Mental Health Information Development

National Outcomes and Casemix Collection *Discussion Paper*

The Health of the Nation Outcomes Scales (HoNOS), General Adult Version:

Towards an agenda for future development

Version 1.0
April 2006

Prepared by Tom Trauer and Bill Buckingham on behalf of the Adult Mental Health Outcomes Expert Group

national mental health strategy

FINAL DRAFT FOR PUBLIC RELEASE

Document information

Title: The Health of the Nation Outcomes Scales (HoNOS), General Adult Version: Towards an agenda for future development
Version: 1.0
File: HoNOS Modifications_v1.0.doc

Document history

Version	Date	Details
Version 0.1	04/3/2003	Distributed for comment to working subgroup of Adult Outcomes Expert Group
Version 0.2	11/3/2003	Distributed to Adult Mental Health Outcomes Expert Group for discussion at meeting of 19/3/2003
Version 0.3	24/5/2004	Distributed to reconvened Adult Mental Health Outcomes Expert Group for discussion at meeting of 8/6/2004. Amendments include: • Title: Previous subtitle of paper ('What revisions are required?') amended to 'Towards an agenda for future development'. • Page 4: Second caveat added to acknowledge the compromises inherent to all brief measurement instruments. • Page 12: Options for changes to Item 8 expanded. • Pages 13-15: Item 10 - substantial rework to address the relationship between basic and complex skills when rating ADLs. Cross tabulated rating guide added. • Page 18: Footnote added to "'Exceptions to the two core HoNOS 'rules'" section to acknowledge useful aspect of HoNOSCA. • Page 19-21: Significant rework of content on 'Cultural and contextual factors', incorporating recent work in Northern Queensland to develop annotated glossary for use with indigenous consumers. Recommendations to amend the general HoNOS instructions modified. • Page 24: Section 6.6 ('Weighting scores for the overall severity index') added. • Page 25: Additional examples given of inconsistencies between members of the 'HoNOS family' • Pages 27-30: Section on 'Options for the future' amended to distinguish potential changes that affect the HoNOS as whole from those applying only to individual items; more clearly separate minor and major modifications; and· provide a summary of potential item-specific changes.
Version 0.4	20/7/2004	Amended following comments from Adult, Child & Adolescent and Older Persons' Mental Health Outcomes Expert Groups. Amendments include: • General removal of reference to New Zealand experience. • Page 13: Clarification added that private sector training continues to use the 'rate the worst' principle for all items.
Version 0.5	17/11/2004	Minor revisions as suggested by New Zealand Ministry, details re Authors added.
Version 1.0	April 2006	Document renamed as 'Final Draft for public release'. Amendment to footnote 5, page 20.

Acknowledgements

Preparation of this paper was commissioned and funded by the Australian Government Department of Health and Ageing upon the advice of the Adult Mental Health Outcomes Expert Group and the recommendation of the National Mental Health Information Strategy Committee. While it focuses specifically on the HoNOS instrument, it forms part a broader set of activities designed to inform discussions on a range of mental health information development and outcome measurement issues.

The views expressed in the paper are those of the authors and do not necessarily reflect the opinions of the Australian Government or any particular State or Territory

About the Authors

Tom Trauer currently holds a number of academic positions in Australian universities. He has more than 35 years experience in the mental health field, including public sector, private practice, medico-legal, and research and evaluation. Trained as a clinical psychologist, his initial work was in mental health services in Melbourne, followed by nearly 20 years in London, where he worked at Guy's hospital, subsequently was head of the psychology department at King's College hospital, and was heavily engaged in the training of clinical psychologists. On his return to Australia in 1989 he worked for several years as head of the psychology department of a psychiatric hospital in Melbourne, following which he took up an academic position in the Department of Psychological Medicine, Monash University. From the early 1990s he has been mainly occupied in teaching (clinical psychology trainees and psychiatric registrars), training (mainly mental health staff in the use of outcome measures), research (primarily quantitative methods, with a speciality in outcome measurement), evaluation studies, supervision of doctoral students, and a variety of national and international consultancies. In recent years he has acted as specialist adviser on outcome measurement at both the State and Commonwealth level, as well as leading and co-leading research teams. He is a regular contributor to mental health conferences and has over 100 publications in the mental health field.

Bill Buckingham is the principal of his consulting company Buckingham & Associates. He has more than 30 years experience in the mental health field, which includes 12 years as a clinical psychologist in the Victorian mental health services and 5 years as the state's Chief Clinical Psychologist. Between 1991-1993, he managed the statewide service planning unit of the former Victorian Office of Psychiatric Services. Since establishing the company in 1993, Bill has been a key figure in the development of projects under the National Mental Health Strategy. From the outset of the Strategy, he has been engaged as a consultant to the Australian Government Department of Health & Ageing and the National Mental Health Information Strategy Committee, designing and guiding much of the information development work program. He has designed and written or co-authored all seven National Mental Health Reports released since 1994. Throughout 1997, he was the consultant to the group commissioned by the Australian Health Ministers' Advisory Council to conduct an independent evaluation of the first five years of national mental health reform. Bill was also the Principal Consultant for the MH-CASC Project, and the New Zealand-Classification and Outcomes Study. More recently, Bill has been working with the Australian Government and states and territories in the implementation of outcome measures and development of new performance measurement frameworks for mental health services.

.

Contents

1. Overview and purpose

This paper reviews experience with the HoNOS (general adult version 4, 1996) with particular attention to summarising the problems encountered in its use within Australia.[1] The purpose is to identify those aspects of the HoNOS which would benefit from revision in future versions.

The paper has been prepared as a discussion document for consideration by the Adult Mental Health Outcomes Expert Group. It is recognized that there is a diversity of views about the nature and extent of possible amendments to the HoNOS. While this document identifies possible ways forward, the views presented should be considered only as contributions to the debate rather than definitive recommendations.

2. Background

It is now eight years since the current version of HoNOS was produced by the United Kingdom Royal College of Psychiatrists (College Research Unit 1996). Considerable experience with the instrument has been accrued and a wide range of research literature produced on the HoNOS family of measures. Australian 'pilots' of the instrument commenced with trials in Victoria in 1996 (Trauer et al. 1999) and continued with its use in a large-scale casemix study involving 18,000 consumers (Buckingham et al. 1998). Subsequently, all States and Territories agreed in 1999 to the introduction of the HoNOS and HoNOSCA as core components of a suite of measures selected for routine implementation in public mental health services (Commonwealth Department of Health and Aged Care 1999). In a parallel development, Australia's private hospitals have also introduced the HoNOS as a standard for the assessment of consumer outcomes (Morris-Yates & The Strategic Planning Group for Private Psychiatric Services Data Collection and Analysis Working Group 2000). In New Zealand, the HoNOS and HoNOSCA were used throughout 2001 by approximately 65% of mental health services in an extension of Australia's earlier casemix study (Gaines et al. 2001) and were followed by a decision by the Ministry of Health that the instruments would be introduced by publicly funded mental health services commencing July 2004.

Collectively, the data accumulated on the performance of the instrument covers approximately 100,000 applications and an estimated 40,000 consumers. This has entailed the instrument being used in increasingly diverse settings, including the public and private health sectors, with culturally and linguistically diverse groups, and in various specialized teams, such as consultation-liaison psychiatric services. Some of these applications would hardly have been envisaged when the HoNOS was first developed, so it is not surprising that a number of problems with it have become apparent.

As part of the commitment to the ongoing development of outcome measures in mental health services it is timely to document and review this experience, with

[1] It is expected that future versions of this paper will incorporate a New Zealand perspective, based on experience in implementing the HoNOS in the cultural context for Maori as part of a major study on outcomes and casemix (See Gaines et al 2003).

particular attention to aspects of the instrument that would benefit from revision. This review is occurring in parallel with similar undertakings in the United Kingdom where the HoNOS is in the process of being implemented as a component of the national minimum data set for mental health services.

3. Method and scope

This paper draws on two lines of evidence.

1. The experiences of those individuals who have been closely involved in the implementation of routine outcome measurement, including research, service evaluation, policy development, and training; and
2. Data on the performance of the HoNOS in various settings, compiled from several sources as summarised in Table 1.

Table 1: Main data sources used for the current report

Source	Description	Number of HoNOS observations
MH-CASC Project, (Buckingham et al. 1998)	Large scale service utilisation and costing study conducted in 1996, involving 25% of Australia's public and private sector mental health services.	About 60,000 assessments of 18,000 patients
Victoria HoNOS field trial, (Trauer et al. 1999)	In a project auspiced by the Victorian Mental Health Branch, five adult mental health services trialled the HoNOS between April and August 1996	About 3,500 assessments of over 2,000 patients
Victorian Phase 1 implementation (Trauer 2003)	In mid-2000, four Victorian adult mental health services began routine outcome measurement. This report analysed the data collected over the first 19 months	Nearly 15,000 assessments of over 6,000 patients

Some of the data we use has been presented previously in the form of journal articles, internal reports, and conference presentations.

The material is organized under the following headings:

- Review of the 12 individual HoNOS items
- Issues affecting scoring and interpretation
    - Scoring guidelines
    - Exceptions to the two core HoNOS 'rules'
    - Cultural and contextual factors
    - The HoNOS collection protocol
    - The computation of summary scales when some ratings are missing
    - Weighting items to derive an overall severity score
- Other issues
    - Relationships within the HoNOS 'family'
    - Training issues
- Options for the future development of the HoNOS

4. Caveats

Three caveats should be kept in mind when reading this paper.

