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

Meredith Harris[1], Caley Tapp[1], Urska Arnautovska[1], Tim Coombs[1], Rosemary Dickson[1], Mark Smith[2], Angela Jury[2], Jennifer Lai[2], Mick James[3], Jon Painter[4], Philip Burgess[1]

[1] Australian Mental Health Outcomes Classification Network (AMHOCN), Australia
[2] Te Pou, New Zealand
[3] Royal College of Psychiatrists, United Kingdom
[4] Sheffield Hallam University, United Kingdom

May 2021

Acknowledgements

The authors gratefully acknowledge the time and effort contributed by experts who participated in this study. We are also grateful to the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) and other colleagues in the sector for assisting with expert nominations, in particular Kate Jackson (NSW Ministry of Health) and Dr. Rod McKay (Health Education & Training Institute). We would also like to thank the NMHIDEAP for their advice and review over previous iterations of the study methodology and survey instrument. 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.

This project was led by the Australian Mental Health Outcomes and Classification Network (AMHOCN). AMHOCN managed the collection of data in Australia. 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+. Te Pou managed the collection of New Zealand data. Te Pou is currently funded by the Ministry of Health to deliver HoNOS 65+ training and reporting of HoNOS 65+ data in New Zealand.

Abbreviations

AD	Average deviation
AMHOCN	Australian Mental Health Outcomes and Classification Network
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
HoNOS	Health of the Nation Outcome Scales
HoNOS OA	Health of the Nation Outcome Scales for Older Adults
HoNOS 2018	Health of the Nation Outcome Scales (2018 version)
HoNOS 65+	Health of the Nation Outcome Scales for people aged 65 years and over
I-CVI	Item-level content validity index
M	Mean
N, n	Number (sample size)
NMHIDEAP	National Mental Health Information Development Expert Advisory Panel
PRIMHD	Programme for the Integration of Mental Health Data
SD	Standard deviation

Contents

Executive Summary iv
About the HoNOS Older Adults iv
Context of this study iv
Method iv
Findings v
Conclusion v
1. Introduction 1
2. Method 3
2.1 Design and participants 3
2.2 Survey instrument 3
2.4 Data analysis 3
3. Results 5
3.1 Sample characteristics 5
3.2 Experts' ratings 6
3.3 Experts' concerns 9
3.4 Experts' summary comments 12
4. Discussion 13
4.1 Summary of findings 13
4.2 Implications 13
4.3 Strengths and limitations 14
4.4 Conclusions 14
References 15
Appendix 17

Executive Summary

About the HoNOS Older Adults

The Health of the Nation Outcome Scales 65+ (HoNOS 65+) is a clinician-rated measure comprising 12 scales that cover the kinds of problems that may be experienced by older adults in contact with specialised mental health services. The HoNOS 65+ measure was first published in 1999, adapted from the HoNOS for working age adults. After 20 years of use in clinical practice, the Royal College of Psychiatrists, as the copyright holder, assembled an advisory board of experts from the United Kingdom, Australia and New Zealand to conduct concurrent reviews of the HoNOS and the HoNOS 65+. Amendments were made to remove sources of ambiguity and inconsistency in the glossaries, and thereby improve reliability, validity and clinical 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 further aligned. The revised HoNOS 65+, published in 2018, was named the HoNOS Older Adults (HoNOS OA). These changes reflect a shift towards later onset of functional impairment and the need to allow for variations in the age cut-offs for older adult services between services and over time.

Context of this study

Despite interest in implementing the HoNOS OA in Australia, New Zealand, and some English services, there is as yet no empirical evidence regarding its measurement properties. The Australian Government Department of Health tasked the Australian Mental Health Outcomes and Classification Network (AMHOCN) to commence this process by investigating the content validity of the HoNOS OA. With guidance from the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) and input from colleagues in England and New Zealand, AMHOCN designed a study to evaluate the content validity of the HoNOS OA scales. New Zealand expressed an interest in undertaking the study locally, supported by relevant study documentation from AMHOCN.

