The Ms. Olsen test: Measurement properties of a short test of nursing staffs' competence in clinical decision-making

Bing-Jonsson, Pia Cecilie[1]; Boman, Erika[2]; Melin, Jeanette[3,4]

[1]Faculty of Health and Social Sciences - University of South-Eastern Norway
[2]Åland University of Applied Sciences - Åland University of Applied Sciences
[3]University of Gothenburg - University of Gothenburg
[4]RISE Research Institutes of Sweden - RISE Research Institutes of Sweden

This is the peer reviewed version of the following article:

Bing-Jonsson, P. C., Boman, E. & Melin, J. (2021). The Ms. Olsen test: Measurement properties of a short test of nursing staffs' competence in clinical decision-making. *Journal of Advanced Nursing, 77*(10), 4268-4278. https://doi.org/https://doi.org/10.1111/jan.15025

which has been published in final form at https://onlinelibrary.wiley.com/doi/10.1111/jan.15025.

ABSTRACT

Aims: To assess the measurement properties of the Ms. Olsen test for registered nurses and assistant nurses, respectively, and suggest cut-off points between competence levels.

Design: Cross-sectional study. The results were analysed by implementing the Rasch Measurement Theory.

Methods: Nursing staff working in various health care settings participated (n = 757). To measure the competence of nursing staff in clinical decision-making, a 19-item scale from the Nursing Older People-Competence Evaluation Tool—the 'Ms. Olsen test'—was used. Data were collected in October 2017, 2018 and 2019.

Results: The Ms. Olsen test showed reasonably good measurement properties for registered nurses and assistant nurses, respectively. Results show slightly better measurement properties for registered nurses than for assistant nurses. The cut-off for registered nurses, 0.62, corresponds to managing approximately two-thirds of the items while, for assistant nurses, the cut-off of 0.01 corresponds to managing approximately half of the items.

Conclusion: The Ms. Olsen test is a short (7- to 10-minute) test measuring competence in clinical decision-making among nursing staff working in older people nursing. Despite reasonably good measurement properties, this should be considered an initial validation in the development of a short test for assessing clinical decision-making among nursing staff in various health care setting.

Impact: Several scales aiming to measure nursing competence have been developed over the last decade, but measurement properties (beyond classical test theory) are seldom evaluated, few scales concern other staff groups than registered nurses and few scales have proposed or established cut-offs for safe practice. The Ms. Olsen test is a short test of clinical decision-making that demonstrates reasonably good measurement properties. Cut-off points for registered nurses and assistant nurses were established. The Ms. Olsen test may be used to measure and evaluate competence in clinical decision-making among nursing staff working in older people nursing and educational settings.

Key words: competence measurement, clinical decision-making, measurement properties, nursing, assistant nurse, older people, Rasch Measurement Theory, registered nurses

INTRODUCTION

Nursing staff competence in clinical decision-making is a key element of quality of care in older people nursing, as in nursing in general (Aiken et al., 2014; Kiljunen et al., 2017; Shaheen et al., 2019). The measurement of nursing staff competence is inherently linked to systematic evaluation and the improvement of quality of care (Donabedian & Bashshur, 2003). To measure competence in older people nursing, the Nursing Older People-Competence Evaluation Tool (NOP-CET) has been developed. The NOP-CET is a self-assessment scale containing 65 items (Bing-Jonsson, Hofoss, et al., 2015). It differentiates from other self-assessment scales by using test items with response categories that are either correct or incorrect: for example, 'Which is the correct procedure for resuscitation?', or 'What is the desired level of blood glucose in patients with diabetes?' The responses to the test items can be analysed according to predefined score sheets, and thus it is possible to rank individuals or groups according to the number of correct responses (test scores).

The development and psychometric evaluation of the NOP-CET is thoroughly described in Bing-Jonsson, Bjørk, et al. (2015) and Bing-Jonsson, Hofoss, et al. (2015). The latter study concluded that the NOP-CET showed acceptable content and construct validity, reliability, precision, interpretability, acceptability, and feasibility according to the principles of classical test theory. However, to be able to translate the test scores into clinically meaningful units, it is necessary to further understand its measurement properties (Cano, 2019) and to decide on defensible cut-off point(s) for what indicates clinically sound competence (Azzarello, 2003). In this study, the focus of interest is on competence in clinical decision-making.