The first concerns the utility of the HoNOS as a broad spectrum tool for standardised outcomes reporting. The investments being made by Australia in the routine collection and reporting of consumer outcomes are founded in part on the documented utility of standardised measurement scales for mental health services including the HoNOS in particular. Because this paper focuses on areas where the HoNOS would benefit from revision, it necessarily addresses problematic aspects of the instrument. These should not be taken as implying the instrument in its current form is unsuitable for routine implementation. The extensive literature and clinical experience in fact points to the contrary conclusion and show the HoNOS to be a valid, reliable and useful summary measure of changes in consumers' health and functioning as they progress through treatment.

The second caveat is that, like all summary measurement instruments, the HoNOS represents a compromise between complexity and practicality. In twelve individual items, and requiring only five minutes of the clinician's time, the instrument attempts to provide a summary picture that covers a wide spectrum of clinical and social phenomena and is relevant across all adult mental health consumer groups. Additionally, it is intended to 'survive' within the busy daily routines of public mental health services and be capable of being used by all mental health professions with minimal training. To achieve this, imperfect solutions to complex issues are inevitable. The question to consider when reviewing the suggestions outlined in the current paper is whether those suggestions might create more problems than they solve, dislodging the instrument from its original purpose. It is possible that the current balance is the right one.

The final caveat concerns the process for taking the current HoNOS to a new version. As discussed below, the HoNOS is a copyrighted instrument released into the public domain for broad use, with copyright held by the United Kingdom government. While undesirable for the purpose of achieving national and international comparability, no restrictions are placed on users from modifying the instrument to suit local requirements. In fact, since the appearance of the current version of HoNOS, a number of variants for special groups have been developed. Australia may decide, for example, to introduce specific changes that may not be agreeable to the original designers. Conversely, Australia is not bound by any changes made elsewhere. However, this paper is premised on the principle that parallel, independent development of the HoNOS by multiple parties will detract from its long term value and ultimately compromise the interests of all parties who use it. To this end, the paper has been prepared with a view to promoting collaborative international development with the HoNOS designers. Based on discussions with the UK College of Psychiatrists, similar papers can be expected from other countries as steps towards building consensus in priority areas for change.

5. Review of the 12 individual HoNOS items

In this section we examine each of the twelve items in turn, summarising the issues and suggesting possible improvements. Three 'tests' are used to describe the performance of each item.

- *Inter-rater reliability* – the performance of each item is summarised, drawing on the sources summarised in Table 1 along with other data available from published studies. Summary data are presented in Appendix 1.
- *Missing ratings* – similarly, relevant data are presented on the extent to which each item is vulnerable to missing ratings. This information is summarised in Appendix 2.
- *Rating criteria* – comments are offered on the adequacy of the glossary rating criteria based on our experience as trainers and from observations made in the field.

It is acknowledged that the HoNOS items may be tested against additional criteria (e.g., sensitivity to change or discrimination between disorders) but these are beyond the scope of the current paper. Further work would consider these aspects.

In relation to the missing ratings test, it is important to note that the standard scoring domain does not distinguish between missing data and ratings of '9' where the clinician specifically indicates 'not known or not applicable'. Recent work in Australia (Commonwealth Department of Health and Ageing 2002) recognises that ratings may be missing for any of the following reasons:

(a) The clinician may be unable to rate due to insufficient information; or
(b) The specific item is treated as not applicable because the data collection protocol does not require it; or
(c) The item is simply not scored.

For the purposes of the current paper, we have treated '9's' (covering a and b above) as equivalent to missing ratings (c). The need to fully specify the HoNOS scoring domain to allow genuine missing data to be distinguished is taken up later in the paper.

Item 1: Overactive, aggressive, disruptive or agitated behaviour

Inter-rater reliability:

- Inter-rater reliability estimates on this item range from acceptable to good (see Appendix 1).

Missing ratings:

- Missing data was found to occur in less than 0.5% of records in inpatient and community settings in the two Victorian studies and just under 1% in the MH-CASC study (Appendix 2).

Rating criteria:

- The issue most often raised with Item 1 concerns its heterogeneity of content. Apart from item 8, the scale covers more clinical ground than all other items. Our experience is that raters typically consider only the aggressive component, and often overlook the other behavioural elements that Item 1 addresses. The option of separating aggressive, antisocial or disruptive behaviours from other rated categories, as occurs in the HoNOSCA, is worth considering in any major rebuild of the instrument. In the absence of a major rebuild, revision of the current item glossary is desirable. In its present form, examples are given at all levels from 1 to 4 of aggressive behaviour but not for overactive, disruptive or agitated behaviour. Suitable examples could be devised and included.

- Item 1 presents the first instance of several where contextual and cultural factors need to be taken into account in order to make a sensible rating. Contextual and cultural issues are raised reasonably frequently. For example, indigenous groups in Australia have questioned whether culturally-prescribed actions that involve aggression to others (e.g., tribal payback) would warrant a rating. Related questions are also raised in how to score violent behaviour where it has occurred as a self defence against attack by others.

- Our usual response is that, as in all clinical judgements, culture and social context need be considered in determining whether a mental health problem is present and that, as a measure of the severity of mental disorder, the same considerations should be applied when making HoNOS ratings. The problem however lies within the item 1 glossary which states *"Include such behaviour due to any cause"*. This is also presented as a general principle at other points throughout the HoNOS glossary and in various training resources produced by the UK College.

- The original 'Guide for Trainers' prepared by the UK College (Wing et al. 1996) recognised the need to emphasise contextual issues and added the following additional note to Item 1: *"… the context must be considered since disagreement, for example, can be expressed more vigorously, but still acceptably, in some social contexts than in others"* (page 8). However, the Trainers' guide then continues to emphasise the overriding general principle: *"Possible causes of the behaviour are not considered in the rating and diagnosis is not taken into account."*(page 8)

- Although not reflected in the inter-rater reliability or missing item estimates, we believe this ambiguity compromises the utility of item 1 and points to an important issue that warrants inclusion in the HoNOS general scoring guidelines. The original purpose of the HoNOS was to serve as a measure of the severity of mental disorder and associated social problems within psychiatric populations, rather than a broad epidemiological screening tool that would be suitable for describing behaviour in the general population. Thus, positive scores on an item should reflect, first and foremost, the presence of a mental health problems rather than behaviours that can be considered normal when context and culture are taken into account. In this sense, cause (or hypotheses about such) are in fact highly relevant considerations when making a rating.

- This issue is relevant across most of the HoNOS scales and is discussed again in section 6.3 of this paper. The possible solutions are:

(a) Add a statement to the HoNOS general instructions that context and culture should be considered when rating all items.

(b) Amend the Item 1 glossary statement that currently reads "Include such behaviour due to any cause" to "include such behaviour that may be associated with any mental or behavioural disorder".

Item 2: Non-accidental self harm

Inter-rater reliability:

- Inter-rater reliabilities are generally good with the only weak result reported from a small study by Brooks which found lower than average reliabilities on most items (Appendix 1).

Missing ratings:

- The "missing" rate is under 1% in both inpatient and community settings in the two Victorian studies, and between 1% and 1.6% in the MH-CASC study.

Rating criteria:

- Comments about cultural interpretation made in relation to Item 1 are also very pertinent to this scale. Many examples exist of deliberate self-harm that is culturally endorsed during specific life stages (e.g., during the period of 'sorry time' that follows the death of a family member in some Australian Aboriginal cultures).

- Clarification would be helpful within the glossary that Item 2 is not a substitute for formal risk assessment as future risk is not considered. While this might be regarded as self evident given the general rule that ratings are based on the past 2 weeks, clinicians can confuse the concepts. It is worth noting that such clarification has been introduced to the most recent version (v3) of the HoNOS65+ glossary.

Item 3: Problem drinking and drug-taking

Inter-rater reliability:

- Inter-rater reliabilities are acceptable to good.

Missing ratings:

- The "missing" rate is around 0.5% to nearly 2%.

Rating criteria:

- An important source of uncertainty on this item is whether tobacco qualifies as a drug for the purpose of rating. This question is contentious and tends to polarise opinion during training sessions. For these reasons, it is desirable that the uncertainty be addressed directly in the glossary entry.
- In favour of regarding tobacco use as 'in scope' is that it is technically a drug as well as a significant health issue for a disproportionate number of people with mental disorders. Against inclusion is that, unlike other illegal drugs, it has little direct relevance to mental health problems and, in our view, is not of the same

order of clinical importance as the use of alcohol or illicit drugs. The fact that a high rating on Item 3 can be earned equally by a heavy smoker and a person who abuses alcohol or illicit drugs that compromise their mental health reduces the value of the item as a measure of co-morbid substance abuse in psychiatric populations. Exclusion of tobacco use in the glossary would improve the item's specificity.

- Complicating this is the fact that there are extreme situations where the pattern of tobacco use, or complications arising from addiction, are an important part of the clinical picture and would warrant a rating, e.g. setting fire to one's bed, begging for cigarettes, or intimidating behaviour toward others in order to extort cigarettes. In such instances, use of the HoNOS Item 3 to capture the problem and track the consumer's progress over time would be appropriate.