Method

This descriptive study involved the completion of an anonymous, web-based survey by experts in Australia and New Zealand. Experts were identified through professional networks and bibliographic database searches. Expertise was defined as any of the following: making or supervising HoNOS 65+ ratings, psychometric or clinical effectiveness research involving the HoNOS 65+, or use of HoNOS 65+ ratings at a macro level (e.g., staff training, monitoring service quality).

The survey gathered information about the experts' professional backgrounds. For each HoNOS OA scale, the survey included 6 'core' questions about the content validity aspects of relevance, comprehensiveness and comprehensibility (giving a total of 72 'core' questions). Experts responded to each question by rating their opinion on a 4-point ordinal Likert scale ranging from negative to positive (e.g., 1=Not important, 2=Somewhat important, 3=Important, 4=Very important). Open-ended questions encouraged experts to elaborate on their ratings of 1 or 2. Experts were given an opportunity to share additional comments about the HoNOS OA at the end of the survey.

An item-level content validity index (I-CVI) value of ≥0.75 indicated 'excellent' content validity, and an average deviation (AD) index value of ≤0.68 indicated 'acceptable and statistically significant agreement' between experts. Open-ended comments were analysed using template analysis and organised into themes.

Findings

Of the 35 experts invited to participate, 25 completed the survey (response rate 71%). Most (72%) were psychiatrists or nurses; the remainder comprised a mix of disciplines. Experts represented the three types of expertise sought and, collectively, had used the HoNOS 65+ across a mix of clinical and non-clinical settings.

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 72 core questions. However, some aspects of content validity were more frequently endorsed than others. For example, all 12 HoNOS OA scales met the *a priori* criterion for excellent content validity (I-CVI ≥0.75) on the question assessing the *importance for determining overall clinical severity* (an indicator of relevance). In contrast, 6 scales met the criterion on the question assessing the *coverage of problems typically seen among older adult mental health service consumers* (an indicator of comprehensiveness).

Some HoNOS OA scales met the criterion for excellent content validity more often than others. For example, several scales met the criterion on all questions; these were 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). In contrast, Scale 2 (Non-accidental self-injury) only met the criterion on only one question (i.e., *importance for determining overall clinical severity*).

Almost all AD index values were equal to or below the critical 0.68 threshold, indicating acceptable and statistically significant agreement between experts.

Themes emerging from experts' open-ended comments provided insights into the variability in ratings for some aspects of content validity across the HoNOS OA scales. For example, some experts suggested modifying or expanding the rating descriptors for some scales - e.g., Scale 7 (Depressed mood), Scale 9 (Problems with relationships), and Scale 10 (Problems with activities of daily living) – to include more older adult-specific examples. Another concern was that some scales may not reflect usual or contemporary clinical thinking about certain clinical problems (e.g., that it may not be clinically meaningful to rate problems such as depressed mood and thought disorder independently of the disorder(s) with which they are associated).

Conclusion

Overall, results lend support for the content validity of the scales. Experts agreed that the HoNOS OA measures important constructs. In the main, the content of the HoNOS OA scales was viewed positively by experts with respect to their ability to capture change, comprehensiveness and comprehensibility. Training may need to orient experienced raters to the rationale underlying some revisions to the measure, including the inclusion of fewer older adult-specific examples. Further psychometric testing of the HoNOS OA, particularly inter-rater reliability and utility, is recommended.

1. Introduction

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] The HoNOS 65+ comprises 12 scales that cover the types of problems experienced by older adults in contact with specialised mental health services, equivalent to the scales in the working age version.[3] Maximum severity is rated (usually) for the previous 2 weeks, with ratings guided by a glossary (Table 1).