After the construction and initial validation of the NOP-CET, feedback from end-users (like Nurse leaders in older people nursing) was that the NOP-CET was too long, too time-consuming to fill in, and resulted in an unmanageable amount of results. The Nurse leaders expressed a need for concrete and manageable results on the nursing staff's competence in clinical decision-making. Careful review of the NOP-CET revealed questions no. 16 and 17, concerning a fictual patient called Ms. Olsen, to be especially relevant in the Nurse leaders' opinion. Questions no. 16 and 17, concerning Ms. Olsen, was developed to cover five of the ten categories of which the NOP-CET was based on (Bing-Jonsson, Bjørk, Hofoss, Kirkevold, & Foss, 2015). These five categories were: "treatment", "assessment and taking action", "cover basic needs", "responsibility and activeness", and "cooperation" (Bing-Jonsson, 2015, p. 41). Question no. 16 and 17, which consisted of nineteen single items, can

be considered an extraction from the NOP-CET, and not a short-version acompassing the totality of what the NOP-CET was intended to measure. The nineteen extracted items which concern clinical decision-making, has been called the Ms. Olsen test, and was pilot-tested on a smaller sample of nursing staff in older people nursing as initial validation (Hopøy, Bakken, & Bing-Jonsson, 2020).

In this study, nineteen items extracted from the NOP-CET concerning clinical decision-making, called the Ms. Olsen test, were explored. To evaluate the measurement properties, the Ms. Olsen test was analysed based on the Rasch Measurement Theory (RMT) for registered nurses (RNs) and assistant nurses (ANs), respectively, and subsequently cut-off points were suggested.

Background

Nursing staff competence in clinical decision-making is a key element in the delivery of safe and high-quality care (Bing-Jonsson, 2016; Morphet et al., 2015; Recio-Saucedo et al., 2018). Generic nursing competence is needed in older people nursing, but there are also specific competence requirements. For example, the nursing staff must possess insight into the ageing process, the diversity of the older population and their health and social needs (Bing-Jonsson, Bjørk, et al., 2015). The nursing staff must be able to apply this knowledge when collecting, interpreting and analysing information in order to make appropriate decisions regarding the older person. They must also continuously create, disseminate, apply and translate their knowledge about older people in their own area of expertise (Dijkman et al., 2016). Moreover, as older people are encountered in a range of contexts, there is a need to develop and assess nurses' clinical competence within their specific context of practice (Lejonqvist et al., 2012).

There are several scales that can assess generic competence at varying stages of nurses' careers; of these, the Nurse Competence Scale (NCS) (Meretoja & Leino-Kilpi, 2003; Meretoja et al., 2004) is the most widely used (Flinkman et al. 2017). However, most generic competence scales are not applied in older people nursing. Some studies have explored nursing professionals' self-assessed competence in relation to current competence requirements in older people nursing: specifically, in nursing and care homes (e.g., Bing-Jonsson et al., 2016; Kiljunen et al., 2019), in municipal care (Karlstedt et al., 2015) and in nursing education (Tohmola, Saarnio, Mikkonen, Kyngäs, & Elo, 2020). To our best knowledge, the NOP-CET is the only scale to assess general competence in older people

nursing throughout the nursing career, including competence in clinical decision-making (the Ms. Olsen test). The NOP-CET, including the Ms. Olsen test, is assumed to have international relevance, as competence demands in older people nursing are similar across borders. Prior research has largely focused on the competence of RNs. Nevertheless, we would argue that the competence of all groups of nursing staff is important to enable the delivery of safe health care to older people. The Ms. Olsen test was therefore designed to measure competence in several groups of nursing staff, including assistant nurses (ANs) and RNs.

THE STUDY

Aim

The aim of the study was to assess the measurement properties of the Ms. Olsen test for registered nurses and assistant nurses, respectively, and suggest cut-off point(s) between competence levels.

Design

A cross-sectional study design was used. The results were analysed using the Rasch Measurement Theory (RMT).

Sample/Participants

Nursing staff working in various health care settings were invited to participate. In Norway, nursing staff mainly consist of registered nurses (RNs), who have a three-year bachelor's degree (or higher), assistant nurses (ANs) who have three years of secondary education, and assistants who have no formal qualifications in health care. In the analyses, as their clinical duties and competence demands are very similar, assistant nurses and assistants were ultimately collapsed into one group labelled 'assistant nurses'(ANs). The nursing staff came from various settings, ranging from emergency departments and hospital wards to nursing homes and home care—all settings with many older patients. In total, 757 nursing staff participated (RNs, n = 495; ANs, n = 262).