- It is worth considering the option of adding the following clarification to the glossary: that tobacco use should only be scored higher than 1 where it is *clinically significant* from a *mental* health assessment perspective, and warrants some form of targeted intervention for a mental health, rather than general health, reason. This is consistent with our recommendations discussed in section 6.1 concerning the need to make overt the overarching scoring principles applying to all HoNOS items.

Item 4: Cognitive problems

Inter-rater reliability:

- Estimates of inter-rater reliability vary greatly between very good and poor. Importantly, Shergill et al. (1999) who used the adult form of the HoNOS with 100 patients over 65 years old (average age 73 years) found good inter-rater reliability.

Missing ratings:

- The missing rate is mostly below 1%.

Rating criteria:

- Two issues compromise this item - firstly, the question of transient versus enduring cognitive impairments, and secondly, the position of formal thought disorder.

- *In relation to transient versus enduring cognitive impairments*, the standard glossary instructions read:

  ***Include** problems of memory, orientation and understanding associated with any disorder: learning disability, dementia, schizophrenia, etc. **Do not include** temporary problems (e.g. hangovers) resulting from drug/alcohol use, rated at Scale 3.*

- The inclusion of the caveat 'do not rate temporary problems', although tied to problems arising from drug or alcohol abuse, creates the problem of certain HoNOS rules working in opposition. On the one hand, the 'rate the worst' principle allows transient problems to be rated. This is reasonable as the capacity of the HoNOS to measure change would be limited if transient problems were

deemed out of scope. On the other hand, the exclusion of transient problems from consideration suggests the intent is to focus the assessment on relatively enduring cognitive impairments such as those arising from learning disability and dementia and not other transient or self-correcting conditions. In our experience, this creates uncertainty for raters about how to score a range of conditions that present with cognitive problems that are expected to be short-term in nature e.g., psychogenic memory loss in dissociative disorders – should this be rated against Item 4 or under Item 8?

- Rewording of the glossary to more clearly confine the caveat about temporary problems to alcohol and drug-induced problems would be beneficial, along the following lines:

  ***Include** problems of memory, orientation and understanding associated with any disorder: learning disability, dementia, schizophrenia, etc. **Do not include** drug or alcohol-induced problems of a transient nature (e.g. hangovers), rated at Scale 3. However, longer term cognitive problems resulting from drug and alcohol use should be rated here.*

- As an additional amendment, the glossary should acknowledge that functional cognitive impairments are in scope and include a few examples.
- *In relation to position of formal thought disorder,* clinicians often ask where such problems should be rated on the HoNOS. Given that approximately one third of patients in public mental health sector have a schizophrenia-spectrum disorder, an unambiguous response on this issue is essential. The approach taken in Australian training sessions is to advise that such problems be rated on Item 4 on the basis that formal thought disorder is a recognised form of cognitive impairment. The alternative is to assign thought disorder to Item 6 (Hallucinations and Delusions) but this would have the effect of the transforming that scale to a measure of psychotic symptoms. In our view this would be inconsistent with the design principles on which the HoNOS was developed.[2]

- At a minimum, the glossary should be explicit in how to rate thought disorder and include examples within the anchor point descriptors. It would also be desirable to caution raters that thought disorder and hallucinations/delusions (item 6) are quite different phenomena, and one may be present without the other.

Item 5: Physical illness and disability

Inter-rater reliability:

- Inter-rater reliability findings are satisfactory to good, with the exception of the Brooks study (Appendix 1).

Missing ratings:

- The missing rate is 1% or below.

[2] Section 7.1 discusses the problem of inconsistency emerging across the 'family' of HoNOS instruments in whether thought disorder is rated on scale 4 or 6.

Rating criteria:

- We see no particular problems with item 5. However, it would be useful to repeat in the glossary the important point that serious medical conditions need not earn a rating if they do not interfere with functioning or cause personal distress.

Item 6: Hallucinations and delusions

Inter-rater reliability:

- Inter-rater reliabilities are consistently good to excellent in all studies.

Missing ratings:

- Missing rates vary around the 1% mark.

Rating criteria:

- We see no problems with item 6, which appears to work well in the field. The only issue of concern is the inconsistency between the general HoNOS and other members of the HoNOS family in how bizarre behaviour associated with hallucinations and delusions is recorded. This issue is discussed in section 7.1.

Item 7: Depressed mood

Inter-rater reliability:

- Inter-rater reliabilities are satisfactory to very good across all studies.

Missing ratings:

- The missing rates are mostly 0.3% and 2%, but are about twice as common in acute inpatient settings compared to community settings.

Rating criteria:

- While this item performs satisfactorily, inconsistency in the glossary terminology causes some difficulties during training. Given that the measurement domain is "depressed mood" and not depression, references to "depression" in glossary entries 2, 3 and 4 should be replaced with "depressed mood". Similarly, a note that reminds the rater how to handle other symptoms of depression (e.g., rate somatic symptoms on Item 8) would be helpful.

Item 8: Other psychological symptoms

Inter-rater reliability:

- Only the Nottingham study cited in Wing et al. (1998) found an acceptable level of inter-rater reliability. In the other studies the reliability ranged from barely acceptable to poor. These reliability coefficients were based simply on the recorded numerical rating, without reference to the assigned letter codes. Thus, two ratings with the same number but different letters would be considered a perfect agreement. Therefore the obtained reliabilities should be considered upper bounds. If the letter codes were factored in, the effective agreements would undoubtedly be even lower.

Missing ratings:

- There is contrasting evidence concerning missing ratings. In the Victorian field trial data the rate is under 1%, but in the "Round One" data it ranges from 7% in assessments conducted in the community to 39% in assessments conducted in acute inpatient settings, and 21% and 9% in acute inpatient and community settings respectively in the MH-CASC study. In the two Victorian studies, these high missing rates are most likely an artefact of the way the HoNOS was used by clinicians. In situations where clinicians did not identify some other problem that might qualify for a rating on item 8, many simply left it blank rather than entering a zero as they should have done. This explanation does not apply in the MH-CASC data, since different codes were used for "not known/ not applicable" (rated as 9) and occasions when no rating was made (entered as 99).

Rating criteria:

- We believe that there are significant problems with item 8 as it currently stands. Apart from the unacceptably low reliability, there are the issues of comparability and interpretation. There is no easy or clinically sensible way to compare, say, a mild phobia with a moderate sleep problem. The code "J – other" is quite frequently used, with the clinician being required to write in what the problem is. Current information systems have no provision for the capture of this qualitative information, rendering the analysis and output of records containing J codes virtually impossible to interpret. The numeric and letter code ratings represent an attempt to simultaneously capture both quantitative and qualitative information. There is a risk of mismatch between the two. Appendix 3 presents item 8 data on the 14,897 HoNOS collected in the first 19 months by four Victorian "Round One" agencies. Three areas have been highlighted with bold borders. The left-most of these represents 2,228 HoNOS administrations (about 15%) in which a letter code was assigned but the numeric rating was zero, and the right-most area a further 269 administrations (about 2%) in which a letter code was assigned but the numeric rating was missing. The middle highlighted area shows a small number of assessments (40) where a non-zero rating was not accompanied by a valid letter code. Thus about 17% of HoNOS administrations, or one in six, had an invalid or incomplete combination of numeric ratings and letter codes. This is largely the fault of the software that allowed such data to be entered, but it draws attention to the need to ensure that clinicians need to understand the way item 8 is meant to work, and the need for computerized validation routines on data entry.

- We suspect that certain problems, principally B (anxiety) and D (stress), are excessively rated, by which we mean that vague, colloquial or non-clinical criteria are accepted. Some trainers have taken to instructing raters that conditions only qualify for rating on item 8 if they constitute separate disorders in their own right. Thus, one should not rate D (stress) on the basis of a consumer saying that he is "stressed", since the scale is intended for actual stress disorders, such as PTSD.

- A final problem to note is the somewhat arbitrary list of choices A to I. As shown in Appendix 3, certain choices, notably Dissociative, Somatoform, and Sexual are very infrequently used (about 1% of records) while more common 'other problems' are not listed. Clinicians frequently ask how, for example, negative

symptoms, or elated mood should be rated on the HoNOS. The only option available is to use code J (other).

- A number of non-exclusive changes to Item 8 could be considered to improve the HoNOS performance.

    1. Extend the list of options beyond the current nine problems.
    2. Add the guideline that a condition meets diagnostic criteria before it can be rated on item 8. Against this is the general instruction, valid for all other items, that diagnostic considerations be put aside when completing the HoNOS (the "any cause" or "irrespective of diagnosis" guidelines).
    3. The descriptive labels for problem categories A to I could be expanded in full on scoresheets to more clearly indicate that they refer to specific disorder categories rather than more casual interpretations. For example, the 'D – Stress' category would be substituted by 'D - Reactions to severely stressful events and traumas'. This in fact was the original approach taken in HoNOS training materials developed by the UK College.
    4. Disallow ratings of 9 (don't know / insufficient information) in order to reduce mismatches between ratings and letter codes, on the basis that if a clinician cannot think of an "other" problem to rate, logically item 8 is a zero.
    5. Information systems that receive HoNOS ratings could have more sophisticated data entry validation routines to prevent invalid rating/code combinations.
    6. On the basis that the HoNOS serves multiple purposes, including the summary profiling of the patient's current set of problems, allow clinicians to rate as many or as few of the item 8 alternatives as they choose (e.g., rate anxiety and sleep problems if they are both present) then derive the item 8 score to be the highest of the ratings. This aim here would be to avoid the frustration experienced by clinicians who have to choose which problem to rate when multiple are present, while also improving the capacity of the HoNOS to profile the patient's current clinical problems.

