# Assessing the content validity of the revised Health of the Nation Outcome Scales 65+: the HoNOS Older Adults

Meredith G. Harris

Caley Tapp

Urska Arnautovska

Tim Coombs

Rosemary Dickson

Mark Smith

Angela Jury

Jennifer Lai

Mick James

Jon Painter

Philip M. Burgess

**Aims and method:** Recently, the Health of the Nation Outcome Scales (HoNOS) 65+ was revised. Twenty-five experts from Australia and New Zealand completed an anonymous web-based survey about the content validity of the revised measure, the HoNOS Older Adults (HoNOS OA).

**Results:** All 12 HoNOS OA scales were rated by most (≥75%) experts as 'important' or 'very important' for determining overall clinical severity among older adults. Ratings of sensitivity to change, comprehensibility and comprehensiveness were more variable, but mostly positive. Experts' comments provided possible explanations. For example, some experts suggested that additional older adult-specific examples be included in the glossary (e.g., for scales measuring depressed mood, problems with relationships, and problems with activities of daily living).

**Clinical implications:** Experts agreed that the HoNOS OA measures important constructs. Training may be needed to orient experienced raters to the rationale for some revisions. Further psychometric testing of the HoNOS OA is recommended.

**Key words:** older adults, routine outcome measurement, content validity, measurement properties, mental health services,

The clinician-rated Health of the Nation Outcome Scales 65+ (HoNOS 65+) was first published in 1999.[1, 2] It was adapted from the HoNOS for working age adults[3] based on feedback that specific content changes were needed to meet the needs of older adults.[4, 5] Like its working age equivalent, the HoNOS 65+ comprises 12 scales that cover the types of problems experienced by older adults in contact with specialised mental health services.[3] Maximum severity is rated (usually) for the previous 2 weeks, with ratings guided by a glossary.

Concurrent reviews of the HoNOS and HoNOS 65+, led by the Royal College of Psychiatrists with participation from Australia and New Zealand, commenced in 2014.[6] Both measures were revised, with the intent of reducing ambiguity and inconsistency in the glossaries, and improving reliability, validity and utility. In addition, for scales where it was considered that presenting needs were the same regardless of age, the wording of the two glossaries was aligned. Further, the revised HoNOS 65+ was named the HoNOS Older Adults (HoNOS OA).[6] These revisions reflect a shift towards later onset of functional impairment[7] and can accommodate variations between services and over time in the age cut-offs for older adult services.

The HoNOS OA was published in 2018 and, as yet, there is no empirical evidence about its measurement properties. When a measure is revised, the assessment of content validity - whether the content of a measure adequately reflects the construct(s) of interest – is recommended as the first step because deficits in content validity may impact other properties.[8] For multi-dimensional measures the content validity aspects of relevance, comprehensiveness, and comprehensibility of each item should be assessed. We designed and conducted a study of the content validity of the 12 HoNOS OA scales.

## Method

This descriptive study involved completion of an anonymous web-based survey by experts from Australia and New Zealand. A minimum of 10 experts was sought from each country. Candidates were identified through database searches and professional networks. Expertise was defined as: making or supervising HoNOS 65+ ratings; psychometric or clinical effectiveness research involving the HoNOS 65+; or using HoNOS 65+ ratings at a macro level (e.g., staff training, monitoring service quality).

Experts were invited to participate via an email containing a link to the survey (one expert subsequently requested a paper-and-pencil version). The survey commenced with an information sheet; written informed consent was obtained from all participants. Consenting participants were asked questions about relevant professional characteristics. They were then presented with each scale of the HoNOS OA and asked for their opinion in response to 6 'core' questions about its relevance, comprehensibility and comprehensiveness.

1. How important is this scale for determining overall clinical severity for older adult mental health service consumers? *(relevance)*
2. How likely are repeat ratings on this scale to capture change in [scale-specific problems] during a period of mental health care? *(relevance)*
3. How well do the descriptors for each rating of 0-4 cover the range of [scale-specific problems] typically seen among older adult mental health service consumers? *(comprehensiveness)*
4. How helpful is the glossary for determining what to include when rating [scale-specific problems]? *(comprehensibility)*
5. How well do the descriptors for each rating of 0-4 correspond to the different levels of severity of [scale-specific problems]? *(comprehensibility)*
6. How consistent is the wording of the glossary with language used in contemporary mental health practice? *(comprehensibility)*

Responses were made on a 4-point Likert scale[9] (e.g., 1=Not important, 2=Somewhat important, 3=Important, 4=Very important). Open-ended questions encouraged experts to elaborate on their 'negative' ratings (i.e., ratings of 1 or 2). At the end of the survey, experts were invited to make additional comments about the content of the HoNOS OA.