Instrument

To measure the competence of nursing staff in clinical decision-making, a 19-item scale (the Ms. Olsen test) from the NOP-CET was used (Bing-Jonsson, Bjørk, et al., 2015; Bing-

Jonsson, Hofoss, et al., 2015). The scale starts with a short description of a typical, but fictional patient in most health care settings: 'Ms. Olsen is 90 years old and generally weakened by age. Imagine that she develops the following symptoms. Please choose how you would respond when Ms. Olsen, your patient, develops the following symptoms. You may choose one option on each line'. Nineteen test items follow, each with descriptions of symptom(s) that Ms. Olsen develops. All 19 test items are listed in Table 1. The respondents may choose between the following response categories for each test item: 1) *no action required*, 2) *observe again the following day*, 3) *consult with an RN*, 4) *nursing-related measure required immediately*, 5) *have patient assessed by physician*, and 6) *requires acute help in hospital*.

PLACE TABLE 1 ABOUT HERE

During the instrument's development in 2013, a score sheet was established by an expert group of three well-respected experts in geriatrics and older people nursing in Norway (one physician and two RNs) (Bing-Jonsson, Hofoss, et al., 2015; Bing-Jonsson et al., 2016). When scoring the responses, response option 1 ('no actions required') was never correct (scored as 0), one or two of the other response options could be the best option depending on the nature of Ms. Olsen's symptom (scored as 2) and the remaining options were neither wrong nor optimal (scored as 1). As their scopes of practice differ, RNs and ANs may have a differing best options for certain items. Please see Appendix 1 for the score sheet containing answers for each item for the RNs and ANs, respectively.

In addition to the Ms. Olsen test, three demographic variables were collected: professional group (RN, AN, assistant), work experience (number of years), and time since last completed education (number of years).

Data collection

Data collection was conducted by master's students in Advanced Practice Nursing at a university in Southern Norway, who were taking a course in research methodology. The master's students all worked alongside their studies; in total 39 (18 + 13 + 8) students were assigned the task of inviting all nursing staff (total number is unknown) at their workplaces to participate in the study and respond to the Ms. Olsen test. The selection criterium for sample inclusion was that the person concerned was a colleague (nursing staff) in the same workplace as the master student. Nursing staff includes RN, AN and assistants. Data were collected in October 2017, 2018 and 2019, and resulted in 757 responses. The sample sizes

for both groups, RNs (n = 495) and ANs (n=262) corresponds recommended sample size for item measures stable ± ½ logit and high-stake decisions (Linacre, 1994) and to provide a good balance for statistical interpretations of fit statistics in RUMM (Hagell & Westergren, 2016; Linacre, 1994). A review of the master students' workplaces over the three consecutive years gave 21 unique workplaces. Of the 21 workplaces one had students in 2017 and 2019, and one had students in all three years. It is therefore possible that some respondents have filled in the Ms. Olsen test more than once, but never in the same year.

The test was administered electronically via Questback and was estimated to take 7 to 10 minutes to fill out. Nursing staff invited to take part in the test received an SMS and/or an e-mail with a link to the test. No personal reminders were sent to the invited participants, but the master's students and heads of departments encouraged their colleges and employees to participate several times during the months in which the test was open.

Ethical considerations

Participation in the test was voluntary and confidential. The master's students who collected the data requested permission from the leaders of their workplaces to distribute the test. All nursing staff received written information about the test and were informed that completing and returning the test was synonymous with informed consent. As the test did not include any items that could identify the respondent, research approval from Norwegian Social Science Data Services was not necessary.

Data analysis

The Ms. Olsen test was analysed using the RMT, a metrological method developed by Danish mathematician Georg Rasch in 1960, based on the same underlying principles as physical measurements (Rasch, 1960). As such, with the RMT, separate values of person and item attributes are estimated and scaled on the same interval logit scale. In the simplest, dichotomous case, it is a logistic regression function:

$$\log\left(\frac{P_{success,i,j}}{1-P_{success,i,j}}\right) = \theta_i - \delta_j \qquad (1)$$

where θ is the person attribute value (hereafter person competence) and δ is the item attribute value (here after task difficulty). The logistic regression function estimates the difference between the person competence and the task difficulty. Thus, persons with high competences

are expected to choose high scoring responses for each item, whereas persons with low competences are expected to consistently choose low scoring responses.