- A more radical approach would entail giving certain current choices separate items of their own. This would have the advantage of extending the coverage to better represent high prevalence conditions, particularly anxiety and phobic disorders, and improve the acceptability of the instrument to clinicians who work with these populations. However, it would lengthen the HoNOS beyond its current twelve items, and effectively create a new instrument.

Item 9: Relationship problems

Inter-rater reliability:

- Inter-rater reliability estimates vary; some are good, another fair, and some are poor (See Appendix 1).

Missing ratings:

- Missing rates range from below 1% to just under 4%, with the rate in acute inpatient settings about double that of community.

Rating criteria:

- Item 9 is the first of the scales where the global HoNOS rule of rating the worst manifestation of a problem in the rating period causes problems for raters. As with the other items in the Social subscale (Activities of daily living, Accommodation, and Occupation), the intention appears to be to capture relatively enduring aspects of the patient's social situation. To illustrate the difficulty for raters arising from the 'rate the worst' rule, compare an individual who is in a stable relationship with a good friendship network but who has experienced a recent, brief relationship crisis (e.g., blazing row with spouse) with another individual with ongoing relationship problems, such as the persistent isolation and loneliness of many people affected by mental illness. The glossary examples, with their emphasis on withdrawal, seem to suggest that it is such "negative symptoms" that are the focus of this item. However, applying the "worst in episode" rule, it is quite possible for the former situation to gain a higher score than the latter. The problem appears to be in defining just what constitutes a relationship problem, and in particularly whether the rater is to judge the "underlying" strength of the individual's relationships.

- The decision to be made is whether to rate according to such "underlying" strength of relationship(s). If so, the 'blazing row' example should be scored low (maybe 1). There are important implications for the sensitivity of the item, subscale, and total score as measures of change. In general, measuring enduring traits will diminish such sensitivity, whereas applying the "worst in episode" criterion will enhance it.

- The solution that has been adopted in Australian training within the public sector[3] is to suggest to clinicians that rating of all of the Social subscale items (9 to 12) should be on the basis of the "usual or typical" situation rather than the conventional "worst in episode" criterion. This is explored further in section 6.2.

- As an additional observation on Item 9, the glossary makes reference to active and passive withdrawal. Our experience has highlighted that this distinction may be unfamiliar to some clinicians, so the glossary could usefully be expanded with the inclusion of some examples of both.

Item 10: Problems with activities of daily living

Inter-rater reliability:

- Inter-rater reliability estimates vary between good and unacceptable. In fact, the same two studies (Orrell et al. (1999) and Brooks (2000)) found the lowest reliabilities on all the items comprising the Social subscale.

Missing ratings:

- The missing rate is quite low, ranging from under 0.5% to 3%.

[3] Current training in the private continues to follow the principle of 'rate the worst' for all 12 HoNOS items.

Rating criteria:

- It is clear from the glossary that there are two main elements involved in the scoring of this item – basic skills (e.g. dressing, toilet) and complex skills (e.g. budgeting, shopping). The main difficulty presented to raters is that performance in both of these areas needs to be *combined* in order to arrive at a single rating. This is in contrast to other HoNOS items that cover multiple elements (e.g., item 1) where the rater is instructed to score on the basis of the most severe element.

- The criteria accompanying each of the rating levels imply that complex skills are more fragile than basic skills, with the deficits in basic skills being the hallmarks of the highest ratings (3 and 4). This causes difficulty in public sector adult mental health services where the 'typical' patient has mild to severe problems in the area of complex skills but has relatively preserved basic skills. In these situations, strict adherence to the glossary would limit the score to a maximum of 2, thus reducing the capacity to monitor improvement.

- The problem can be distilled to the fact that the glossary gives little advice on how to combine the assessment of basic and complex skills into a single rating. This is a particular issue when they are divergent. There are also situations (e.g. in consultation-liaison psychiatry in medical inpatient settings) where the typical relativity between basic and complex is reversed – a medically ill patient may not be able to dress himself but may be able to manage his finances.

- While all HoNOS scales cover important domains, it is arguable that scale 10, with its focus on functioning, is especially critical, given that problems associated with psychiatric disability are a major focus of the care provided by specialist mental health services.[4] For this reason, several options are outlined below that are designed to clarify how to integrate judgements of basic and complex skills when ratings on item 10 are made.

  Option 1: Detailed map of the relationship between basic and complex skills

  This approach explicitly maps the relationship between the two, as shown in Table 2. The suggested ratings in the body of the table are based as far as possible on the glossary guidelines.

[4] In Australia, the standard assessment suite of measures complements the HoNOS with a dedicated measure of functioning, the short form of Life Skills Profile (LSP-16, Rosen et al. 1989; Buckingham et al. 1998).

Table 2: Cross-tabulation of ratings of basic and complex skills

		BASIC SKILLS				
		No problems	Minor	Mild	Moderate	Severe
COMPLEX SKILLS	No problems	0	1	2	3	4
	Minor	1	1	3	3	4
	Mild	2	2	3	3	4
	Moderate	2	2	3	3	4
	Severe	2	2	3	3	4

Option 2: Summary guide to relationship between basic and complex skills

One limitation of the cross-tabulation shown in Table 2 is that it may be excessively complicated for the busy clinician. Table 3 provides an alternative summary guide.

Table 3: Summary rating guide for item 10

	RATING				
	0	1	2	3	4
Basic skills	No problems	Nil or minor problems	Mild problems	Moderate problems	Severe problems
Complex skills	No problems	Minor problems	Mild, moderate or severe problems		

Option 3: Revise guidelines to allow an either-or approach to rating of basic and complex skills

The limitation of both options 1 and 2 is that each works within the current ‘rule’ implicit within the glossary - that higher scores are almost entirely ‘driven’ by problems in basic skills. As noted above, under these arrangements an individual with no more than minor problems in their basic skills could never be rated higher than 2, even if they had severe problems in their complex skills.

An alternative approach is to follow the guideline used in other HoNOS items that cover multiple elements and allow the rater to rate the most severe problems evident in either basic or complex skills

Option 4: Separate items for basic and complex skills

This is the more radical option and involves splitting item 10 into two separate items, one rating basic skills and the other rating complex skills. This would largely overcome the contortions required to combine the two into one rating.

However, in lengthening the instrument by one item, this would be a major change, and would require strong justification.

Item 11: Problems with living conditions

Inter-rater reliability:

- Acceptable inter-rater reliabilities were reported only by Wing et al. (1998) in the Nottingham and Manchester studies. All other studies have found poor to very poor reliability.

Missing ratings:

- Missing rates overall are around 1% to 2% in the two Victorian studies but much higher (around 7.5%) in the MH-CASC data. However this masks a significant disparity between acute inpatient and community settings, with the former missing rates being three to ten times those of the latter.

Rating criteria:

- Previous work has alluded to the conceptual complexity of items 11 and 12. For example:

  "The attribute to be rated is 'Opportunities for using and improving abilities, where the patient is living (item 11) or, occupational and recreational activities (item 12)'. This means that the same client should get different scores in areas where opportunities differ. This introduces uncertainty and ambiguity, since the ratings of these items is based upon the fit between the client's needs and the available opportunities. Whatever the reason, items 11 and 12 are problematic". (Trauer et al. 1999 p. 385-386).

  Subsequent work and experience has not suggested a different conclusion.

- The tendency for item 11 (and item 12) to pose difficulties in acute inpatient settings was recognized early. The UK College report on HoNOS research (College Research Unit 1996 p.81) indicated that in such settings items 11 and 12 might be omitted, and another study (Allan & McGonagle 1997) excluded both items from their analysis of long-stay inpatients. In the glossary, both items 11 and 12 carry the note to rate the patient's usual situation.

- It is likely that the poor reliabilities and high rates of missing ratings are a function of primarily acute inpatient unit staff not having sufficient knowledge of the patient's accommodation status prior to their admission. So long as raters are bound by the strict rule to only consider information relating to the last two weeks, these effects are predictable. The problem would be ameliorated by a relaxation of the two week rule for both items 11 and 12. Thus, in addition to extending the "usual or typical" variation to these items, a further variation is required to the effect that the rater may need to go back beyond two weeks to establish what is usual or typical. In fact, this is implicit within the glossary itself with the instruction *'If in acute ward, rate the home accommodation'*.

- The need to make more overt that item 11 (and 12) may require exceptions to the 'last two weeks' principle is taken up in section 6.2.

Item 12: Problems with occupation and activities

Inter-rater reliability:

- The inter-rater reliability problem with this item is serious - no published data have demonstrated a level better than 0.56.

Missing ratings:

- The overall missing rates are between 1% and 2%, but, as with item 11, much higher in the MH-CASC data (7.5%). In all three data sets, the rate is three to nine times greater in acute inpatient units than community settings.

Rating criteria:

- Comments made in relation to item 11 apply to item 12. Like item 11, item 12 also carries the instruction to *'rate the patient's usual'* situation but is more overt in its advice that this may require ratings to be made outside the two week period: *"If in acute ward, rate activities during period before admission'.*

- As discussed in section 6.2, clarification is required that makes more explicit that the "worst in episode" and "two weeks" rules need to be varied when rating item 12, as for item 11.