Table 1. The HoNOS 65+/HoNOS OA scales

Scale titles	Range of scale scores [a]
1. Overactive or aggressive or disruptive or agitated behaviour [b]	0 – 4
2. Non-accidental self-injury	0 – 4
3. Problem drinking or drug-taking	0 – 4
4. Cognitive problems	0 – 4
5. Physical illness or disability problems	0 – 4
6. Problems associated with hallucinations and /or delusions [c]	0 – 4
7. Problems with depressed mood	0 – 4
8. Other mental and behavioural problems	0 – 4
9. Problems with relationships	0 – 4
10. Problems with activities of daily living	0 – 4
11. Problems with housing and living conditions [d]	0 – 4
12. Problems with occupation and activities	0 – 4

Notes. [a] Scales are rated on a 5-point scale: 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. [b] In the original HoNOS 65+, the title for Scale 1 is 'Overactive, aggressive, disruptive or agitated behaviour'. [c] In the original HoNOS 65+, the title for Scale 6 is 'Problems associated with hallucinations and delusions'. [d] In the original HoNOS 65+, the title for Scale 11 is 'Problems with living conditions'.

The HoNOS 65+ is widely used in routine mental health service contexts in several countries. For example, in Australia, the HoNOS 65+ has been mandated for collection in all specialised public sector mental health services as part of the National Outcomes and Casemix Collection (NOCC) which was implemented from 2001. The HoNOS 65+ is also used to monitor outcomes in private hospitals with psychiatric beds.[6] The HoNOS 65+ is an important component of the Australian Mental Health Care Classification,[7] which will eventually be used for activity-based funding in the public mental health service sector. In New Zealand, the HoNOS 65+ has been mandated for collection by mental health services since 2008 and is part of the Programme for the Integration of Mental Health Data (PRIMHD) national data collection.[8] Across England, use of the HoNOS 65+ is more varied, and it is used in both its original form and a tabulated version.[9, 10]

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.[9] Both measures were revised, with the intent of reducing ambiguity and inconsistency in the glossaries, and improving reliability, validity and utility. For scales where it was considered that consumers' presenting needs were the same regardless of age, the wording of the two glossaries was further aligned. The revised HoNOS 65+ was named the HoNOS Older Adults (HoNOS OA).[9] The revisions to the HoNOS 65+ reflect a shift towards later onset of functional impairment[11] and allow for variability in the age cut-offs for older adult services between services and over time.

The HoNOS OA was published in 2018. There is interest in implementing the HoNOS OA in Australia, New Zealand and some English services. However there is, as yet, no empirical evidence about the measurement properties of the HoNOS OA to help inform such decisions. To address this gap, the Australian Mental Health Outcomes and Classification Network (AMHOCN) was tasked by the Australian Government Department of Health to investigate key measurement properties of the HoNOS OA. Content validity (whether the content of a measure adequately reflects the construct(s) of interest) was identified as a priority for investigation, as it can impact all other measurement properties.[12, 13] For multi-dimensional measures such as the HoNOS OA, the content validity aspects of relevance, comprehensiveness, and comprehensibility should be assessed for each item.

With guidance from the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) in Australia, AMHOCN designed a study to evaluate the content validity of the 12 HoNOS OA scales. New Zealand participated in the study with support from AMHOCN.

2. Method

2.1 Design and participants

This descriptive study involved completion of an anonymous web-based survey by experts from Australia and New Zealand. Experts were identified through bibliographic 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).

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).

2.2 Survey instrument

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[14] (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.

2.4 Data analysis

The main body of this report presents results for the total sample. Country-level results are provided in Appendix Tables A.1 through A.6.

An item-level content validity index (I-CVI)[15, 16] shows the proportion of experts who rated each scale positively on each core question. The I-CVI is calculated by dividing the total number of 'positive' ratings

(i.e., ratings of 3 or 4) by the number of raters. At the 5% significance level, an I-CVI value ≥0.75 indicates 'excellent' content validity when there are ≥16 raters.[15] An average deviation (AD) index was used to measure the dispersion of responses around the median, with lower values indicating less dispersion.[17] At the 5% significance level with a 4-point response scale, AD index values ≤0.68 indicate 'acceptable and statistically significant agreement' when there are ≥15 raters.[17] 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.[18, 19] 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),[20] then refined iteratively as the comments were coded. The final template was applied across all comments.