The software Rasch Unidimensional Measurement Model 2030 (RUMM) was used for the analyses with a partial credit model, i.e. allowing all items to have its own rating scale structure. The two sets of RNs and ANs were analysed separately and structured around three questions outlined by Hobart and Cano (2009):

1. Is the scale-to-sample targeting adequate for making judgements about the performance of the scale and the measurement of people? This was assessed via scale-to-sample targeting of item locations, i.e. persons and item locations should be equally distributed (normally ranging from -3.00 to 3.00 logits). The mean of item locations is always 0.00, thus, if the mean person location is higher or lower than 0 it indicates whether the sample is off centred from the items.
2. Are the people in the sample being successfully measured? For assessments of successfully measured persons, the item-person distributions, Person Separation Index (PSI) and person fit residuals. We used the following guidance:
    a. The PSI is a reliability indicator, where 0 implies all error and 1 implies no error.
    b. The person fit residual should ideally lie within -2.5 to +2.5.
3. Has a measurement ruler successfully been constructed? For assessment of successfully constructed rulers, response categories were assessed if they worked as intended and items were judged according to their clinically logical order and by assessment of fit residuals, chi-square, item characteristic curve (ICC), residual correlations and unidimensionality. We used the following recommendations:
    a. The thresholds should show monotonicity and be sequentially ordered (Andrich, 1978).
    b. A priori, an independent nurse who has been involved in the early development of the Ms. Olsen test categorized each item difficulty level as either easy, middle or hard for RNs and ANs, respectively. Those categorizations were used to assess how well the theory and the estimated task difficulty values corresponded.
    c. The individual item fit residuals should ideally lie between -2.5 and +2.5; the chi-square values should not be statistically significant (Bonferroni correction

was applied); and the dots of the class intervals should follow the ICC to support good fit.

d. Residual correlations above a relative cut-off greater than 0.20 above the average correlations indicate local dependency (Marias, 2013).
e. The first factor in the principal component analysis of the residuals was used to divide items into two subsets (positively and negatively correlated items). Person abilities for each subset were then compared by an independent t-test where the percentage of persons outside the range of -1.96 to 1.06 should not exceed 5% to support unidimensionality (Smith, 1996).

Several techniques exist to determine cut-off scores. In line with Azzarello's (2003) emphasis on the importance of making a rational and defensive cut-off point between different levels of competence, we build on the Contrasting Groups Method, with which distributions of known persons with low and high competence, respectively, were explored. We started with samples of at least 15 persons (Azzarello, 2003), i.e., the 15 persons with the lowest competence measures and the 15 persons with the highest competence measures. Subsequently, we assessed the expanded measurement uncertainties ($k = 2$) and set the initial cut-off at the mean location between the uncertainty ranges. Furthermore, the proposed cut-offs were also judged based on their clinical implications and relevance (Azzarello, 2003).

Rigour

Psychometric evaluation of the NOP-CET is thoroughly described in Bing-Jonsson, Hofoss, et al. (2015). The NOP-CET demonstrates acceptable content and construct validity, reliability, precision, interpretability, acceptability, and feasibility according to the principles of classical test theory. The Ms. Olsen test was pilot tested on a small sample and concluded that further testing should include larger samples like this study (Hopøy et al., 2020). Internal and external validity was sought enhanced through sequential data collection over three consecutive years ensuring relevance over time, and a sample that encompassed a variety of settings for older people nursing ensuring relevance across settings. Selection bias may however not be ruled out as there may have been preexisting differences in clinical decision-making between groups in different settings.

To strengthen the reporting we followed the STROBE guidelines.

RESULTS

The results section presents the measurement properties of the Ms. Olsen test for RNs and ANs, respectively, followed by proposed cut-off scores.

Measurement properties of the Ms. Olsen test for registered nurses (RNs)

There were very few scorings (80 of the total 9405, or <1%) of *no action required* (which is always wrong and scored as 0). As such, 0 and 1 were collapsed, even though disordered thresholds were only present for one item (Q10). This significantly improved the targeting—the person competence mean value was changed from 2.44 to 0.86 logits—and had no major impact on the other fit statistics.

Figure 1 shows a Rasch histogram with person competence values on the top ranging from low to high abilities and task difficulty values on the bottom ranging from easy to more challenging items. There are some gaps in the items, especially among the more challenging items. This implies that RNs with higher competencies have larger measurement uncertainties and are measured with less precision (PSI = 0.45). One person showed a fit residual outside the desirable range of $\pm$2.5, but 2.57 is negligible. No persons showed extreme values.

PLACE TABLE 2 AND FIGURE 1 ABOUT HERE

Item locations ranged from -2.67 (2SE 0.48) to 2.81 (2SE 0.26), where the easiest item (Q10) and the most difficult item (Q3) corresponded to the theoretical classification (Table 2). Item Q11 had a high task difficulty value (0.92, 2SE 0.19) when compared to a theoretically classified easy item, whilst items Q7, Q15 and Q18 were less challenging than classified. As shown in Table 3, no items had fit residuals outside the desirable range of $\pm$2.5 and no significant chi-square values. However, items Q4 and Q15 showed dots for the class intervals, deviating from their ICC.

