to ROM were guided by the comments, requests and questions raised by the clinical team members.

The initiative was assessed both quantitatively and qualitatively. The quantitative aspect consisted of asking team members to complete a questionnaire both before (Time 1), at the end (Time 2) of the 4-month trial, and again 5 months later (Time 3). The questionnaire was adapted from one used in an earlier study that evaluated mental health services using ROM⁸ and covered personal details, training status, attitudes to OM, preferences for feedback (text/graphics/numbers), usefulness of the measures, and ease of use of the measures (questionnaire available from the first author on request). The qualitative aspect con-

sisted of taking field notes after each clinical meeting on occurrences that were considered to be of interest, important or relevant. In order to promote as much open discussion as possible, assurance was given to members of the participating teams that comments and questionnaire responses made by individual staff would not be individually identified.

Results

Participation

Sixty-one staff from the four agencies returned a questionnaire at some point: 49 at Time 1, 46 at Time 2, and 41 at Time 3. The distribution across the four teams is shown in Box 1.

Of the 61 respondents there were 28 nurses, 12 doctors, 9 social workers, 7 psychologists, and 5 occupational therapists. Fifty-six indicated how many years they had been a mental health professional. The mean was 12.7 years (SD, 10.8; range, 1 to 42 years). Forty-two respondents (69%) indicated that they had had training in the rating of outcome measures and 19 (31%) that they hadn't. Seven respondents (12%) indicated that they had attended an advanced outcomes measurement training session and 54 (88%) that they hadn't. The mean interval between the two completions for the 39 who completed at both Time 1 and Time 2 was 117 days (about 17 weeks; range, 86 to 153 days), the interval for the 32 who completed at both Time 2 and Time 3 was 152 days (about 22 weeks; range, 112 to 195 days), and for the 33 who completed at both Time 1 and Time 3 was 264 days (about 38 weeks; range, 112 to 312 days).

Attitudes to outcome measurement

On the first page of the questionnaire, there were nine statements relating to attitudes to outcome measurement generally. In the interest of parsimony we explored whether these nine responses could be summarised into a single score. Cronbach's alpha, a measure of internal consistency, at Times 1, 2 and 3 was 0.68, 0.71 and 0.67, indicating that there is a reasonable

I Numbers of staff who completed the questionnaire

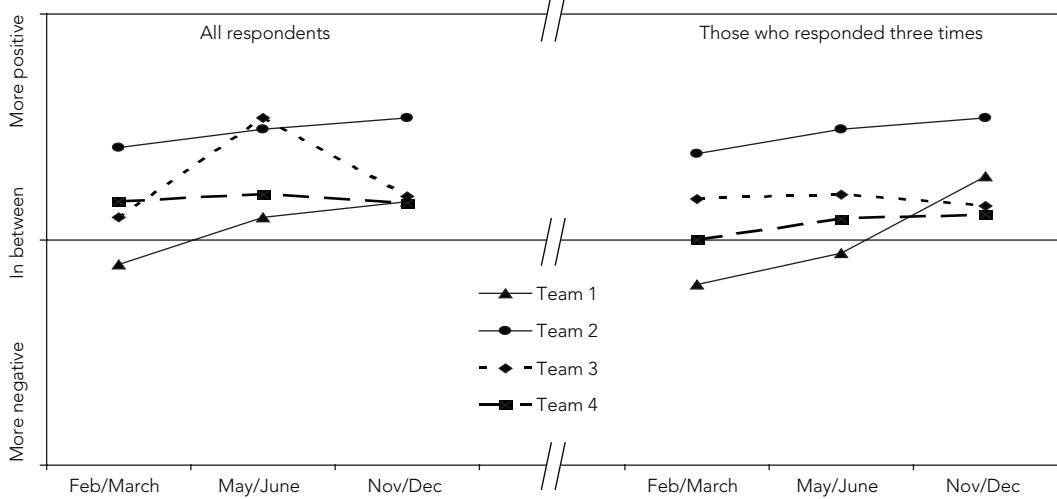
	Team				All
	1	2	3	4	
Time 1 only	1	1	1	3	6
Time 2 only	1	0	1	3	5
Time 3 only	0	2	1	2	5
Times 1 and 2	2	1	2	4	9
Times 1 and 3	1	1	0	2	4
Times 2 and 3	1	0	0	1	2
All three times	4	13	6	7	30

degree of consistency in responses to the 9 items at all time points. Therefore, responses that were positive to ROM were coded as 2, don't know as 1, and negative to ROM as 0, and these scores were averaged across the nine statements at each time point. Box 2 shows the means of all respondents and of the thirty who responded at all three time points by time point and team.