3. Results

3.1 Sample characteristics

The overall response rate was 71% (25 completed/35 invited).[a] Most (72%) were psychiatrists or nurses; the remainder comprised a mix of disciplines. Experts represented the three types of expertise sought and, collectively, had used the HoNOS 65+ across a mix of settings. One-quarter said they had used the HoNOS OA in their work (Table 2).

Table 2. 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 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	24.1 (10.8)	2-42
Years worked with the HoNOS	13.9 (7.1)	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.

[a] Response rates were 71% (12 completed/17 invited) in Australia and 72% (13/18) in New Zealand.

3.2 Experts' ratings

Tables 3 and 4 show statistics summarising experts' responses to the 6 core questions about 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 1 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).

However, some aspects of content validity were more frequently endorsed than others. For example, 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.

Conversely, some HoNOS OA scales met the criterion for excellent content validity more often than others. For example, 3 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 further 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 criterion for one question (i.e., importance for determining overall clinical severity).

AD index values indicated acceptable and statistically significant agreement between experts, with 3 exceptions relating 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).

Table 3. 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. Bolded I-CVI values meet the criterion 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 4. 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. Bolded I-CVI values meet the criterion 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?

3.3 Experts' concerns

Experts' open-ended comments on their 'negative' ratings were organised into 9 themes - 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.

3.3.1 Themes related to comprehensiveness

3.3.1.1 Incomplete coverage

A recurring concern was that the rating descriptors for some scales 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*).

3.3.2 Themes related to comprehensibility

3.3.2.1 Lack of fit with clinical thinking

For some scales, experts identified that rating problems separately from the disorders with which they are associated might not fit with usual clinical thinking.

> "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*).

Experts also identified divergence from usual or contemporary conceptualisations for some clinical phenomena.

> "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*).

3.3.2.2 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*).

3.3.2.3 Ambiguity

Some experts indicated ambiguity in the glossary wording.

> "Ending it all is open to interpretation" (*Scale 2. Non-accidental self-injury*).
>
> "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*).

3.3.2.4 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*).
>
> "Including thought disorder with few specifiers is problematic" (*Scale 4. Cognitive problems*).
>
> "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*).
>
> "… it seems poorly specified. It could do with more actual examples" (*Scale 12. Problems with occupation and activities*).

3.3.2.5 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*).

3.3.3 Themes related to relevance

3.3.3.1 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*).

> "I think the functional anchor points are too gross to measure change but adequate for casemix." (*Scale 10. Problems with activities of daily living*).

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*).

> "Capturing change over time is problematic especially where more than one clinician involved (e.g. in-patient unit) and they don't agree on the most severe problem" (*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*).

3.3.3.2 Lack of relevance

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

> "These type[s] of problems are not an indicator for severity of mental illness - people without mental illness are subject to the same issues" (*Scale 12. Problems with occupation and activities*).

> "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*).

3.3.4 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*).

3.4 Experts' summary comments

Approximately half of all respondents (12/25) made some form of summary comment at the end of the survey. 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 the comprehensiveness of the glossary.

> "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."

4. Discussion

4.1 Summary of findings

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+,[21, 22] 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. With respect to comprehensiveness, for example, experts suggested additional older adult-specific examples for some scales, such as: 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 improve alignment between 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).

With respect comprehensibility, for example, one concern was that some scales might not reflect usual or contemporary clinical thinking about certain clinical problems. Specifically, 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) with which they are associated. These issues may have attracted comment because the revision increased the emphasis on rating these phenomena. For Scale 2 (Non-accidental self-injury), experts commented on how self-injury should be conceptualised. This may reflect, at least in part, an acknowledged lack of consistent terminology in the conceptualisation and description of non-accidental self-injury[23] and/or difficulties identifying non-accidental self-injury in older adults.[24]

4.2 Implications

Experts rated all HoNOS OA scales as important; this may give clinicians confidence in the measure's relevance 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 expand the older adult-specific examples might raise concerns about the utility of the HoNOS OA. Conversely, including 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 could helpfully orient experienced clinicians to the rationale for certain revisions, including the reduced emphasis on age-specific examples.