Local dependency was only present for 3 out of 171 residual correlations: items Q14, Q15 and Q16, where all were above the relative cut-off of 0.15. This is likely due to the nature of Ms. Olsen's symptoms in those items where acute symptoms require acute help in the hospital compared to the other items with less acute symptoms. The PCA of the residuals divided items into two subsets of items (Table 3) used for the t-test for unidimensionality, giving 6.26% outside the desired range of $\pm$1.96.

Measurement properties of the Ms. Olsen test for assistant nurses (ANs)

The same low scorings of 0 were also present for ANs—83 of 4978 (1.6%)—and scores of 0 and 1 could be collapsed with improved targeting (mean person competence improved from 1.86 to -0.07) and no meaningful impact on the other fit statistics.

Figure 2 shows a Rasch histogram for the ANs. Similar to what Figure 1 shows for RNs, items are close to each other but do not completely cover the persons with the lowest and highest competences, respectively. The reliability was low (PSI = 0.50), but there were no extremes and only two persons with undesirable fit residuals (2.51 and 2.64).

PLACE TABLE 3 AND FIGURE 2 ABOUT HERE

As shown in Table 3, one (Q14) of the two easiest items was correctly theoretically defined as an easy item, but not the other one (Q11). The spread among the other items were limited, but some of the theoretically most challenging items were observed as being much easier (Q1, Q15 and Q18). There were no fit residuals outside the desirable range of $\pm$2.5 and no significant ChiSq values, at the same time, misfit was indicated by several items dots for the class intervals deviated from their ICC. Local dependency was present for 6 out of 171 residual correlations and the t-test had 8.02% outside the desired range of $\pm$1.96.

Cut-offs for the Ms. Olsen test

For RNs, the 15 persons with the lowest competence were within the range of -2.52 to -0.70, and a further one person also had a competence of 0.70. For this group, the upper limit of measurement uncertainty was 0.38. At the other end of the scale, the 15 persons with the highest competence ranged from 2.55 to 2.34, and a further three persons also had a competence of 2.55. For those with the high competence, the lower limit of measurement uncertainty was 1.00. Consequently, the initial cut off was set to 0.62: i.e., the person should have a 50% chance of a correct answer on an item at that location corresponding to the two closest items. As shown in Table 2, a cut-off of 0.62 falls in between Q5 at 0.25$\pm$0.20 and Q11 at 0.92$\pm$0.19. From a clinical point of view, this implies that the person has to pass on approximately two-thirds of the items, including most of the items predefined as easier. Due to the large measurement uncertainties, only 67 (14%) of RNs could be classified as either below or above the cut-off without measurement noise, while the others had measurement uncertainties overlapping.

The same assessment of ANs resulted in 24 persons in the group with the lowest competence (range -2.17 to -1.09), with the upper limit of measurement uncertainty at 0.00, and 18

persons in the group with the highest competence (range 1.10 to 2.14), with the lower limit of measurement uncertainty at 0.02. In turn, this gave an initial cut-off at 0.01 corresponding to a 50% chance of managing 8 out of 19 items—which, from a clinical perspective, can be seen as a minimum requirement rather than a cut-off for safe clinical decision-making. It was only the 42 (16%) persons with the lowest and highest competence, respectively, who did not have overlapping measurement uncertainties.

DISCUSSION

As nursing staff competence in clinical decision-making is a key element in quality of care, the measurement of clinical decision-making is linked to the systematic evaluation and improvement of quality of care. Research by Aiken et al. indicates that competent nursing staff have a positive impact on quality of care in hospital care settings (Aiken et al., 2014; Aiken et al., 2017) and in the geriatric context (e.g., Castle & Anderson, 2011; Shaheen et al., 2019). Several scales aiming to measure nursing competence have been developed over the last decade, but according to our best knowledge, measurement properties (beyond classical test theory) are seldom evaluated, few scales concern other staff groups than RNs and few scales have proposed or established cut-offs for safe practice. In order to translate test scores into clinically meaningful units, it is necessary to further understand the measurement properties and to decide on defensible cut-off points for what indicates clinically sound competence, which this study aim to contribute.

Measurement properties

The Ms. Olsen test demonstrated reasonably good measurement properties for RNs and ANs, respectively. However, this should be considered only as an initial validation in the development of a short test for assessing clinical decision-making among nursing staff in various health care setting. Our results show slightly better measurement properties in terms of item-fit to the model and a priori categorisation of easy, middle, and hard items for RNs than for ANs. This is understandable, as the Ms. Olsen test measures clinical decision-making concerning relatively complex symptoms, which is the core of nursing. ANs encounter the same complex symptoms as RNs, but do not have the same level of educational preparedness for responding to these symptoms. However, there were several items that the ANs found easier than expected: Q11, Q18, Q1 and Q15. These are all symptoms that require quick

responses—for example, Q15 (Ms. Olsen has changes in sight, hearing, speech and comprehension)—and may indicate adequate clinical decision-making in acute situations among ANs.

