

Analysis of Outcome Measurement Data from the Four Victorian 'Round One' Agencies

January 2003

A Report prepared by

Associate Professor Tom Trauer

Department of Psychiatry, University of Melbourne

Mental Health Research Institute

St Vincent's Hospital, Melbourne

for

Mental Health Branch

Department of Human Services, Victoria

EXECUTIVE SUMMARY

Four agencies (St Vincent's, Barwon, Grampians, and Maroondah) have been collecting routine outcome measures since training in mid-2000. Health of the Nation Outcome Scales (HoNOS), Life Skills Profile (LSP-16), Behaviour and Symptom Identification Scale (BASIS-32) and Focus of Care (FoC) assessments were entered into specially designed software. The Mental Health Branch commissioned this project to analyze the accumulated data in order to provide feedback to the agencies, identify trends, and discover what can be learned about the instruments themselves and about the process of routine outcome measurement, particularly as it might inform the training and implementation in the remaining agencies in the State.

Over 6,000 consumers were assessed across the four agencies. The diagnostic composition of the group seemed broadly representative of the public adult mental health services. Nearly 23,000 separate administrations of the instruments were recorded, over half of which were HoNOSs. There were sizable differences between agencies in how many assessments they completed.

Looking at collection rates over time, there were differences between inpatient and community settings both between and within agencies. One agency recorded no inpatient assessments whatever, another recorded few in the early stages but built up over time, another showed a reverse trend, and a fourth recorded few assessments on admission but a moderate number on discharge. Differences between agencies in monthly rates of collection of outcome measures in the community were less marked. The trend in monthly rates of completion suggested that in some settings early enthusiasm was overtaken by subsequent inactivity.

Certain items of the instruments were more prone to being omitted than others. The most frequently omitted item of the HoNOS (item 8) is particularly complex and would be a candidate for revision. The omission rate on the LSP-16 was lower than on the HoNOS, suggesting that it may be generally easier for clinicians to complete. Over 82% of consumers responding to the BASIS-32 omitted no items, and most items with potentially objectionable or sensitive content (sex, alcohol, drugs) were not especially prone to omission.

The internal consistency of an instrument is often taken as a measure of its reliability. The internal consistency of the HoNOS is quite low, that of the BASIS-32 quite high, and the LSP-16 is intermediate. This suggests that the HoNOS is not a measure of a single underlying construct, and operates more like a checklist of weakly related features. On the other hand, the BASIS-32 is highly internally consistent, indicating that most consumers responded to its items in a consistent fashion, and did not discriminate much between items.

Average scores on the clinician-rated instruments (HoNOS, LSP-16, and FoC) varied according to sex, age, diagnosis, context of assessment, and discipline of the clinician. On the HoNOS, scores from men, older consumers, consumers with organic and substance abuse disorders, admission and intake contexts, and assessments conducted by medical staff, tended to be high. However, there were complex interactions between these factors, rendering the effects of each factor conditional upon some or all of the others. Similar associations were found for the LSP-16, except that personality disorder and learning disability were associated with high scores, assessments at community intake were not, and there was a different pattern of clinician discipline associations. Age and sex differences were apparent with the FoC. There were some unexpected findings

on the FoC that call into question whether the instrument was always being used as intended.

There were also clear age and sex differences on the BASIS-32, and a tendency for consumers with non-psychotic disorders to assess themselves as worse than those with psychotic disorders. Completion of a BASIS-32 was associated with lower age, lower levels of severity and disability, and with non-organic psychotic diagnoses. It would clearly be desirable to know if non-completion was because the form was not offered, or if it was offered and refused. A recommendation that should facilitate this has been made.

The correlations between the instruments were examined. There was a strong correlation between the HoNOS and the LSP-16. Consumers assigned to the Maintenance FoC category obtained the lowest HoNOS scores, but the LSP-16 scores were more discriminative between the FoC categories. As expected, the relationships of the BASIS-32 and the HoNOS and LSP-16 were much weaker, confirming the conclusion that the self-assessment measure constitutes a quite separate form of information. Consumers assigned the Acute and Functional Gain FoC categories obtained higher BASIS-32 scores than those assigned the Intensive Extended and Maintenance categories, which may mean that there is greater expectation of improvement in those cases where subjective distress is higher.