Some experts raised concerns about 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 diagnostic tool or screener.

4.3 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.[13, 25] We used standardised methods to determine whether excellent content validity and acceptable agreement among experts were reached.[16, 17, 26] The inclusion of a qualitative component enabled us to explore possible explanations for patterns in the experts' quantitative responses. However, some limitations should be noted. First, there may have been selection biases. We drew on multiple sources to identify experts and made efforts to confirm their expertise, but we did not apply measurable criteria.[27, 28] However in the survey, all experts self-identified at least one area of HoNOS expertise. Second, there may have been non-response bias, as 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 participating experts expressed a range of views, both positive and negative. Third, to minimise respondent burden, the open-ended questions focused on experts' concerns. Therefore, any interpretation of the findings should consider the qualitative and quantitative results in tandem.

4.4 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 could 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. Burgess P, Pirkis J, Coombs T. Routine outcome measurement in Australia. Int Rev Psychiatry. 2015;27(4):264-75.

7. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Res Social Adm Pharm. 2019;15(2):214-21.

8. Smith M, Baxendine S. Outcome measurement in New Zealand. Int Rev Psychiatry. 2015;27(4):276-85.

9. 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.

10. Royal College of Psychiatrists. Health of the Nation Outcome Scales (HoNOS) [Internet]. 2021 [cited 2021 22/04]. Available from: https://www.rcpsych.ac.uk/events/in-house-training/health-of-nation-outcome-scales.

11. World Health Organization. World Report on Ageing and Health. Geneva, Switzerland 2015. Available from: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf;jsessionid=84F1F08EA158E6029C38C833AC45567F.

12. 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.

13. 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.

14. 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.

15. 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.

16. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382-5.

17. Burke MJ, Dunlap WP. Estimating interrater agreement with the average deviation index: A user's guide. Organ Res Methods. 2002;5(2):159-72.

18. 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.

19. 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.

20. 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). Australia 2021. Available from: https://www.amhocn.org/publications/assessing-content-validity-revised-health-nation-outcome-scales-honos-2018-0.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25(5):508-18.

27. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health. 1997;20(3):269-74.

28. 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.

Appendix

Appendix Table A.1. Characteristics of experts who completed the survey, by country

	Australia (N = 12)		New Zealand (N = 13)	
	n	**%**	**n**	**%**
Main professional background				
Nurse	2	17	5	38
Psychologist	0	0	0	0
Clinical psychologist	2	17	1	8
Social worker	1	8	0	0
Psychiatrist	7	58	4	31
Psychiatric registrar	0	0	0	0
Occupational therapist	0	0	2	15
Other	0	0	1[a]	8
Expertise with HoNOS 65+ [b]				
Rating HoNOS 65+ or reviewing HoNOS 65+ ratings made by others	11	92	12	92
Research in the measurement properties of the HoNOS 65+ and/or measuring clinical effectiveness	1	8	2	15
HoNOS 65+ staff training and/or using HoNOS 65+ results at a macro level	6	50	9	69
Other expertise working with HoNOS 65+	2	17	2	15
Mental health settings worked with HoNOS 65+ [b]				
Inpatient	9	75	8	62
Residential	3	25	- [c]	-
Community	10	83	13	100
Other, non-clinical setting	0	0	1	8
Aware of HoNOS OA prior to survey				
No, I was not aware of the HoNOS OA at all	6	50	6	46
Yes, I was aware of the HoNOS OA, but have not used it in my work	2	17	4	31
Yes, I have used the HoNOS OA in my work	4	33	2	15
Not Sure	0	0	1	8
Other	0	0	0	0
	M (SD)	**Range**	**M (SD)**	**Range**
Years worked in mental health	24.4 (10.5)	4-40	23.8 (11.6)	2-42
Years worked with the HoNOS	16.3 (6.7)	4-25	11.8 (7.0)	2-28

HoNOS, Health of the Nation Outcome Scales. M, mean. N, number. 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.