Thanks to the RMT's unique properties of scaling person and item attributes values on the same interval logit scale, we are provided with actionable information about which items require more training and/or nursing development activities: for example, a person with a location of 0 logits should first focus on practising skills in the close to 0 logits and then move up to more challenging tasks. For an RN, this corresponds to first focusing on activities related to tasks in Q7 (Ms. Olsen has reduced appetite and food intake), Q15 (Ms. Olsen has changes in sight, hearing, speech and comprehension) and Q18 (Ms. Olsen has short attention span and delusions), rather than focusing on more complex and challenging items.

In this work we tested unidimensionality by Smiths t-test, which showed more than desired persons outside the $\pm$1.96 range. However, the t-test should not, as claimed by Hagell (2014) be *viewed as a "definite" test of unidimensionality and does not replace an integrated quantitative/qualitative interpretation based on an explicit variable definition in view of the perspective, context and purpose of measurement*. Thus, we also assessed the item fit as well as how well items a priori expected to be easy, middle, or hard items. In this initial study, we find the support for unidimensionality acceptable, but would encourage forthcoming research – after refinement and in larger sample – to further assess the unidimensionality.

In this study, we assessed RNs and ANs who are already in practice; as such, it was expected that many of them would already have higher abilities and be highly qualified in their work. An additional field for use of the Ms. Olsen test would therefore be to study learning progression during nursing education and other preparatory activities. In such cases, we would expect that beginners would have lower abilities and that, toward the end of their education, they will have acquired new and more challenging abilities.

Defensible cut-off points

The cut-off for RNs (0.62) corresponds to pass on approximately two-thirds of the items, while for ANs the cut-off was approximately half of the items (0.01). The initial assessments of cut-off scores should be interpreted with some caution—rather than used as definite defensible cut-off points—due to the low reliability and large measurement uncertainties in this study. However, for both RNs and ANs, with the expanded measurement uncertainties

(i.e., $k = 2$, corresponding to a 95% confidence interval) not overlapping each other, we could separate two distinct groups of persons with low and high competence, respectively. The Contrasting Groups Method that we used presupposes that the sample is representative in terms of various levels of competence. Fair normal distributions were evident for both RNs and ANs' competences (see upper pink bars in the histograms in Figures 1 and 2), which is what would be assumed in clinical settings; however, this still needs additional testing. Such testing is recommended to also include aspects such as requirements for safe clinical decision-making. In addition to this initial assessment of cut-off scores, there is a risk of consequences of miss-classifications, i.e., false positives or false negatives (Azzarello 2003). The ranking of individuals that follows a measurement with cut-off points may, however, have ethical implications. Results from a competence measurement can have consequences if a measure is developed to rank individuals or classify someone as competent or incompetent. To ensure high measurement precision, we would recommend further evaluations of defensible and optimal cut-offs with additional samples as well as extended clinical judges, as recommended by Azzarello (2003).

Despite recommendations of further evaluations of defensible and optimal cut-offs, in Appendix 2 we provide a version 0.1 with transformed raw scores to logits and associated cut-offs.

Limitations

This is an initial validation of the Ms. Olsen test and there are some limitations to bear in mind when interpreting the results. In particular, more challenging items are needed to improve the targeting for RNs and ANs with high abilities if the test should be used to monitor change. Moreover, the PSIs were low for both RNs and ANs, which implies that there is a high probability that estimated person abilities is not a reproducible person competence value (Fisher 1992; Wright 1996). A low reliability often relates to there being too few items to cover the whole sample. By using the Spearman-Brown prophecy formula (Spearman, 1910), a total of 95 and 76 items (respectively) are required to reach a reliability of 0.8. However, this many items might not be feasible in practice unless computer-adaptive testing is used. Another option could be to further consider other ways of scoring that potentially gives thresholds spread across the continuum. Despite the fact that we had to collapse 0 and 1 in this study to improve the targeting and reduce the measurement uncertainties, one must nevertheless determine whether any person has scores of 0 in their

test scores using a ‘kid map’: this might be an indication that this person lacks necessary nursing competencies, as no action is never a good action. A further limitation is that we could not properly assess differential item functioning (DIF), i.e., if persons from different groups have different expected values for their responses, due to lack of background characteristics to create relevant and meaningful groups. This is warranted in forthcoming studies, e.g., between workplaces and length of work experience from working with older people nursing.