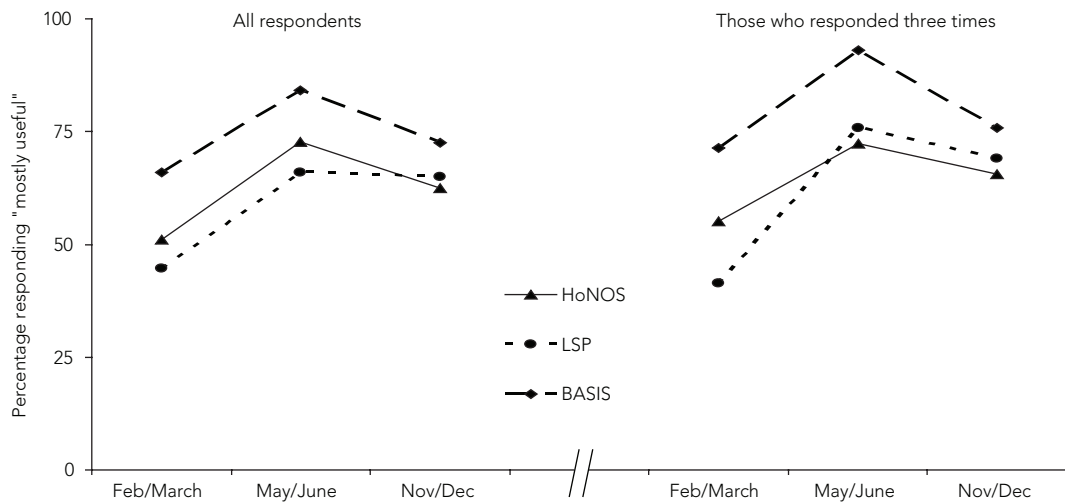
Box 2 shows that there were small improvements in attitude across the three time points in teams 1 and 2 whether one looks at all respondents or only those who provided responses on all three occasions. In team 4 there was no overall change when comparing each separate time cohort, but there was a small improvement in those who returned three questionnaires. Team 3 shows an apparent improvement between times 1 and 2, but this must have been due to the composition of respondents at time 2, because the results of those who returned three questionnaires show no change. Differences between the four teams were significant at time 1 ($F_{(3,45)} = 4.00$; $P = 0.01$) and time 3 ($F_{(3,37)} = 3.64$; $P = 0.02$) and nearly significant at time 2 ($F_{(3,42)} = 2.55$; $P = 0.07$).

A limitation of these comparisons is that different respondents contributed to the team averages at each time point. An approach that accommodates missing data on repeated measures is the mixed model.⁹ A mixed model regression, using standard software¹⁰ was per-