Appendix Table A.2. Australian 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**	**Median**	**% positive**	**n**	**Range**	**Median**	**% positive**	**n**	**Range**	**Median**	**% positive**
Scale 1. Overactive or aggressive or disruptive or agitated behaviour	12	1-4	3.5	75	12	1-4	3	75	12	2-4	3	75
Scale 2. Non-accidental self-injury	12	2-4	3.5	83	12	2-4	3	67	12	1-4	3	67
Scale 3. Problem drinking or drug-taking	11	2-4	3	82	11	1-4	3	64	11	1-4	3	55
Scale 4. Cognitive problems	12	2-4	3	83	12	1-4	3	75	12	2-4	3	67
Scale 5. Physical illness or disability problems	12	2-4	3	92	12	2-4	3	67	12	1-4	3	58
Scale 6. Problems associated with hallucinations and /or delusions	12	2-4	3.5	92	12	2-4	3	92	12	2-4	3	92
Scale 7. Problems with depressed mood	12	2-4	4	92	12	2-4	3	83	12	1-4	3	75
Scale 8. Other mental and behavioural problems	11	2-4	3	91	12	1-4	3	67	12	2-4	3	83
Scale 9. Problems with relationships	12	2-4	3	83	12	2-4	3	67	12	2-4	3	83
Scale 10. Problems with activities of daily living	11	2-4	4	91	12	2-4	3	67	11	2-4	3	64
Scale 11. Problems with housing and living conditions	12	2-4	3	75	12	2-4	3	83	12	2-4	3	83
Scale 12. Problems with occupation and activities	12	1-4	3	83	12	1-4	3	67	12	2-4	3	58

% positive, percentage of ratings of 3 or 4. HoNOS, Health of the Nation Outcome Scales. n, number. [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?

Appendix Table A.3. Australian experts' ratings of the content validity of the HoNOS OA scales: comprehensibility

HoNOS OA scale	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?			
	n	Range	Median	% Positive	n	Range	Median	% Positive	n	Range	Median	% Positive
Scale 1. Overactive or aggressive or disruptive or agitated behaviour	12	2-4	3	92	12	2-4	2.5	50	10	1-3	3	60
Scale 2. Non-accidental self-injury	12	1-4	3	75	12	1-4	3	67	12	1-3	3	75
Scale 3. Problem drinking or drug-taking	11	2-4	3	82	11	1-4	3	64	11	2-3	3	64
Scale 4. Cognitive problems	12	1-4	3	83	12	1-4	3	67	12	1-3	2.5	50
Scale 5. Physical illness or disability problems	12	2-4	3	75	12	1-4	3	67	12	1-3	3	75
Scale 6. Problems associated with hallucinations and /or delusions	12	1-4	3	83	12	2-4	3	83	12	3-4	3	100
Scale 7. Problems with depressed mood	12	2-4	3	75	11	2-4	3	73	12	2-4	3	83
Scale 8. Other mental and behavioural problems	12	1-4	3	75	12	1-4	3	67	12	2-4	3	75
Scale 9. Problems with relationships	12	2-4	3	83	12	2-4	3	75	12	1-4	3	83
Scale 10. Problems with activities of daily living	12	2-4	3	58	12	1-4	3	67	12	2-4	3	83
Scale 11. Problems with housing and living conditions	12	2-4	3	83	12	2-4	3	75	11	2-4	3	82
Scale 12. Problems with occupation and activities	12	2-4	3	67	12	2-4	3	75	12	1-4	3	75

% positive, percentage of ratings of 3 or 4. HoNOS, Health of the Nation Outcome Scales. n, number. [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?