An item-level content validity index (I-CVI)[10, 11] shows the proportion of experts who rated each scale positively on each core question. The I-CVI is the total number of 'positive' ratings (i.e., ratings of 3 or 4), divided by the number of raters. At the 5% significance level, an I-CVI value $\geq$0.75 indicates 'excellent' content validity when there are $\geq$16 raters.[10] An average deviation (AD) index was used to measure the dispersion of responses around the median, with lower values indicating less dispersion.[12] At the 5% significance level with a 4-point response scale, AD index values $\leq$0.68 indicate 'acceptable and statistically significant agreement' when there are $\geq$15 raters.[12] Statistical analyses were conducted in Stata 16.0 (StataCorp, College Station, TX, USA). Open-ended comments were analysed independently by two members of the research team using Template Analysis.[13, 14] The initial coding template was based on themes arising from a concurrent study of the content validity of the revised HoNOS for working age adults (HoNOS 2018),[15] then refined iteratively as the comments were coded. The final template was applied across all comments.

Each site received approval to conduct the study and to pool the data for analysis - Australia (University of Queensland Medicine, Low and Negligible Risk Ethics Committee, 2019/HE002824; Research Ethics and Integrity, 2021/HE000113); New Zealand (ethics review not required; Ministry of Health, Health and Disability Ethics Committees).

## Results

Of 35 invited experts, 25 completed the survey (71% response rate). Most (72%) were psychiatrists or nurses; the remainder comprised a mix of disciplines. Experts represented the 3 types of expertise sought and, collectively, had used the HoNOS 65+ across a mix of settings. One-quarter had used the HoNOS OA in their work (Table 1).

### Experts' ratings of relevance, comprehensiveness and comprehensibility

The I-CVI values show that 'positive' ratings were made by at least half (i.e., I-CVI≥0.5) of experts on all but one of the core questions, and by three-quarters of experts (i.e., I-CVI≥0.75) on nearly 70% of core questions (Tables 2 and 3).

All 12 scales met the *a priori* criterion for excellent content validity (I-CVI≥0.75) for the question assessing importance for determining overall clinical severity (an indicator of relevance) (Tables 2 and 3). Between 6 and 9 scales met the criterion for all other questions.

Three scales met the criterion for all questions: Scale 5 (Physical illness or disability problems), Scale 6 (Problems associated with hallucinations and /or delusions), and Scale 11 (Problems with housing and living conditions). Three scales met the criterion for all but 1 question: Scale 4 (Cognitive problems), Scale 7 (Problems with depressed mood) and Scale 10 (Problems with activities of daily living). Conversely, Scale 2 (Non-accidental self-injury) met the threshold only for the question assessing importance for determining overall clinical severity.

AD index values indicated acceptable and statistically significant agreement between experts, with 3 exceptions related to scales that measure behavioural problems - Scale 1 (Overactive or aggressive or disruptive or agitated behaviour), Scale 2 (Non-accidental self-injury) and Scale 3 (Problem drinking or drug-taking).

### Experts' concerns

Analysis of experts' elaborations on their 'negative' ratings revealed 1 theme related to comprehensiveness, 5 related to comprehensibility and 2 to relevance. A further theme highlighted the important role of HoNOS training. The themes are summarised below, with illustrative quotations.

#### *Themes related to comprehensiveness*

*Incomplete coverage*

A recurring concern was that some scale descriptors were not sufficiently specific to older adults:

> "[In] older adults self-harm is often more subtle - not taking medications or accepting required health interventions, isolating or withdrawing from supports." (*Scale 2. Non-accidental self-injury*).

> "…might be worth specifying beyond recommended limits adjusted for age. Perhaps more specifiers for adverse effects including effects on relationships, self-care, falls" (*Scale 3. Problem drinking or drug-taking*).

> "I think this item is too limited in its scope. It does not mention the common types of elder abuse encountered in clinical practice" (*Scale 9. Problems with relationships*).