There are two pairs of contexts, admission/discharge in inpatient settings and intake/closure in community settings, where we would expect to see changes for the better on outcome measures. Large improvements, especially on items reflecting depressive problems, were apparent on the HoNOS between admission and discharge. The few instances of actual worsening occurred with very short (less than 9 days) admissions. Differences between intake and closure, while smaller, were still significant, and were also greater on those items related to depression.

HoNOS and LSP-16 retests at review in the community were analyzed with a view to deriving thresholds that could be regarded as "significant". That is, we asked the question: How large does the difference in total scores have to be in order to be judged unlikely to have arisen by chance? It was found that it depended in part upon whether the retest was conducted by the same or a different clinician from the one who had conducted the initial assessment; the difference needed to be slightly greater to be judged significant if the retest was conducted by a different clinician. This suggests that it is advisable to routinely record the identity of the clinician completing a measure. Provisional minimum differences were calculated, but these should be regarded as tentative.

A few graphical examples were presented of how changes on available assessments can be inspected to elucidate changes or trends over time. While not always showing clear or consistent patterns, the capacity to look at results in this way affords an opportunity to discern trends that may be present.

The disciplines of staff completing the measures were examined and considerable differences between agencies were found. Apart from in a consultation-liaison service, medical staff rarely completed instruments, and there were large difference between nursing and allied health staff between certain agencies. It appears that the outcome measurement workload is shared differently in different agencies, reflecting different customs and practices.

A set of recommendations is presented at the end of this report.

CONCLUSIONS AND RECOMMENDATIONS

Throughout this report the findings have been summarized in “Comment” sections at the end of each main section, and they are recapitulated in the Executive Summary, so we will not reproduce them again here. Rather, we will extract those conclusions that lend themselves to recommendations.

1. Agencies should review the numbers of assessment completed with reference to the numbers obtained in the other agencies. St Vincent's should review its process for obtaining or recording self-assessments on the BASIS-32.
2. Certain items on certain instruments were subject to relatively high rates of omission. Training of new staff and refresher training of existing staff may need to pay special attention to these items.
3. The LSP-16 was less prone to item omission than the HoNOS, and appears to be an “easier” test for clinicians to use than the HoNOS or FoC. Compared to the balance of time devoted to the different instruments in the Round I training, consideration should be given to reducing the amount of time allocated to the LSP-16.
4. Scores on instruments were typically found to be associated with several factors (e.g. diagnosis, setting). This means that the interpretation of individual assessments needs to be conducted cautiously, with due regard to the complex interplay of factors.
5. Although the FoC is not an outcome measure in the conventional meaning of the term, for the purposes of data handling it should be treated as if it were.
6. The FoC produced several anomalous findings. The way this measure is presented and explained in training needs to be reviewed.
7. It was found that consumers who completed the BASIS-32 were on average younger, less ill, and less disabled than those who did not. The possibility arises that opportunities for self-assessment were not as equally distributed as they should have been. Agencies should review their standard documentation with a view to ensuring that clinicians indicate (a) whether or not the BASIS-32 was offered, (b) if it was not offered, why not, and (c) if it was offered but not completed, what was the reason. RAPID/CMI should have provision for the entering of this information. Regular audit of this information should maximize appropriate self-assessment.
8. In judging the “significance” of retest assessment, it was found that differences needed to be larger when a different clinician completed the second assessment. This translated into different criteria for judging significance. For this reason, assessments conducted by clinicians should be required to include some form of clinician identifier, and this should be a mandatory field in RAPID/CMI.
9. It was found that plotting OM scores on a horizontal time axis allowed trends over time (where present) to be readily discerned. Consideration should be given to making such graphical displays available as standard reports in RAPID/CMI.

10. In looking at which discipline groups completed the clinician measures, large differences were found between agencies. In general, medical staff rarely completed outcome measures. Agencies should review their policies and practices in this regard. Agencies yet to commence routine OM also need to anticipate the role of their medical staff in forthcoming training and beyond.
11. Standard reports available in HBL and RAPID/CMI do not provide distributional information on ad hoc groupings (e.g. means and medians of HoNOS scores of consumers with a given diagnosis assessed in a given setting). Such reports should assist clinicians in interpreting the scores they obtain on individual consumers. Consideration should be given to making such reports available as ad hoc queries.