2 Mean attitude scores, by team and time



3 Mean ratings of usefulness, by time



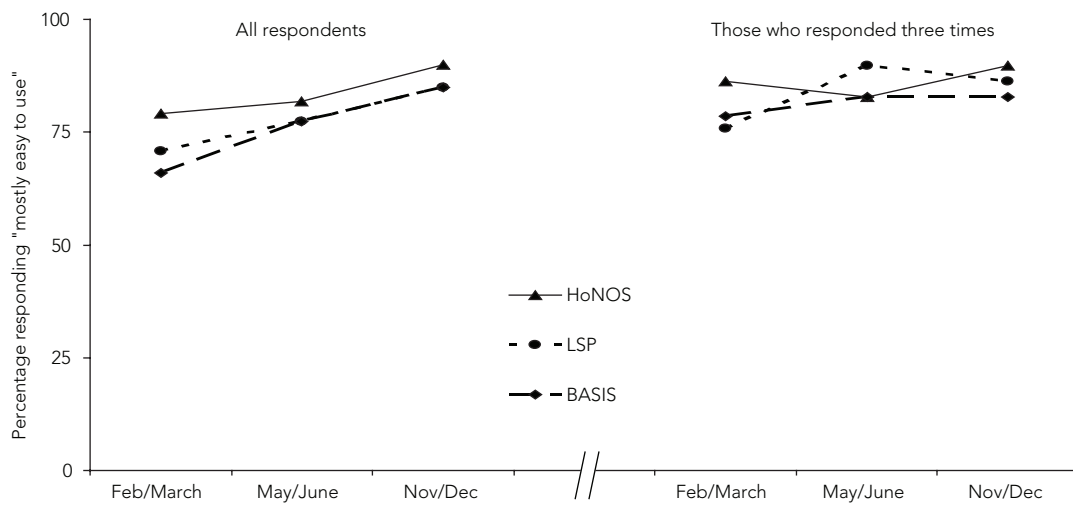
HoNOS = Health of the Nation Outcome Scales. LSP = Life Skills Profile. BASIS = Behaviour and Symptom Identification Scale.

formed. Attitude score was the dependent variable, the three time points was the repeated factor, and respondents and teams were set as random effects, with respondents nested within teams. This showed a near significant positive effect of time ($P = 0.07$) and significant variation between respondents ($P < 0.001$) and between teams ($P = 0.03$).

Perceived usefulness and ease of use of the measures

Respondents were asked whether they didn't like, liked a bit, or liked a lot, outcome measures reported as numbers, graphs or text. The percentages of respondents at each time point who indicated that they liked numbers "a lot" were 31.2%, 18.2%, and 35.0%. The corresponding

4 Mean ratings of ease of use, by time



percentages for graphs were 64.6%, 72.1% and 67.5%, and for text 27.7%, 23.3%, and 35.1%. It is clear that there was a consistent preference for graphs over text and numbers at all three time points.

Respondents were asked how useful they thought the HoNOS, LSP-16 and BASIS-32 were, using the responses of "mostly not useful", "don't know/unsure", and "mostly useful". Box 3 shows the mean responses at each time point.

This shows that the BASIS-32 was typically seen as the most useful instrument, followed about equally by the HoNOS and LSP-16. We performed three mixed model regressions, one for each instrument, with a similar specification to that used for the attitude data. These showed that only for the LSP-16 was there a significant improvement over time ($P=0.05$), there were significant ($P=0.03$) differences between teams on both the LSP-16 and BASIS-32, and there were significant ($P \leq 0.01$) differences between respondents on all three instruments.

Respondents were asked how easy to use they thought the HoNOS, LSP-16 and BASIS-32 were, using the responses of "mostly not easy to use", "don't know/unsure", and "mostly easy to use". Box 4 shows the mean responses at each time point.

The majority of respondents rated all three instruments as mostly easy to use at all three time points. We again performed three mixed model regressions, one for each instrument, with similar specifications to those used for the attitude and usefulness data. These showed that for the LSP-16 there was a significant improvement over time ($P=0.05$), and near significant for the BASIS-32 ($P=0.07$). For no instrument were the differences between teams significant, and for all three the differences between respondents were highly significant ($P<0.001$).

Qualitative observations

We identified certain elements that appeared to facilitate the use of OMs. One of these was the capacity to compare ratings of a consumer with national aggregate data through the Decision Support Tool (only possible for HoNOS during this project). Another was the ability of OMs to direct attention to aspects of consumers' lives that were not always routinely addressed (eg, social needs). In one consumer, very gradual deterioration had been occurring and comparison with previous ratings provoked review of medication and increased contact. On several occasions ROM scores were rated higher (worse) than clinicians had expected. In one of these cases, where staff

expressed strong feelings concerning a consumer, the team came to notice the high level of sympathy and disability.