#### *Themes related to comprehensibility*

*Lack of fit with clinical thinking*

For some scales, experts identified challenges in rating problems separately from the disorders with which they may be associated.

> "Severity of neurocognitive disorder is not just determined by cognitive impairment […] it should include behaviour, self-care, etc." (*Scale 4. Cognitive problems*).

> "…it would make more sense to include [thought disorder] with other positive psychotic symptoms such as delusions" (*Scale 4. Cognitive problems*).

> "Include a sentence to clarify that it is depressed mood not clinical depression that is being rated" (*Scale 7. Problems with depressed mood*).

For one scale, experts identified different concerns.

> "It would be more consistent with clinical reasoning for assessing suicidal risk by adding more risk factors into the descriptors, such as whether having suicidal plans, access to suicidal means, intention to act …" (*Scale 2. Non-accidental self-injury*).

> "There is a move away from 'accidental' vs 'intentional' and more towards self-harm in general" (*Scale 2. Non-accidental self-injury*).

*Too many phenomena*

Several experts noted that some scales combine too many different phenomena together:

> "I have two issues with this item. the first is the conflation of deliberate self-harm with suicidal behaviour…" (*Scale 2. Non-accidental self-injury*).

> "The difficulty is clumping together a range of cognitive problems which may not correspond e.g. language might be good memory might be poor. Thought disorder might be prominent, problem solving might be intact" (*Scale 4. Cognitive problems*).

with not all included phenomena mentioned in the descriptors for each severity level:

> "Discuss[es] suicide in step 2 but not in step 3 - language needs to be consistent" (*Scale 2. Non-accidental self-injury*).

> "Inconsistent exclusion of adverse consequences from rating 3 (included in 2 and 4-5)" (*Scale 4. Cognitive problems*).

*Ambiguity*

Some experts indicated ambiguity in the glossary wording.

> "Not clearly identified what the psychological effects of excessive alcohol or substance use may be" (*Scale 3. Problem drinking or drug-taking*).

> "Occupation and activities: rating the 'quality of meaningful' activities seems rather subjective. This may prove difficult to rate consistently" (*Scale 12. Problems with occupation and activities*).

*Need for more description or examples*

Comments around multiple phenomena and ambiguity often corresponded to suggestions for more descriptions or examples to be added to the glossary.

> "It may be useful to expand on what constitutes non-compliant or resistive behaviour" (*Scale 1. Overactive or aggressive or disruptive or agitated behaviour*).

> "The scale should have more about IADLs than ADLs, in psychiatric care the former are very important - the latter are important but of greater issue for long term residential care" (*Scale 10. Problems with activities of daily living*).

*Assessment challenges*

Assessment challenges were noted for some scales.

> Sometimes it is difficult to determine what is the most severe problem when there are multiple and almost equally severe problems." (*Scale 8. Other mental and behavioural problems*).

> "The problem with the scale is that it requires an independent observation to be rated - that is often not possible, not relevant to the case or occasionally refused" (*Scale 11. Problems with housing and living conditions*).

> "Too many judgements here that are likely based on inadequate information" (*Scale 12. Problems with occupation and activities*).

***Themes related to relevance***

*Challenges to capturing change*

Some experts expressed concern that some scales lack sensitivity to describe the subtle, delayed or rapid changes often seen in clinical practice.

> "Presentation of a person can change very rapidly, clinical assessment and documentation is more useful in tracking changes of a person's presentation" (*Scale 1. Overactive or aggressive or disruptive or agitated behaviour*).

> "Change in dementia is slow and change will not be noticeable within the typical period of clinical contact" (*Scale 4. Cognitive problems*).

Others commented on other challenges to capturing change.

> "It could be hard to show change, for example, a patient may be elated, with poor sleep and appetite and marked anxiety. Three of the 4 might improve but the 4th is unchanged - the scale does not alter" (*Scale 8. Other mental and behavioural problems*).

> "Some elements of this scale may not be modifiable or changeable if communities have sparse resourcing and groups and transportation is an issue" (*Scale 12. Problems with occupation and activities*).

*Lack of relevance*

Some experts considered Scale 12 to be less relevant because of its focus on the environment:

> "In my view, this item is not needed in the scale… Availability of activities is not a patient issue, it's a social system issue" (*Scale 12. Problems with occupation and activities*).

or because the instructions about what to include when rating the scale did not cover all relevant treatment contexts:

> "would be good to have more mention of residential care situations" (*Scale 11. Problems with housing and living conditions*).