Appendix Table A.4. New Zealand 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**	**Median**	**% Positive**	**n**	**Range**	**Median**	**% Positive**	**n**	**Range**	**Median**	**% Positive**
Scale 1. Overactive or aggressive or disruptive or agitated behaviour	13	1-4	4	85	13	1-4	3	54	12	2-4	3	75
Scale 2. Non-accidental self-injury	11	2-4	4	91	12	1-4	3	67	11	1-4	2	27
Scale 3. Problem drinking or drug-taking	12	1-4	3	83	13	1-4	3	69	12	2-4	3	58
Scale 4. Cognitive problems	13	2-4	4	92	12	2-4	3	75	13	3-4	3	100
Scale 5. Physical illness or disability problems	13	2-4	4	85	13	1-4	3	85	13	1-4	3	92
Scale 6. Problems associated with hallucinations and /or delusions	12	2-4	3.5	92	11	1-4	3	82	12	2-4	3	83
Scale 7. Problems with depressed mood	13	3-4	4	100	13	1-4	3	85	13	2-4	3	69
Scale 8. Other mental and behavioural problems	13	1-4	3	92	13	1-4	3	77	13	1-4	3	69
Scale 9. Problems with relationships	13	2-4	4	85	13	2-4	3	69	13	1-4	3	54
Scale 10. Problems with activities of daily living	13	3-4	4	100	13	1-4	3	85	13	2-4	3	77
Scale 11. Problems with housing and living conditions	13	1-4	3	85	12	1-4	3	92	13	1-4	3	69
Scale 12. Problems with occupation and activities	13	1-4	3	77	13	2-4	3	85	13	1-4	3	77

% positive, percentage of ratings of 3 or 4. HoNOS, Health of the Nation Outcome Scales. n, number. [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?

Appendix Table A.5. New Zealand experts' ratings of the content validity of the HoNOS OA scales: comprehensibility

HoNOS OA scale/section	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?			
	n	Range	Median	% Positive	n	Range	Median	% Positive	n	Range	Median	% Positive
Scale 1. Overactive or aggressive or disruptive or agitated behaviour	13	2-4	3	77	13	2-4	3	69	13	2-4	3	77
Scale 2. Non-accidental self-injury	13	1-4	3	69	12	2-4	2	42	12	1-4	3	67
Scale 3. Problem drinking or drug-taking	13	2-4	3	69	12	2-4	3	75	13	2-4	3	92
Scale 4. Cognitive problems	13	2-4	3	92	12	2-4	3	83	13	2-4	3	85
Scale 5. Physical illness or disability problems	13	2-4	4	92	13	2-4	4	92	13	2-4	3	92
Scale 6. Problems associated with hallucinations and /or delusions	12	2-4	3	75	12	2-4	3	75	12	2-4	3	92
Scale 7. Problems with depressed mood	13	2-4	3	77	13	2-4	3	85	13	2-4	3	92
Scale 8. Other mental and behavioural problems	13	2-4	3	69	13	1-4	3	69	12	2-4	3	75
Scale 9. Problems with relationships	13	2-4	3	69	13	2-4	3	62	13	2-4	3	77
Scale 10. Problems with activities of daily living	13	2-4	3	92	12	2-4	3	92	13	2-4	3	85
Scale 11. Problems with housing and living conditions	13	1-4	3	77	13	1-4	3	85	13	1-4	3	92
Scale 12. Problems with occupation and activities	13	1-4	3	62	12	1-4	3	75	13	2-4	3	77

% positive, percentage of ratings of 3 or 4. HoNOS, Health of the Nation Outcome Scales. n, number. [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?

Appendix Table A.6. Summary of themes identified through the qualitative assessment, by country

Themes	Australia	New Zealand
Experts' concerns		
Incomplete coverage	✓	✓
Lack of fit with clinical thinking	✓	✓
Too many phenomena	✓	✓
Ambiguity	✓	✓
Need for more description or examples	✓	✓
Assessment challenges	✓	✓
Challenges to capturing change	✓	✓
Lack of relevance	✓	✓
Need for training	✓	✓
Experts' summary comments		
Updated title is an improvement	✓	✓
Measure not improved	✓	
Positive views about comprehensiveness		✓
Negative views about comprehensiveness	✓	✓