6. Issues affecting scoring and interpretation

6.1 Scoring guidelines

The 'simple HoNOS' is in reality a complex instrument. While the general rating rules and criteria presented at the beginning are helpful, these are qualified and elaborated in subtle ways by the glossary entries to each of the twelve items. We have found that, despite the criteria and glossary examples, raters sometimes have difficulty in deciding the most appropriate rating.

We believe that the instrument would benefit from greater elaboration of the scoring rules to give clearer guidance to clinicians. The current scoring system provides the following guide:

0 = no problem
1 = minor problem requiring no action
2 = mild problem but definitely present
3 = moderately severe problem
4 = severe to very severe problem

Implicit in these criteria is that ratings of 0 and 1 require no action (not clinically significant), but that ratings of 2, 3, and 4 do (clinically significant). This is an important organizing principle that we have found to be useful and intuitive to clinicians when it is made overt.

The question then arises: What constitutes 'action'? Broadly, this comprises monitoring and/or intervention. Lesser problems may be adequately dealt with by monitoring alone (with a view to intervention should they become more severe), while

greater problems will require actual intervention plus ongoing monitoring. In effect, these operational criteria serve to relate the HoNOS ratings to the clinical judgement of the rater.

We have found that presenting a tabulated summary of these clarifications of the rating criteria to be helpful. The table below, used in many Australian training sessions, would be a useful addition to the 'up front' section of the HoNOS.

Table 4: Clarification of the scoring criteria

				Monitor ?	**Active treatment or management plan ?**
Not clinically significant	**0**	No problem	Problem not present.	No	No
	1	Minor problem	Requires no formal action. May or may not be recorded in clinical file.	Maybe	No
Clinically significant	**2**	Mild problem	Warrants recording in clinical file. May or not be incorporated in care plan.	Yes	Maybe
	3	Moderate problem	Warrants recording in clinical file. Should be incorporated in care plan.	Yes	Yes
	4	Severe to very severe problem	Most severe category for patient's with this problem. Warrants recording in clinical file. Should be incorporated in care plan. *Note – patient can get worse.*	Yes	Yes

6.2 Exceptions to the two core HoNOS 'rules'

In our review of the twelve HoNOS items, we noted that the rating task would be made easier if variations to the 'core' rating rules that are mentioned in the glossary entries were presented more overtly.

The two core rules are:

- Rate the worst manifestation of the problem in the rating period, and
- Rate only problems that were present in the last two weeks.

Based on our examination of the twelve items, the qualification of these rules applies to the four 'social' items (Items 9-12). By their nature, these items relate to the usual

or typical situation and sometimes require the rater to look beyond the previous 2-week period.[5]

We have found it to be helpful during training, and in working with clinicians in a wide variety of contexts, to make overt these relatively implicit qualifications by putting then 'up front' as a clarification of the core rules. Thus, for items 9 to 12, the problems need to be rated according to the *usual or typical situation.* This is already noted for items 11 and 12 in the glossary ('Rate the patient's usual situation.') but should be extended to items 9 and 10 as well.

For items 11 and 12, especially in in-patient settings, the 'usual situation' often requires the rater to extend the period under examination back beyond the two weeks that apply to the other items. In fact, this was recognized by the instrument's developers at an early stage (Wing et al. 1996, p. 81).

These two clarifications are presented in tabular form in the following table. We have found that presentation in this form assists clinicians to make the necessary adjustments.

Table 5: Qualification of the two 'core' HoNOS rules

	CORE SCORING RULES	
	RATE THE WORST MANIFESTATION	**RATE OVER THE PAST TWO WEEKS**
Scales 1 to 8	Always	Always
Scales 9 and 10	Based on usual or typical	Always
Scales 11 and 12	Based on usual or typical	May need to go back beyond two weeks to establish the usual or typical

6.3 Cultural and contextual factors[6]

The importance of cultural and contextual factors was identified in the preceding discussion, particularly in relation to items 1 and 2 but affecting the HoNOS more globally. In general, the HoNOS instructions include the advice to rate the presence of the behaviour or problem in question *regardless of cause*. For example, the notes accompanying item 1 begin "Include such behaviour due to any cause (e.g. drugs, alcohol, dementia, psychosis, depression, etc.)". The examples imply that "any cause" is to be interpreted *diagnostically* (i.e. that the rating should be made irrespective of

[5] It is relevant to note here that the HoNOSCA is the only member of the HoNOS family that recognised the need to vary the 'rate the worst' rule in relation to social items. For example, the Trainer's Guide (September 1995 – March 1997, Gowers et al 1997) included the following statement: *"For items 1-9, the worst problem occurring during the chosen period is rated to give a measure of 'present state'. The rater should not attempt to rate each item as an average over the period. Items 10 to 13, however, require a more general rating over the chosen period' (our emphasis).* This caveat was not, however, included in the final glossary published in the British Journal of Psychiatry in 1999.

[6] As noted earlier, future versions of this paper are expected to incorporate a New Zealand perspective, based on experience in implementing the HoNOS in the cultural context for Maori as part of a major study on outcomes and casemix (See Gaines et al 2003).

the type of psychiatric disorder) and serve to remind the rater that similar problems may arise in the context of widely differing disorders.

What the instructions do not make clear is how to handle behaviours that are understandable or common in certain cultural or subcultural contexts. Examples include culturally-sanctioned violence (item 1), self-harm associated with religious ceremonies or periods of mourning (item 2), paranormal experiences associated with cultural beliefs or events (item 6) or the expression of "normal" sadness associated with bereavement (item 7). To the extent that these examples are not *prima facie* evidence, nor consequences, of mental health problems, it can be argued that they should not be rated on an instrument designed to capture the breadth and severity of psychiatric disorder.

Three broad options are available to address the issue.

The first option entails no action, on the assumption that such issues should be self-evident to trained clinicians who, as part of their daily work, are required to distinguish culture from illness. This might be described as the 'optimistic view' that common sense will prevail in guiding consistent use of the HoNOS.

The second option takes the view that separating cultural contributions from mental health issues in HoNOS ratings is best handled through training and does not require any adjustment to the current instructions or item glossaries.

The third option is to make amendments to the HoNOS instruction set to clarify that behaviours, experiences, and aspects of the person's circumstances that are culturally-based should not be included when making ratings.

Based on our experience, we strongly favour this third approach for several reasons. There is considerable evidence that cultural competence amongst mental health clinicians cannot be assumed. Further, inaction promotes the view that the HoNOS is not appropriate for specific groups of consumers due to its lack of attention to the cultural aspects of measurement. And finally, good practice in the development of any measurement instrument entails making the implicit explicit to remove known or foreseeable sources of ambiguity.

A lead in this area is given by the *'Measuring Mental Health Outcomes for Adult Indigenous Consumers'* project being undertaken in Far North Queensland through a collaborative initiative involving local Aboriginal-controlled councils, academic researchers and mental health providers. The project is aimed at evaluating the validity and reliability of two instruments – the HoNOS and LSP-16 – as mental health outcomes measures for adult indigenous consumers. As a first step, additional guidelines have been developed for the HoNOS glossary to assist clinicians in rating a range of circumstances where cultural considerations are fundamental. For example, the guidelines include the following principles:

- *" … that scoring of behaviours that are socially and culturally unacceptable should not be influenced by how common such behaviours are in the community. That is, scoring should objectively reflect behaviours not sanctioned or accepted even if they are common …"; and*

- *" … socially and culturally acceptable behaviours, experiences and beliefs associated with funerals, religious or traditional activities should not be included in any assessment items. Therefore you must identify whether the reported / observed findings are consistent with social or cultural practices that are recognised and accepted within the community. If your discussions with a family member/carer and the local practitioner indicate that the behaviours are socially and culturally acceptable, they should not be included in the scoring."*

The enhanced version of the HoNOS glossary and notes is being trialled throughout 2004 by mainstream mental health services working with indigenous consumers in Far North Queensland and results will inform the extent to which further changes re needed.

In summary, we believe there is a case to amend both the general instructions and relevant item-specific glossary notes to clarify that clinicians must consider culture and context when making ratings. The clarification could take the following form:

1. A reminder that the primary purpose of the instrument is to provide a summary measure of the severity of mental and behavioural disorders and associated problems.
2. As such, scores greater than zero on the majority of HoNOS items[7] should be used to indicate, first and foremost, the degree of severity of a mental or behavioural disorder rather than behaviours that can be considered normal when context and culture are taken into account.

6.4 The HoNOS collection protocol

Developing a clear protocol to guide the collection and reporting of the HoNOS is as important as the psychometric properties of the instrument if the data are to be used for comparative purposes across agencies and countries. A collection protocol should specify:

- when the instrument is to be applied;
- the observation period on which ratings are based; and
- who should use the instrument.

The HoNOS instructions cover the second and third aspects in a rudimentary way but are silent on the issue of when the ratings should be conducted. This is relatively straightforward for brief and discrete episodes of care, such as acute inpatient admissions, where a protocol would provide for collection at admission and discharge. However, even in such episodes there is a need to clarify how discharge ratings are made when the episode is very brief. There is greater ambiguity in resolving when to apply the HoNOS in the more typical pattern of care of public sector mental health services where care is provided over protracted periods in community settings.