***Need for training***

Some comments from experts reinforced the need for training.

> "In New Zealand the cultural context should be emphasized. […] This is important for Māori and Pacific peoples." (*Overarching rating instructions*).

> "I find some confusion in the glossary where it states 'rate what the person is capable of doing' but then also states 'include any lack of motivation'. A person may be capable of doing something but is not doing it because of low motivation" (*Scale 10. Problems with activities of daily living*).

**Experts' summary comments**

The survey tasks did not involve comparing the HoNOS OA to the original HoNOS 65+. Nonetheless, some experts endorsed the revised title.

> "Well I notice it's no longer "65+" … I think that's an improvement! I like Older Adult rather than older persons for example and 65 is stigmatising and misleading…"

Others felt the measure had not improved, regardless of revisions.

> "This OA version is not much of an improvement on the 65+ version."

These mixed views were reflected in comments about ambiguity and consistency in the descriptors.

> "The content of HoNOS OA includes more detailed descriptions and examples for some of the scales, which are very helpful to rate with confidence."

> "It is too narrow in its focus and some of the items are poorly specified or lacking in range."

## Discussion

A key finding was that experts held the HoNOS OA scales to be important for determining clinical severity among older adults in contact with specialised mental health services. This accords with studies of the HoNOS 65+,[16, 17] and provides reassurance that the glossary revisions have not adversely affected this core aspect of content validity.

Results of the thematic analysis may help explain why ratings of other aspects of content validity were more variable. Experts suggested additional older adult-specific examples for some scales – for example, not taking medications as a form of self-harm in Scale 2 (Non-accidental self-injury); elder abuse in Scale 9 (Problems with relationships). This issue may have attracted comment among this sample of experts with a high level of familiarity with the HoNOS 65+ glossary, because the wording of some examples was revised to be the same for the HoNOS OA and HoNOS 2018. However, it is important to note that, even in the absence of these older adult-specific examples, the revised glossary provides the opportunity to rate the phenomena of interest (e.g., passive forms of self-harm in Scale 2 and problematic relationships in Scale 9).

Another concern was that some scales might not reflect usual or contemporary clinical thinking about certain clinical problems. For example, some comments suggested it may not be clinically meaningful to rate thought disorder on Scale 4 (Cognitive problems) and depressed mood on Scale 7 (Depressed mood) independently of the disorder(s) in which they manifest. These issues may have attracted comment because the revision increased the emphasis on rating these phenomena. For Scale 2 (Non-accidental self-injury), experts had different views about how self-injury should be conceptualised. This may reflect an acknowledged lack of consistent terminology for non-accidental self-injury[18] and/or difficulties identifying non-accidental self-injury in older adults.[19]

## Implications

Experts rated all HoNOS OA scales as important; this may give clinicians confidence that the measure provides information relevant to clinical decision making and care planning. The findings may help inform services to make decisions about implementing the HoNOS OA, noting that other sources of evidence (e.g., inter-rater reliability, utility and infrastructure costs) are also likely to be needed.

Experts' suggestions to include more older adult-specific examples might raise concerns about the utility of the HoNOS OA. Conversely, the inclusion of more examples could adversely affect utility, for example by encouraging raters to rely on the descriptors as an exhaustive checklist or by making the measure longer and less acceptable to clinicians. Studies of the measures' utility could explore these possibilities. Given the breadth of problems covered by the HoNOS OA, training remains critical. Training may also need to orient experienced clinicians to the rationale for certain revisions, including the inclusion of fewer older age-specific examples.

Some comments highlighted the challenge of rating some clinical phenomena independent of disorder. It remains important to emphasise (through training and other means) that the HoNOS OA is not intended to be used as a screening or diagnostic tool.

## Strengths and limitations

This study included experts from 2 countries with a long history of using the HoNOS 65+, lending support for the 'real-world' relevance of the results. Survey questions were designed from best practice principles[8, 20] and we included a qualitative component, which enabled us to explore possible explanations for patterns in the experts' quantitative responses.[21] A limitation was the sample size, which fell just short of 'adequate' (i.e., ≥30) for a quantitative content validity study, but is 'very good' (i.e., ≥7) for a qualitative study.[21] More than one-quarter of invited experts did not complete the survey. We do not know whether those who did not participate may have held different views from those who did; however, the responses represented a range of views, both positive and negative. Although we drew on multiple sources to identify experts, there may have been selection biases. We relied on bibliographic evidence and informants to identify experts, rather than quantified criteria.[22, 23] However in the survey, all experts self-identified at least one area of HoNOS expertise. As the open-ended questions focused on experts' concerns, any interpretation of the findings should consider the qualitative and quantitative results in tandem.