CONCLUSION

The Ms. Olsen test is a short test that takes 7 to10 minutes and measures clinical decision-making among nursing staff working in older people nursing. The Ms. Olsen test was extracted from a larger measurement instrument called the NOP-CET on request from nursing leaders. There was an expressed need for concrete and manageable results on the nursing staff’s competence in clinical decision-making. The Ms. Olsen test was analysed using the RMT, and demonstrated reasonably good measurement properties for RNs and ANs, respectively. This should be considered as an initial validation in the development of a short test for assessing clinical decision-making among nursing staff in various health care setting. Scaling person and item attribute values on the same interval logit scale provides actionable information about which competencies require more training and/or nursing development activities. Despite somewhat large measurement uncertainties, two distinct groups of persons with low and high competence (respectively) can be separated based on the initial cut-offs proposed. An implication for nursing practice is thus that a short test of nursing staffs’ competence in clinical decision-making may be explored further in order for easy and reliable measurement of clinical decision-making. Future development of this initial validation study of the Ms. Olsen test will also include further evaluations of defensible and optimal cut-offs, to ensure safe clinical decision-making. In this study, the Ms. Olsen test was tested on nursing staff working in various health care settings in older people nursing. Further development of the Ms. Olsen test would be to analyse the measurement properties of testing learning progression during nursing education and other preparatory activities for older people nursing.

Conflict of Interest statement

There are no conflicts of interest.

References

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M. D., Moreno-Casbas, M. T., Rafferty, A. M., Schwendimann, R., Scott, P. A., Tishelman, C., van Achterberg, T., Sermeus, W., & RN4CAST Consortium (2014). Nurse staffing and education and hospital mortality in nine European countries: A retrospective observational study. Lancet, 383(9931), 1824–1830. https://doi.org/10.1016/S0140-6736(13)62631-8

Aiken, L. H., Sloane, D., Griffiths, P., Rafferty, A. M., Bruyneel, L., McHugh, M., Maier, C. B., Moreno-Casbas, T., Ball, J. E., Ausserhofer, D., Sermeus, W., & RN4CAST Consortium (2017). Nursing skill mix in European hospitals: Cross-sectional study of the association with mortality, patient ratings, and quality of care. BMJ Quality & Safety, 26(7), 559–568. https://doi.org/10.1136/bmjqs-2016-005567

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43(4), 561–573. https://doi.org/10.1007/BF02293814

Azzarello, J. (2003). Setting valid cut-off scores for competency. Journal for Nurses in Staff Development—JNSD, 19(5), 238–243.

Bing-Jonsson, P. C. (2015). Nursing Older People - Competence Evaluation Tool. Development and psychometric evaluation of an instrument measuring community-based nursing staff competence. (PhD). University of Oslo, Oslo. Retrieved from https://www.duo.uio.no/handle/10852/43201

Bing-Jonsson, P. C., Bjørk, I. T., Hofoss, D., Kirkevold, M., & Foss, C. (2015). Competence in advanced older people nursing: development of 'Nursing older people - Competence evaluation tool'. International Journal of Older People Nursing, 10, 59-72. doi:10.1111/opn.12057

Bing-Jonsson, P. C., Hofoss, D., Kirkevold, M., Bjørk, I. T., & Foss, C. (2015). Nursing older people—competence evaluation tool: Development and psychometric evaluation. Journal of Nursing Measurement 23(1), 127–153. https://doi.org/10.1891/1061-3749.23.1.127

Bing-Jonsson, P. C., Hofoss, D., Kirkevold, M., Bjørk, I. T. & Foss, C. (2016). Sufficient competence in community elderly care? Results from a competence measurement of nursing staff. BMC Nursing, 15(5). https://doi.org/10.1186/s12912-016-0124-z

Cano, S. J., Pendrill, L. R., Melin, J., Fisher, W. P. (2019). Towards consensus measurement standards for patient-centered outcomes. Measurement, 141, 62–69. https://doi.org/10.1016/j.measurement.2019.03.056

Castle, N. G., & Anderson, R. A. (2011). Caregiver staffing in nursing homes and their influence on quality of care: Using dynamic panel estimation methods. Medical care, 49(6), 545–552. https://doi.org/10.1097/MLR.0b013e31820fbca9

Dijkman, B., Roodbol, P., Aho, J., Achtschin-Stieger, S., Andruszkiewicz, A., Coffey, A., Felsmann, M., Klein, R., Mikkonen, I., Oleksiw, K., Schoofs, G., Soares, C., & Sourtzi, P. (2016). European Core Competences Framework for Health and Social Care Professionals Working with Older People. Retrieved from http://ellan.savonia.fi/

Donabedian, A., & Bashshur, R. (2003). An introduction to quality assurance in health care. Oxford University Press.