Our experience also identified factors that represented barriers to the productive use of OMs. One of these was lack of access to data projection facilities. Initially in some teams, paper copies of reports were used for clinical discussion until computer network access to reports became available. Use of paper reports was found to be inefficient and confusing as staff took time to determine which report was being discussed. The project team reached the conclusion that consideration of graphical information by a team was best done with a data projector showing graphs in the course of the meetings. This method meant that everyone was "on the same page" when looking at the graphs. Some clinicians were initially reluctant to set up and use a data projector. One service had a ceiling-mounted data projector and this option provided the greatest ease of use. Typically, only one or two members in any one team were confident to access reports electronically and use the data projectors. One team had excessive delay in accessing reports due to network system issues, and the rural service experienced occasional periods where connection to the computer server dropped out.

Another factor which was sometimes a barrier, but at other times not, was the way the clinical meetings were organised and led. In one team, use of ROM seemed to assist with streamlining and structuring the clinical review process. Leadership by team managers, team leaders and consultant psychiatrists also impacted significantly on other team members' interest in outcome measures.

Discussions in clinical review meetings revealed that consumers had varying rates of having been offered and of completing the BASIS-32, ranging from some who had never completed a BASIS-32 to others who had completed several. Our impression was that where there were no BASIS-32s completed, this was more often a reflection of the culture of the service and clinicians rather than the unwillingness of the consumer. Privacy and consent issues concerning the

BASIS-32 were raised and discussed on several occasions in several teams, as was the idea of using the BASIS-32 with family and carers. One service had a number of BASIS-32s completed when consumers were inpatients, and these had been used to inform the development of the inpatient unit group program. Some clinicians had used various innovative strategies to offer and assist consumers with completing the BASIS-32. These included: mailing them out to consumers with a cover letter (one consumer had completed a BASIS-32 by email), working through several initial questions with the consumer and then allowing the consumer to complete the remainder, completing the BASIS-32 through a structured interview process (on one occasion this had been done over the telephone), and completing the BASIS 32 with the consumer over several sessions where concentration was poor. There were several occasions where ratings on the BASIS 32 revealed different problems to those rated by the clinician. For example, on two occasions, consumers rated themselves as having difficulty with drug and alcohol issues, however the clinician had not rated any problem in this area.

Depending on which instruments had been completed, we found that outcome measure ratings were not that useful for some consumers and very useful for others. In general, the more complete the set of measures, the more useful they were. As already mentioned, the presence of the consumer self-rating measure added an often unexpected dimension; similarly, the availability of previous assessments allowed teams to focus on the consumer's progress in a systematic fashion. This was particularly the case when some staff were comparatively recent to the team and did not know the earlier progress of the consumer, and when changes in consumers over extended periods had been gradual.

Discussion

In this study, new forms of feedback on ratings for outcome measures were trialled within community mental health clinical review meetings to

determine whether there was any effect on staff perception and clinical use of outcome measurement data. Sixty-one staff from four adult community mental health services completed questionnaires on their attitudes toward outcome measures. As well, they allowed the project team to be part of their clinical review processes for 3 months and to make notes on the meetings. ROM reports available on individual consumers were presented and explained in order to discover more about which reports and what information clinicians found useful.

Judging by answers to the attitude statements in the questionnaire, there was a significant change in attitudes toward outcome measures in a positive direction at the end of the 3 months, most of which was maintained at follow-up. These findings are consistent with previous research findings where the provision of technical support led to more positive attitudes toward outcome measures among clinicians.⁵ Questionnaire results also revealed that most staff preferred information presented in graphical format over text or numbers. The fact that many of the changes were maintained, and even in some cases extended, at follow-up, suggests that the intervention was not simply effective while it lasted. This is in contrast to the study by Close-Goedjen and Saunders⁵ who found that the positive effects of feedback dissipated once the support was withdrawn.

One clear finding was that most respondents considered the consumer self-rating measure to be more useful than the two clinician-completed measures. This echoes the finding of Callaly et al.¹¹ It is ironic, therefore, that, in most adult settings, the self-rated measure was typically not completed.

Qualitatively, the project team noted that there were many variables in each of the teams that could influence the extent to which teams use outcome measures. Along with factors such as team size, experience, discipline mix and previous training, additional factors such as team systems and processes for clinical review (eg, in one team there were occasions when only 2–3 staff attended), interest or lack of it by consultant

psychiatrists in using outcome measures, supervision of clinicians, and other changes in the organisation affecting staff were all thought to contribute to attitudes toward use of outcome measures.