**Conclusions**

Findings indicate that the HoNOS OA scales remain important for determining clinical severity among older adults in contact with specialised mental health services. Given the decreased emphasis on age-specific examples in the glossary, training may need to include a focus on orienting experienced raters to the changes to the glossary. Overall, findings support progression to inter-rater reliability and utility of the HoNOS OA.

**References**


1. Burns A, Beevor A, Lelliott P, Wing J, Blakey A, Orrell M, et al. Health of the Nation Outcome Scales for elderly people (HoNOS 65+). Br J Psychiatry. 1999;174:424-7.
2. Burns A, Beevor A, Lelliott P, Wing J, Blakey A, Orrell M, et al. Health of the Nation Outcome Scales for elderly people (HoNOS 65+). Glossary for HoNOS 65+ score sheet. Br J Psychiatry. 1999;174:435-8.
3. Wing JK, Beevor AS, Curtis RH, Park SB, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS). Research and development. Br J Psychiatry. 1998;172:11-8.
4. Ashaye K, Knightley S, Shergill S, Orrell M. Do the Health of the Nation Outcome Scales predict outcome in the elderly mentally ill? A 1-year follow-up study. J Ment Health. 2009;8(6):615-20.
5. Shergill SS, Shankar KK, Seneviratna K, Orrell MW. The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly. J Ment Health. 1999;8(5):511-21.
6. James M, Buckingham B, Cheung G, McKay R, Painter J, Stewart MW. Review and update of the Health of the Nation Outcome Scales for Elderly People (HoNOS65+). BJPsych Bull. 2018;42(6):248-52.
7. World Health Organization. World Report on Ageing and Health 2015. Available from: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf;jsessionid=84F1F08EA158E6029C38C833AC45567F.
8. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.
9. Likert R, Roslow S, Murphy G. A Simple and Reliable Method of Scoring the Thurstone Attitude Scales. The Journal of Social Psychology. 1934;5(2):228-38.
10. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489-97.
11. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382-5.
12. Burke MJ, Dunlap WP. Estimating interrater agreement with the average deviation index: A user's guide. Organ Res Methods. 2002;5(2):159-72.
13. King N. Using Templates in the Thematic Analysis of Text. In: Cassell C, Symon G, editors. Essential Guide to Qualitative Methods in Organizational Research. London: SAGE Publications Ltd; 2004. p. 256-70.
14. Brooks J, McCluskey S, Turley E, King N. The Utility of Template Analysis in Qualitative Psychology Research. Qual Res Psychol. 2015;12(2):202-22.

15. Harris M, Tapp C, Arnautovska U, Coombs T, Dickson R, James M, et al. Assessing the content validity of the revised Health of the Nation Outcome Scales (HoNOS 2018). Under review.
16. Burgess P, Trauer T, Coombs T, McKay R, Pirkis J. What does 'clinical significance' mean in the context of the Health of the Nation Outcome Scales? Australas Psychiatry. 2009;17:141-48.
17. National Mental Health Information Development Expert Advisory Panel. Mental Health National Outcomes and Casemix Collection: NOCC Strategic Directions 2014 – 2024. Canberra 2013. Available from: https://www.amhocn.org/sites/default/files/publication_files/nocc_strategic_directions_2014-2024.pdf.
18. Silverman MM. Challenges to Defining and Classifying Suicide and Suicidal Behaviors. In: O'Connor RC, Pirkis J, editors. The International Handbook of Suicide Prevention. Second ed. Hoboken: Wiley; 2016. p. 11-35.
19. De Leo D, Arnautovska U. Prevention and Treatment of Suicidality in Older Adults. In: O'Connor RC, Pirkis J, editors. The International Handbook of Suicide Prevention. Second ed. Hoboken: Wiley; 2016. p. 323-45.
20. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. Value Health. 2011;14(8):978-88.
21. Terwee C, Prinsen C, Chiarotto A, de Vet H, Bouter L, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMS. User manual version 1.0. February 2018. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf.
22. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health. 1997;20(3):269-74.
23. Leyden KN, Hanneman SK. Validity of the Modified Richmond Agitation-Sedation Scale for use in sedated, mechanically ventilated swine. J Am Assoc Lab Anim Sci. 2012;51(1):63-68.