Australia has developed a national protocol that covers these issues and establishes a basis for comparative benchmarking (Commonwealth Department of Health and

[7] The reference to the 'majority of HoNOS items' is necessary as three items are rated independently of mental disorders – items 5 (Physical illness), 11 (Living conditions) and 12 (Occupation and activities).

Ageing 2002). The Australian protocol identifies three 'collection occasions' where the HoNOS (and other outcome measures) should be used at:

- Start of new episode of care; and
- End of an episode of care; and
- Review – defined as being three monthly intervals where people are in ongoing care, or earlier if clinically indicated.

Specification of a standard review interval is essential for comparability between services or studies in the provision of continuing care as typically occurs in the community. The requirement for 3-monthly reviews is tied to the Australian National Standards for Mental Health Services and is argued on 'best practice' principles rather than from purely psychometric considerations.

Two additional aspects of the Australian protocol are worth considering for incorporation in any revised instrument.

- **Acute inpatient episodes** – In Australia, the majority of acute inpatient episodes are completed within 11 days, less than the two-week default period over which the HoNOS is meant to be rated. Consequently, if a second discharge assessment is applied that follows the standard rating period, it has little scope to measure change because it would incorporate the patient's presentation at admission. For this reason, we have revised the rule by specifying that, in the case of acute inpatient episodes, the discharge assessment *should only cover the 72 hours prior to the discharge*.

- **'Assessment only' and other brief episodes** – very brief episodes raise special challenges from an outcome monitoring perspective. When clinicians are instructed to use the instrument 'at the beginning and end of care', the question is inevitably raised as to when to apply the HoNOS in such episodes. The question is significant – approximately 15% of discharges from Australia acute psychiatric units have a length of stay of less than 3 days and 20% of people seen in the community are seen only once or twice and referred elsewhere for care. The approach we have taken begins with the premise that the HoNOS is a useful standardised measure to apply at all assessments and that this is should be done regardless of the expected period of care. But where the episode is brief (defined as less than 3 days for inpatient episodes, definition yet to be agreed for community care) discharge ratings are not mandatory.

It is important to note that these recommendations for when HoNOS assessments should be conducted should be understood as minimal requirements, and are not intended to detract from clinicians using their judgement to conduct other, *ad hoc* assessments as they may consider appropriate from time to time.

6.5 The computation of summary scales when some ratings are missing

The official HoNOS documentation gives no guidance on how missing values should be treated when computing subscale and total scores.[8] There are two aspects to this issue:

[8] 'Missing items' includes items left blank and ratings of '9' for the purpose of this discussion.

- Establishing the maximum number of missing items allowable for total scores and subscale scores to be considered valid.
- Determining how to derive subscale and total scores when the number of missing items is within tolerable limits.

6.4.1 Limit for maximum missing items

We examined above the susceptibility of each item to attract a missing rating. Table 6 uses the same data sources and summarises the numbers and percentages of HoNOS assessments with zero, one, two, and more than two missing ratings.

The results from the three data sources are not particularly similar, with the fewest number of missing ratings per assessment in the Victorian field trial and the most in the MH-CASC study. However, no more than 4% of assessments had more than two items with missing ratings in any data set. We may therefore tentatively suggest that assessment with up to and including two missed items may be accepted, while those with three or more be rejected. In suggesting rejection of the latter it is not implied that the individual ratings that are present are invalid or lacking in utility, only that the assessment be regarded as unsuitable for the computation of subscale and total scores.

Table 6: Numbers and percentages of HoNOS assessments with missing ratings

N Missing items	Victorian HoNOS field trial, 1996[1]			Victorian "Round One" agencies, 2000-2002[2]			Australian MH-CASC study, 1996[3]		
	N	%	Cum %	N	%	Cum %	N	%	Cum %
0	2,035	95.3	95.3	13,010	87.3	87.3	36,699	78.6	78.6
1	83	3.9	99.2	1,540	10.3	97.7	5,647	12.1	90.7
2	14	0.7	99.8	143	1.0	98.6	2,473	5.3	96.0
3-11	4	0.2	100.0	204	1.4	100.0	1,870	4.0	100.0

1 Trauer et al. (1999) Victorian HoNOS field trial.
2 Trauer (2003)
3 Buckingham et al. (1998).

6.4.2 Calculating total and subscale values when data are missing

Without resorting to complex imputation procedures, there are two options for dealing with missing ratings:

1. treat the missing items as zeros, or
2. replace the missing items with the mean of the valid ratings.

Both approaches have their strengths and weaknesses. In favour of the "zero" option, are firstly, probable greater validity, since the "true" rating is more likely to be zero, on the basis that if there were a problem in that domain the rater would probably have known about it, and secondly, the subscale and total scores will continue to be whole numbers. In favour of the "replace" option is increased robustness of the subscale and total scores (because of the positive intercorrelations among the items), but a weakness is the introduction of awkward fractions. A further issue with the "replace"

option is whether one averages over the whole scale or just the subscale in which the missing ratings occur. That is, if item 1 is the only one attracting a missing rating, does one replace with the mean of all other (eleven) items, or the mean of just items 2 and 3, which comprise the Behaviour subscale?

The attractiveness of the "replace" option depends to a large degree on the extent to which the HoNOS items are intercorrelated, with high inter-item correlations justifying the method more than low inter-item correlations. There are four available sources of this information.

- Stedman et al. (1997 Table B3.17) alpha = .73 mean r = .18
- MH-CASC (Buckingham et al. 1998) alpha = .79 mean r = .23
- Victorian HoNOS field trial (Trauer 1999) alpha = .75 mean r = .20
- Victorian "Round One" agencies (Trauer 2003) alpha = .72 mean r = .18

The results of these four independent studies are quite consistent, with the mean inter-item correlations ranging from .18 to .23. While the alpha levels are respectable the mean inter-item correlations are low, which vitiates to some extent the basis for replacing missing items with the mean of the non-missing items. Trauer (1999) previously used this finding to argue that the HoNOS should not be regarded as measuring a single underlying continuous construct, which would also be an assumption implicit in the "replace" option.

Weighing up these considerations, we favour the "zero" option. An implication of this option, which is explained during training sessions, is that a missing rating will be handled as if is were a zero by the software that assessments are entered into, and that by rating an item as missing the clinician is in effect declaring that he/she is not aware of any problem in the domain covered by that item.

6.4.3 Domain specification to distinguish genuine missing data

This is a technical but important point in relation to missing data. As noted earlier (see page 5 section 5), the HoNOS scoring domain does not distinguish between genuinely missing data and ratings of '9' where the clinician specifically indicates 'not known or not applicable'. This is contrary to international conventions. For the purpose of coding the data within various software applications, future revision should separate '9's' from blank entries.

6.6 Weighting scores for the overall severity index

The HoNOS total score provides the highest level summary of overall severity. Summary scores of this type are common throughout clinical measurement instruments and serve a number of useful purposes

The approach taken to calculating total scores rests on the crude assumption that all items are of equal importance. This lacks both clinical credibility and is not supported by evidence from large scale service utilisation studies in both Australia and New Zealand (Buckingham et al, 1998; Gaines et al, 2003). Further work is needed on

development of item weightings to construct a weighted total score that reflects the differential contributions of each item. Weightings could be constructed for different purposes. For example, one set of weights could be developed empirically, based on the relative contributions of each item to service utilisation while another set might be designed to reflect consumer concerns about optimal outcomes.

7. Other issues

7.1 Relationships within the HoNOS 'family'

While the current paper focuses on the general adult version of the HoNOS, any consideration of future revision needs to also be mindful of developments elsewhere in the HoNOS 'family'.

Since the appearance of the current adult form of the HoNOS (Wing et al. 1998) a number of variations have been developed to accommodate the needs of other clinical populations. There are forms for children and adolescents (Gowers et al. 1999), the elderly (Burns et al. 1999) and a patient self-assessment form (College Research Unit 1996; Trauer & Callaly 2002), as well as forms for acquired brain injury (HoNOS-ABI), learning difficulties (HoNOS-LD), and mentally disordered offenders (HoNOS-MDO).

Inspection of the versions applicable to adult and aged populations indicates some loss of consistency and comparability among the versions. It is unclear whether some of the versions are measuring the same domains and constructs as the general HoNOS, and there is no guarantee that obtained severity levels are comparable between versions. For some versions, there is little psychometric or other supporting information publicly available.

We mention here several examples of lack of consistency to illustrate the point.

- The HoNOS-LD glossary gives the ratings 1 to 4 as 'mild', 'moderate', 'severe' and 'very severe' respectively (Roy et al. 2002) which differs from the convention used in the general HoNOS (Wing et al. 1999), and indeed the HoNOSCA (Gowers et al. 1999) and the HoNOS65+ (Burns et al. 1999).

- Unlike all other members of the HoNOS family, the HoNOS-LD explicitly assesses future risk. The developers characterize the HoNOS as providing " . . . professionals with a framework to measure risk and vulnerability" (Roy et al. 2002) and the glossary (Roy et al. 2002) advises the clinician to consider risk in the rating of the first two behavioural items. The HoNOS-LD concept of prospective risk contrasts with the consistent focus of all other HoNOS instruments on the immediately preceding period.