Observations from the project team were that many staff were interested to learn more about what the outcome measure ratings meant and were interested in comparing ratings for their consumers to the national averages through the use of the recently developed Decision Support Tool. When exposed to this tool, clinicians then expressed interest in being able to compare more demographic features of consumers with matched averages.

We noted that it took a number of sessions for staff to become familiar with the reports, and repetition of material was necessary due to staff absences. Team members seemed to respond more positively when the most senior doctor present and/or the service manager was interested and engaged. Features and effects of leadership in the use of ROM would be an instructive area for further inquiry. Interestingly, sometimes the most junior staff had the greatest interest in ROM.

Ease of access to and timely production of reports were essential for engaging clinicians and maintaining their interest. Use of a data projector for discussion of the results in meetings was pivotal to ensuring that everyone was “on the same page” during discussions. The presence of staff confident with information technology and using data projectors was also thought to affect team perceptions of ease of use.

We conclude by acknowledging certain limitations of this study. First, it is important to recognise that attitudinal surveys usually reveal what respondents think and believe, and not necessarily what they do. In a study of this kind we were unable to determine whether clinician practices changed in any way. Secondly, the researchers were known to be supporters of routine ROM, and there is the possibility that some of the positive findings were a form of social desirability response. Thirdly, the number of teams, and the numbers of staff in each team at each time point

were quite small, and we cannot be sure how robust our results are.

Further studies in this area, notably on effects of leadership, effects of access to reports, education about the meaning and correct interpretation of OM scores, and longer term studies to ensure that effects of interventions are enduring would be recommended.

Acknowledgements

The project team would also like to acknowledge the work of Peta Slattery and Rowan Chipchase in their roles as project officers with QUATRO North East throughout parts of this project. We are also appreciative of the goodwill of management and staff of the clinical services that took part in this study.

Competing interests

Tom Trauer is Chair of the Australian Adult Mental Health Outcomes Expert Group. Glenda Pedwell was Team Leader, QUATRO, Department of Human Services, Victoria.

References

- 1 Wing JK, Beevor AS, Curtis RH, et al. Health of the Nation Outcome Scales (HoNOS). *Brit J Psychiatry* 1998; 172: 11-18.
- 2 Buckingham W, Burgess P, Solomon S, et al. Developing a casemix classification for mental health services. Canberra: Commonwealth Department of Health and Family Services, 1998.
- 3 Eisen SV. Behavior and Symptom Identification Scale (BASIS-32). In: Sederer LI, Dickey B, eds. Outcomes assessment in clinical practice. Williams & Wilkins, 1996.
- 4 Pedwell G, Slattery P, Gill L. Quatro North East Progress Report 2006: March.
- 5 Close-Goedjen JL, Saunders SM. The effect of technical support on clinician attitudes toward an outcome assessment instrument. *J Behav Health Serv Res* 2002; 29: 99-108.
- 6 Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: what's the use? *J Behav Health Serv Res* 2003; 30: 393-405.
- 7 Eagar K, Buckingham B, Coombs T, et al. Final report on the implementation of outcome measurement in adult area mental health services in Victoria. Wollongong: Centre for Health Service Development, University of Wollongong, 2001.
- 8 Trauer T, Callaly T, Herrman H. Attitudes of mental health staff to routine outcome measurement. *J Mental Health* (in press).
- 9 Brown H, Prescott R. Applied mixed models in medicine. Chichester: Wiley, 1999.
- 10 StataCorp. Stata Statistical Software, release 9. College Station, Texas: 2005.
- 11 Callaly T, Hyland M, Coombs T, et al. Routine outcome measurement in public mental health — results of a clinician survey. *Aust Health Rev* 2006; 30: 164-73. Available at: http://www.aushealthreview.com.au/publications/articles/issues/ahr_30_2_0506/ahr_30_2_164.asp (accessed Nov 2008).

(Received 14/01/08, revised 23/05/08, accepted 21/09/08) □