**Table 1.** Characteristics of experts who completed the survey (N = 25)

| | n | % |
|---|---|---|
| **Main professional background** | | |
| Nurse | 7 | 28 |
| Psychologist | 0 | 0 |
| Clinical Psychologist | 3 | 12 |
| Social worker | 1 | 4 |
| Psychiatrist | 11 | 44 |
| Occupational therapist | 2 | 8 |
| Other | 1 [a] | 4 |
| **Expertise with HoNOS 65+** [b] | | |
| Rating HoNOS 65+ or reviewing HoNOS 65+ ratings made by others | 23 | 92 |
| Research in the measurement properties of the HoNOS 65+ and/or measuring clinical effectiveness | 3 | 12 |
| HoNOS 65+ staff training and/or using HoNOS 65+ results at a macro level | 15 | 60 |
| Other expertise working with HoNOS 65+ | 4 | 16 |
| **Mental health settings worked with HoNOS 65+** [b] | | |
| Inpatient | 17 | 68 |
| Residential [c] | 3 | 12 |
| Community services | 23 | 92 |
| Other, non-clinical setting | 1 | 4 |
| **Aware of HoNOS OA prior to survey** | | |
| No, I was not aware of the HoNOS OA at all | 12 | 48 |
| Yes, I was aware of the HoNOS OA, but have not used it in my work | 6 | 24 |
| Yes, I have used the HoNOS OA in my work | 6 | 24 |
| Not sure | 1 | 4 |
| Other | 0 | 0 |
| | **M (SD)** | **Range** |
| **Years worked in mental health** | 24.08 (10.84) | 2-42 |
| **Years worked with the HoNOS** | 13.92 (7.06) | 2-28 |

HoNOS, Health of the Nation Outcome Scales. M, mean. SD, standard deviation.

[a] "Consumer and Family Leader". [b] Categories not mutually exclusive. [c] 'Residential' category included only in the Australian version of the survey.

**Table 2.** Experts' ratings of the content validity of the HoNOS OA scales: relevance and comprehensiveness

| | Relevance | | | | | | | | Comprehensiveness | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | How important is this scale for determining overall clinical severity for older adult mental health service consumers? | | | | How likely are repeat ratings on this scale to capture change in [scale-specific problems] during a period of mental health care? | | | | How well do the descriptors for each rating of 0-4 cover the range of [scale-specific problems] typically seen among older adult mental health service consumers? [a] | | | |
| **HoNOS OA scale** | **n** | **Range** | **I-CVI** | **AD** | **n** | **Range** | **I-CVI** | **AD** | **n** | **Range** | **I-CVI** | **AD** |
| Scale 1. Overactive or aggressive or disruptive or agitated behaviour | 25 | 1-4 | **0.80** | 0.76 | 25 | 1-4 | 0.64 | 0.68 | 24 | 2-4 | **0.75** | 0.50 |
| Scale 2. Non-accidental self-injury | 23 | 2-4 | **0.87** | 0.57 | 24 | 1-4 | 0.67 | 0.58 | 23 | 1-4 | 0.48 | 0.83 |
| Scale 3. Problem drinking or drug-taking | 23 | 1-4 | **0.83** | 0.52 | 24 | 1-4 | 0.67 | 0.63 | 23 | 1-4 | 0.57 | 0.65 |
| Scale 4. Cognitive problems | 25 | 2-4 | **0.88** | 0.60 | 24 | 1-4 | **0.75** | 0.50 | 25 | 2-4 | **0.84** | 0.40 |
| Scale 5. Physical illness or disability problems | 25 | 2-4 | **0.88** | 0.60 | 25 | 1-4 | **0.76** | 0.60 | 25 | 1-4 | **0.76** | 0.56 |
| Scale 6. Problems associated with hallucinations and /or delusions | 24 | 2-4 | **0.92** | 0.58 | 23 | 1-4 | **0.87** | 0.52 | 24 | 2-4 | **0.88** | 0.38 |
| Scale 7. Problems with depressed mood | 25 | 2-4 | **0.96** | 0.44 | 25 | 1-4 | **0.84** | 0.48 | 25 | 1-4 | 0.72 | 0.60 |
| Scale 8. Other mental and behavioural problems | 24 | 1-4 | **0.92** | 0.50 | 25 | 1-4 | 0.72 | 0.68 | 25 | 1-4 | **0.76** | 0.48 |
| Scale 9. Problems with relationships | 25 | 2-4 | **0.84** | 0.64 | 25 | 2-4 | 0.68 | 0.60 | 25 | 1-4 | 0.68 | 0.64 |
| Scale 10. Problems with activities of daily living | 24 | 2-4 | **0.96** | 0.50 | 24 | 1-4 | **0.76** | 0.52 | 24 | 2-4 | 0.71 | 0.42 |
| Scale 11. Problems with housing and living conditions | 25 | 1-4 | **0.80** | 0.60 | 25 | 1-4 | **0.88** | 0.46 | 25 | 1-4 | **0.76** | 0.44 |
| Scale 12. Problems with occupation and activities | 25 | 1-4 | **0.80** | 0.68 | 24 | 1-4 | **0.76** | 0.52 | 25 | 1-4 | 0.68 | 0.60 |