- The third example involves the nature of item 1 and the rating of bizarre behaviour on the HoNOS65+ (v3) and the HoNOS-MDO. Wing et al's original HoNOS (v4) clearly states that "odd and bizarre behaviour associated with hallucinations and delusions" be recorded under scale 6 ('problems associated with hallucinations and delusions'). This instruction was maintained in the original

HoNOS65+ glossary (Burns et al. 1999) but subsequently amended with the release of the tabulated glossary issued by the Royal College in 2002 (http://www.rcpsych.ac.uk/cru/honoscales/honos65/index.htm). The tabulated version defines item 1 simply as "Behavioural disturbance" and provides for bizarre behaviour (e.g. posturing) to be rated under this item. In a similar vein, the HoNOS65+ tabulated version also instructs raters to include thought disorder under Item 6.This example captures two forms of inconsistency: firstly between the HoNOS 65+ and the HoNOS, and secondly between a previous and a later version of the glossary within a single instrument.

- The HoNOS-MDO continues this variation, also instructing clinicians to rate bizarre behaviours and thought disorder on Item 6 with the following glossary entry: "Include odd and bizarre behaviour associated with hallucinations or delusions, *such as thought disorder*' (our emphasis).

- The HoNOS 65+ tabulated version (version 3) introduces the concept of patient compliance for rating item 12 (Problems with work and leisure activities). Specific instructions are given to include consideration of the degree of patient cooperation with daily activities when rating the item. This is in direct conflict with the general adult version instruction that this item (along with item 11) be based solely on an assessment of the environment rather than the patient's functioning.

While additions to the HoNOS family may be welcomed as evidence of the attractiveness of the HoNOS as a prototype, the variations emerging on fundamental scoring rules undermine the value of the instruments. It is important to note that the inconsistency between versions is becoming a significant issue as many clinicians are required to work with more than one version of the 'family'.

Any revision to the general adult HoNOS should therefore be approached as a broader exercise that establishes the fundamentals across all members of the family. Included here should be the requirement for adequate development work and independent trialing of any version 'badged' under the HoNOS logo.

7.2. Training issues

Another source of potential inconsistency is in the way HoNOS users are trained. In the main, training methods and materials in Australia have been developed separately from those in the UK. A set of training materials, including video vignettes, were developed for the implementation of routine outcome measurement in Victorian public mental health services in 2000 (Eagar et al. 2000), and, independently, a graded set of paper vignettes was developed for use in private mental health services (Morris-Yates et al. 1999). Both of these resources are predicated on a basic training session of approximately half a day; this contrasts with the approach of the Royal College, which offers an initial HoNOS training package that lasts one day, and refresher sessions of half a day.

Thus there does not exist at the present time such a thing as a standard HoNOS training package. All of the existing packages and associated resources have their advantages and disadvantages, and it might be contended that having a range of

training resources facilitates adapting training for specific needs. On the other hand, the drawback of this diversity is that one cannot be sure to what level any particular clinician has been trained. Revision of the HoNOS should take the opportunity to introduce a degree of standardisation into training.

In many cases, the training that HoNOS users received was some time ago. The UK Royal College recommends refresher training every two years, but in Australia, there is no agreement on either the need or the optimum frequency of refresher training. Ideally, refresher training should use different materials from initial training, not least in order to maintain trainees' interest. A project to develop such materials will soon be undertaken by the recently created Australian Mental health Outcomes and Classification Network. New vignettes will be developed, with special attention to common problems and mistakes in HoNOS scoring.

8. Options for the future

This paper identifies a number of possible modifications to the way the HoNOS is presented and used. In resolving how and whether to take these forward we need to consider the scale of change that is acceptable. On the one hand, minor changes could be made relatively easily, which might involve such things as clarification of the rating rules and improvements to the glossary. On the other hand, major changes could be pursued, such as the addition or deletion of items.

The pros and cons of each are obvious: minor changes may leave certain problems uncorrected but retain the essential nature of the current version of HoNOS, while major change has the potential to correct more problems, but may lead to the *de facto* creation of a new instrument.

Major changes will affect certain parties differently. Australian mental health services are now well advanced in the introduction of routine outcome measurement, which has required the development of training materials, information systems, and documentation to support various scales, including HoNOS. Major changes are likely to be more disruptive to services and programs that are well advanced compared to those that are at an early stage in implementation. The creation of a new instrument may also:

- render the accumulated knowledge base less relevant for future use;
- necessitate retraining of staff;
- require the re-establishment of basic psychometric properties; and
- require retooling of information systems.

Clearly, the magnitude of changes is a difficult but important question that will need to be handled judiciously. As a general statement, we think that the benefits of major change will need to clear and great in order to justify them.

Against this background, we summarize the possible changes, beginning with minor changes, followed by potential substantial amendments of a more substantial nature.

- Table 7 presents minor changes that apply to the HoNOS as a whole. In general, many of these changes have been, by necessity, adopted informally throughout Australia and are reflected in various resource manuals and training materials, but not in the version of the HoNOS itself. Acceptance of these would primarily formalise these recent developments by incorporating the amendments to relevant sections of the HoNOS general instructions and various glossary items.

- Table 8 lists minor changes at the individual scale level. Again, most of these have been included in various training materials used by Australian states and territories but have not been incorporated formally in the HoNOS glossary.

- Table 9 summarizes the possible major changes.

Table 7: Possible minor changes for the HoNOS as a whole

1. Clarifications and enhancements to general instructions a. Revise the "due to any cause" advice to reflect specifically mental health causes and effects. b. Introduce the concept of clinical significance to assist raters in deciding rating level (see Table 3). c. Amend the "rate worst in period" rule to "usual or typical" for items 9 to 12 (see Table 4) d. Relax the "two week" rule for items 11 and 12 for acute inpatient and similar settings (see Table 4). e. Provide general instructions as to how deal with cultural and contextual factors
2. Computation of summary scores when some scale scores are missing a. For the scoring of subscale and total scores, treat missing data as zeros. b. Disallow computation of subscale and total scores for assessment with more than two scales with missing values.
3. Minimize inconsistencies between versions of the HoNOS a. Introduce a system of official recognition of HoNOS versions b. Resolve such inconsistencies that have been identified to date (see Section 7.1)

Table 8: Minor changes at the individual scale level

Scale 1:	Include examples of overactive, disruptive and agitated behaviour at all rating levels of 1 and above.
Scale 2:	Make explicit that it is self-harm behaviour in the previous period that is being assessed, and not prospective risk.
Scale 3:	Provide advice about how to handle tobacco use.
Scale 4:	Clarify that functional impairments and thought disorder are in scope for this item.
Scale 5:	No specific suggestions.
Scale 6:	If thought disorder is to be rated at Scale 4, then it should not be rated again on Scale 6.
Scale 7.	For consistency, replace glossary entries of "depression" with "depressed mood".
Scale 8.	Add a diagnostic threshold criterion; disallow ratings of 9.
Scale 9.	Clarify the distinction between active and passive withdrawal.
Scale 10.	Use a tabular presentation to clarify the relationship between basic and complex skills for each of point of the score range (see Table 2 and Table 3).
Scale 11.	Amend rating rules as per 1 c. and 1 d. in Table 6.
Scale 12.	Amend rating rules as per 1 c. and 1 d. in Table 6.

Table 9: Potential major changes

a.	Separate the composite domain represented by Scale 1 into a number of discrete scales.
b.	Revise the list of choices for Scale 8.
c.	Create new items for some of the more prevalent choices for Scale 8.
d.	Split the rating of basic and complex skills for rating ADLs (item 10) into separate items.
e.	Develop a tabulated version along the lines of the HoNOS65+.
f.	Develop a weighted total score.

We note that certain changes need not be contingent upon others, and it may be that further targeted work needs to be undertaken to assess their desirability. In general, implementing minor changes to the HoNOS instrument will be least disruptive, and maintain the continuity of the HoNOS as a key outcomes measure. We repeat a point made earlier that there are multiple and serious problems with Scale 8 (any other psychological problems) that minor changes will not address.

The decisions required to make even minor enhancements to an instrument that is in current heavy use are complex and, desirably, should involve all user-countries of the

HoNOS. This of course begs the question of how international collaboration would be coordinated and resourced. This would initially need to be discussed with colleagues in the UK and elsewhere, although the option of Australia taking independent action remains a possibility.

References

Allan, S. & McGonagle, I. (1997) A comparison of HoNOS with the Social Behaviour Schedule in three settings. *Journal of Mental Health* 6(2): 117-124

Brooks, R. (2000) The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry* 34: 504-511

Buckingham, W., Burgess, P., Solomon, S., Pirkis, J. & Eagar, K. (1998) Developing a casemix classification of mental health services. Canberra, Commonwealth Department of Health and Family Services.

Burns, A., Beevor, A., Lelliott, P., Wing, J., Blakey, A., Orrell, M., Mulinga, J. & Hadden, S. (1999) Health of the Nation Outcome Scales for Elderly People (HoNOS65+): Glossary for the HoNOS65+ score sheet. *British Journal of Psychiatry* 174: 435-438

Burns, A., Beevor, A., Lelliott, P., Wing, J., Blakey, A., Orrell, M., Mulinga, J. & Hadden, S. (1999) Health of the Nation Outcome Scales for Elderly People (HoNOS 65+). *British Journal of Psychiatry* 174: 424-427

College Research Unit (1996) Health of the Nation Outcome Scales: Report on Research. London, Royal College of Psychiatrists.