AD, average deviation. I-CVI, item-level content validity index. n, number. Bold CVI values meet criteria for excellent content validity (i.e., I-CVI ≥ 0.75). [a] To fit the wording of Scale 8, the equivalent question for Scale 8 was: How well do problems A-O cover the range of other mental and behavioural problems typically seen among older adult mental health service consumers?

**Table 3.** Experts' ratings of the content validity of the HoNOS OA scales: comprehensibility

| | Comprehensibility | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | How helpful is the glossary for determining what to include when rating [scale-specific problems]? [a, b] | | | | How well do the descriptors for each rating of 0-4 correspond to the different levels of severity of [scale-specific problems]? | | | | How consistent is the wording of the glossary with language used in contemporary mental health practice? | | | |
| **HoNOS OA scale** | **n** | **Range** | **I-CVI** | **AD** | **n** | **Range** | **I-CVI** | **AD** | **n** | **Range** | **I-CVI** | **AD** |
| Scale 1. Overactive or aggressive or disruptive or agitated behaviour | 25 | 2-4 | **0.84** | 0.36 | 25 | 2-4 | 0.60 | 0.56 | 23 | 1-4 | 0.70 | 0.43 |
| Scale 2. Non-accidental self-injury | 25 | 1-4 | 0.72 | 0.64 | 24 | 1-4 | 0.54 | 0.63 | 24 | 1-4 | 0.71 | 0.54 |
| Scale 3. Problem drinking or drug-taking | 24 | 2-4 | **0.75** | 0.46 | 23 | 1-4 | 0.70 | 0.70 | 24 | 2-4 | **0.79** | 0.33 |
| Scale 4. Cognitive problems | 25 | 1-4 | **0.88** | 0.36 | 24 | 1-4 | **0.75** | 0.50 | 25 | 1-4 | 0.68 | 0.52 |
| Scale 5. Physical illness or disability problems | 25 | 2-4 | **0.84** | 0.48 | 25 | 1-4 | **0.80** | 0.64 | 25 | 1-4 | **0.84** | 0.44 |
| Scale 6. Problems associated with hallucinations and /or delusions | 24 | 1-4 | **0.79** | 0.46 | 24 | 2-4 | **0.79** | 0.50 | 24 | 2-4 | **0.96** | 0.29 |
| Scale 7. Problems with depressed mood | 25 | 2-4 | **0.76** | 0.60 | 24 | 2-4 | **0.79** | 0.50 | 25 | 2-4 | **0.88** | 0.36 |
| Scale 8. Other mental and behavioural problems | 25 | 1-4 | 0.72 | 0.56 | 25 | 1-4 | 0.68 | 0.64 | 24 | 2-4 | **0.75** | 0.46 |
| Scale 9. Problems with relationships | 25 | 2-4 | **0.76** | 0.48 | 25 | 2-4 | 0.68 | 0.64 | 25 | 1-4 | **0.80** | 0.52 |
| Scale 10. Problems with activities of daily living | 25 | 2-4 | **0.76** | 0.44 | 24 | 1-4 | **0.79** | 0.54 | 25 | 2-4 | **0.84** | 0.36 |
| Scale 11. Problems with housing and living conditions | 25 | 1-4 | **0.80** | 0.48 | 25 | 1-4 | **0.80** | 0.48 | 24 | 1-4 | **0.88** | 0.29 |
| Scale 12. Problems with occupation and activities | 25 | 1-4 | 0.64 | 0.60 | 24 | 1-4 | **0.75** | 0.50 | 25 | 1-4 | **0.76** | 0.44 |

AD, average deviation. I-CVI, item-level content validity index. n, number. Bold CVI values meet criteria for excellent content validity (i.e., I-CVI ≥ 0.75). [a] Question text differed across scales; depending on the glossary, "what to rate and include" or "what to rate and consider" was substituted for the phrase "what to include". [b] To fit the wording of Scale 8, the equivalent question for Scale 8 was: How helpful is the glossary for determining which other mental and behavioural problem to rate on this scale?

**Acknowledgements**

The authors gratefully acknowledge the time and effort contributed by the experts who participated in this study. We are grateful to the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) in Australia for their advice and review over previous iterations of the study methodology and survey instrument. We thank the NMHIDEAP and other colleagues for assisting with the nomination of experts in Australia, in particular Kate Jackson (NSW Ministry of Health) and Prof. Rod McKay (Health Education & Training Institute). We acknowledge the national stakeholder group for the Programme for the Integration of Mental Health Data (PRIMHD) for assisting with participant recruitment in New Zealand. [NZ to confirm]

**Declaration of Interest**

MH, CT, UA, TC, RD and PB undertook this project in their roles with the Australian Mental Health Outcomes and Classification Network (AMHOCN). AMHOCN received funding from the Australian Government Department of Health to support the implementation, training and public reporting of the National Outcomes and Casemix collection, which includes the HoNOS 65+. MS, AJ and JL undertook this project in their roles with Te Pou. Te Pou is currently funded by the Ministry of Health to deliver HoNOS 65+ training and reporting of HoNOS 65+ data in New Zealand. MJ was chair of the advisory board for the review and update of the HoNOS 65+; JP was a member of the advisory board. MJ and JP were authors on the publication of the HoNOS OA. [NZ and England to confirm]

**Funding statement**

This project was led by the Australian Mental Health Outcomes and Classification Network which is funded by the Australian Government Department of Health. Te Pou is funded by the Ministry of Health in New Zealand.

**Ethics statement**

Each site received approval to conduct the study and to pool the data for analysis - Australia (University of Queensland Medicine, Low & Negligible Risk Ethics Sub-Committee, 2019/HE002824; Research Ethics and Integrity, 2021/HE000113) and New Zealand (ethics review not required; Ministry of Health, Health and Disability Ethics Committees). Written informed consent was obtained from all participants.

**Author Contribution Statement**

MH, UA, TC, RD, and PB designed the study with input from all authors. MH, TC, MS, AJ and JL managed the collection of data. CT and MH analysed the data and interpreted results. MH and CT drafted the manuscript, and all authors made revisions to the intellectual content and approved the final version.

**Data availability**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**About the authors**

**Meredith G. Harris** is a Principal Research Fellow at the School of Public Health at the University of Queensland, Australia. **Caley Tapp** is a Research Fellow at the School of Public Health at the University of Queensland, Australia. **Urska Arnautovska** is a Research Fellow at the School of Public Health at the University of Queensland, Australia. **Tim Coombs** is Consultant to the Training and Service Development, Australian Mental Health Outcomes and Classification Network, Australia. **Rosemary Dickson** is the Network Coordinator for the Australian Mental Health Outcomes and Classification Network, Australia. **Mark Smith** is Programme Lead, Outcomes and Information, at Te Pou, New Zealand. **Angela Jury** is a Researcher at Te Pou, New Zealand. **Jennifer Lai** is a Researcher at Te Pou, New Zealand. **Mick James** is the National HoNOS Advisor at the Royal College of Psychiatrists, London, UK. **Jon Painter** is a Senior Lecturer in Mental Health Nursing at Sheffield Hallam University, Sheffield, UK. **Philip M. Burgess** is a Professor at the School of Public Health at the University of Queensland, Australia.

**Corresponding author:**

Correspondence to Meredith G. Harris (meredith.harris@uq.edu.au)