College Research Unit (1996) HoNOS - Health of the Nation Outcome Scales. Brief Report. London, Royal College of Psychiatrists.

Commonwealth Department of Health and Aged Care (1999) Mental Health Information Development: National Information Priorities and Strategies under the National Second Mental Health Plan 1998-2003. Canberra, Commonwealth of Australia.

Commonwealth Department of Health and Ageing (2002) National Outcomes and Casemix Collection: Technical specification of State and Territory reporting requirements for the otucomes and casemix components of 'Agreed data' under National Mental Health Information Development Funding Agreements. Canberra.

Eagar, K., Buckingham, B., Coombs, T., Trauer, T., Graham, C., Eagar, L. & Callaly, T. (2000) Outcome Measurement in Adult Area Mental Health Services: Implementation Resource Manual. Melbourne, Victoria, Department of Human Services Victoria.

Gaines, P., Bower, A. & Buckingham, W. (2001) Mental Health Classification and Outcomes Study: Study Resource Manual. Auckland, Health Research Council of New Zealand.

Gaines P, Bower A, Buckingham W, Eagar K, Burgess P. & Green J. (2003) New Zealand Mental Health Classification and Outcomes Study: Final Report. Auckland, Health Research Council of New Zealand.

Gowers, S. G., Harrington, R. C., Whitton, A., Beevor, A., Lelliott, P., Jezzard, R. & Wing, J. K. (1997) Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA): *Trainer's Guide September 1995 - March 1997*. College Research Unit, Royal College of Psychiatrists, London.

Gowers, S. G., Harrington, R. C., Whitton, A., Beevor, A., Lelliott, P., Jezzard, R. & Wing, J. K. (1999) Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA): Glossary for the HoNOSCA score sheet. *British Journal of Psychiatry* 174: 428-431

Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J. & Jezzard, R. (1999) Brief scale for measuring the outcomes of emotional and behavioural disorders in children. *British Journal of Psychiatry* 174: 413-416

Morris-Yates, A., Barber, H., Harris, M. & Zapart, S. (1999) Trainer's manual for the Health of the Nation Outcome Scales, Version 5, Adult. Canberra, Commonwealth Department of Health and Aged Care.

Morris-Yates, A. & The Strategic Planning Group for Private Psychiatric Services Data Collection and Analysis Working Group (2000) A National Model for the Collection and Analysis of a Minimum Data Set with Outcome Measurement for Private, Hospital-based, Psychiatric Services. Canberra, Commonwealth of Australia.

Orrell, M., Yard, P., Handysides, J. & Schapira, R. (1999) Validity and reliability of the Health of the Nation Outcome scales in psychiatric patients in the community. *British Journal of Psychiatry* 174: 409-412

Rosen, A., Hadzi-Pavlovic, D. & Parker, G. (1989) The Life Skills Profile: a measure assessing function and disability in schizophrenia. *Schizophrenia Bulletin* 15(2): 325-337

Roy, A., Matthews, H., Clifford, P., Fowler, V. & Martin, D. M. (2002) Health of the Nation Outcome Scales for People with Learning Disabilities (HoNOS-LD). *British Journal of Psychiatry* 180: 61-66

Roy, A., Matthews, H., Clifford, P., Fowler, V. & Martin, D. M. (2002) Health of the Nation Outcome Scales for People with Learning Disabilities (HoNOS-LD). Glossary for HoNOS-LD score sheet. *British Journal of Psychiatry* 180: 67-70

Shergill, S. S., Shanker, K. K., Seneviratna, K. & Orrell, M. W. (1999) The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly. *Journal of Mental Health* 8(5): 511-521

Stedman, T., Yellowlees, P., Mellsop, G., Clarke, R. & Drake, S. (1997) Measuring consumer outcomes in mental health. Canberra, A.C.T., Department of Health and Family Services.

Trauer, T. (1999) The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health* 8(5): 499-509

Trauer, T. (2003) Analysis of outcome measurement data from the four Victorian "Round One" agencies, Unpublished report.

Trauer, T. & Callaly, T. (2002) Concordance between mentally ill clients and their case managers using the Health of the Nation Outcome Scales (HoNOS). *Australasian Psychiatry* 10(1): 24-28

Trauer, T., Callaly, T., Little, J., Shields, R. & Smith, J. (1999) Health of the Nation Outcome Scales (HoNOS): results of the Victorian field trial. *British Journal of Psychiatry* 174: 380-388

Wing, J., Curtis, R. & Beevor, A. (1996) Health of the Nation Outcome Scales. Trainers' Guide. London, Royal College of Psychiatrists.

Wing, J., Curtis, R. H. & Beevor, A. (1999) Health of the Nation Outcome Scales (HoNOS) Glossary for HoNOS score sheet. *British Journal of Psychiatry* 174: 432-434

Wing, J., Curtis, R. H. & Beevor, A. S. (1996) HoNOS: Health of the Nation Outcome Scales: Report on Research and Development July 1993 - December 1995. London, Royal College of Psychiatrists Research Unit.

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B. G., Hadden, S. & Burns, A. (1998) Health of the Nation Outcome Scales (HoNOS). *British Journal of Psychiatry* 172: 11-18

Appendix 1: HoNOS inter-rater reliability coefficients

Study	1	2	3	4	5	6
N	100	97	50	100	100	20
1 Aggression	.97	.80	.73	.60	.88	.61
2 Self-harm	.88	.92	.83	.82	.64	.52
3 Alcohol & drug use	.99	.61	.86	.65	---	.81
4 Cognitive problems	.81	.92	.41	.55	.81	.50
5 Physical problems	.88	.89	.62	.67	.77	.44
6 Hallucinations/delusions	.87	.92	.83	.80	.71	.88
7 Depression	.84	.89	.79	.61	.69	.60
8 Other symptoms	.95	.52	.61	.27	.66	.27
9 Relationship problems	.74	.78	.60	.39	.77	.33
10 Activities of daily living	.71	.90	.68	.58	.76	.20
11 Accommodation	.83	.72	.47	.33	.59	.28
12 Occupational problems	.49	.51	.38	---	.56	.37

1 Wing et al. (1998) Nottingham study. Coefficients are intraclass correlation coefficients.
2 Wing et al. (1998) Manchester study. Coefficients are intraclass correlation coefficients.
3 Trauer et al. (1999) Victorian HoNOS field trial. Coefficients are intraclass correlation coefficients.
4 Orrell et al. (1999). Coefficients are unweighted kappas.
5 Shergill et al. (1999) Used HoNOS with elderly patients. Coefficients are unweighted kappas
6 Brooks (2000). Coefficients are intraclass correlation coefficients.

Appendix 2: Missing items on HoNOS: percentages of assessments

	Victorian HoNOS field trial, 1996[1]			Victorian "Round One" agencies, 2000-2002[2]			Australian MH-CASC study, 1996[3]		
	All	Acute inpat	Comm	All	Acute inpat	Comm	All	Acute inpat	Comm
N	3,535	458	3,076	14,897	1,865	7,188	48,799	8,910	34,318
1	0.2	0.0	0.2	0.3	0.4	0.5	0.9	0.8	0.9
2	0.3	0.0	0.3	0.4	0.8	0.5	1.6	0.9	1.1
3	1.0	0.4	1.1	0.8	1.9	0.8	1.6	1.4	1.4
4	0.4	0.7	0.4	0.7	1.6	0.7	0.8	0.8	0.7
5	0.3	0.2	0.4	0.7	1.0	0.9	0.9	0.9	0.9
6	0.6	1.3	0.5	0.9	1.2	1.1	1.7	1.3	1.2
7	0.4	0.9	0.3	0.6	1.3	0.6	1.6	2.0	1.0
8	0.7	0.9	0.7	9.7	38.9	6.7	11.9	21.4	9.5
9	0.4	0.9	0.4	1.2	2.2	1.1	2.0	3.8	1.7
10	0.3	0.2	0.3	1.1	2.2	1.2	2.0	2.9	1.9
11	2.2	10.5	1.0	1.4	3.2	1.1	7.8	15.1	4.8
12	1.0	4.6	0.5	1.6	4.0	1.3	7.5	15.0	4.7

1 Trauer et al. (1999) Victorian HoNOS field trial.
2 Trauer (2003) unpublished manuscript.
3 Buckingham et al. (1998). Unlike the other two studies, the MH-CASC study identified three locations: acute inpatient settings, community/outpatient settings, and non-acute inpatient settings.

Appendix 3: Item 8 data from Victorian "Round One" agencies, Trauer (2003)

	0	1	2	3	4	Missing	Total
A Phobic	371	62	94	59	22	26	634
B Anxiety	619	918	1,341	683	186	111	3,858
C Obs-comp	6	66	110	87	61	3	333
D Stress	79	555	820	330	82	17	1,883
E Dissociative	6	18	47	28	12	6	117
F Somatoform	5	42	60	31	17	4	159
G Eating	20	139	210	95	38	5	507
H Sleep	141	534	826	370	120	25	2,016
I Sexual	17	54	65	38	14	4	192
J Other	964	149	228	143	58	68	1,670
Invalid entry	337	0	0	1	0	11	349
Missing	1,967	13	10	13	3	1,173	3,179
Total	4,532	2,250	3,871	1,878	613	1,453	14,897