Fisher, W (1992). Reliability, separation, strata statistics. Rasch Measurement Transactions, 6(3), 238.

Flinkman, M., Leino-Kilpi, H., Numminen, O., Jeon, Y., Kuokkanen, L., & Meretoja, R. (2017). Nurse Competence Scale: A systematic and psychometric review. Journal of Advanced Nursing, 73(5), 1035–1050.

Hagell, P. (2014). Testing Rating Scale Unidimensionality Using the Principal Component Analysis (PCA)/t-test Protocol with the Rasch Model: The Primacy of Theory over Statistics. Open Journal of Statistics, 04(06), 456–465. doi:https://doi.org/10.4236/ojs.2014.46044

Hagell, P., & Westergren, A. (2016). Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. J Appl Meas, 17(4), 416-431.

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. Health Technology Assessment, 13(12), iii–177. https://doi.org/10.3310/hta13120

Hopøy, A.-C., Bakken, L. N., & Bing-Jonsson, P. C. (2020). A survey of clinical competence among health personnel in care homes: a pilot study. Sykepleien forskning, 15. doi:10.4220/Sykepleienf.2020.83096en

Karlstedt, M., Wadensten, B., Fagerberg, I., & Pöder, U. (2015). Is the competence of Swedish Registered Nurses working in municipal care of older people merely a question of age and postgraduate education? Scandinavian Journal of Caring Sciences, 29(2), 307–316.

Kiljunen, O., Välimäki, T., Kankkunen, P., & Partanen, P. (2017). Competence for older people nursing in care and nursing homes: An integrative review. International Journal of Older People Nursing, 12(3). https://doi.org/10.1111/opn.12146

Kiljunen, O., Partanen, P., Välimäki, T., & Kankkunen, P. (2019). Older people nursing in care homes: An examination of nursing professionals' self-assessed competence and its predictors. International Journal of Older People Nursing, 14(2), e12225. https://doi.org/10.1111/opn.12225

Lejonqvist, G. B., Eriksson, K., & Meretoja, R. (2012). Evidence of clinical competence. Scandinavian Journal of Caring Sciences, 26(2), 340–348. https://doi.org/10.1111/j.1471-6712.2011.00939.x

Linacre, J. M. (1994). Rasch Measurement Transactions. In Sample Size and Item Calibration Stability (Vol. 7): American Educational Research Association.

Marias, I. (2013). Local dependence. In K. B. Christensen, S. Krainer, & M. Mesbah (Eds.), Rasch models in health (pp. 111–130). ISTE and John Wiley & Sons.

Meretoja, R., & Leino-Kilpi, H. (2003). Comparison of competence assessments made by nurse managers and practising nurses. Journal of Nursing Management, 11(6), 404–409.

Meretoja, R., Leino-Kilpi, H., & Kaira, A.-M. (2004). Comparison of nurse competence in different hospital work environments. Journal of Nursing Management, 12(5), 329–336.

Morphet, J., Innes, K., Griffiths, D. L., Crawford, K., & Williams, A. (2015). Resident transfers from aged care facilities to emergency departments: Can they be avoided? Emergency Medicine Australasia, 27(5), 412–418. https://doi.org/10.1111/1742-6723.12433

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. MESA Press.

Recio-Saucedo, A., Dall'Ora, C., Maruotti, A., Ball, J., Briggs, J., Meredith, P., Redfern, O. C., Kovacs, C., Prytherch, D., Smith, G. B., & Griffiths, P. (2018). What impact does nursing care left undone have on patient outcomes? Review of the literature. Journal of Clinical Nursing, 27(11-12), 2248–2259. https://doi.org/10.1111/jocn.14058

Shaheen, A.M., El-Hneiti, M., Albqoor, M., & Ahmad, M. (2019). Predictors of quality of care provided for older adults. Journal of Nursing Management, 27(8), 1747–1755. https://doi.org/10.1111/jonm.12869

Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. Structural Equation Modeling, 3(1), 25–40. https://doi.org/10.1080/10705519609540027

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

Tohmola, A., Saarnio, R., Mikkonen, K., Kyngäs, H., & Elo, S. (2020). Development and psychometric testing of the Gerontological Nursing Competence (GeroNursingCom) instrument. *Journal of Advanced Nursing, n/a*(n/a). doi:https://doi.org/10.1111/jan.14609

Wright, B. (1996). Reliability and separation. Rasch Measurement Transactions, 9(4), 472. Retrieved from https://www.rasch.org/rmt/rmt94n.htm

Table 1

The test items of the Ms. Olsen test

Item no.	Item wording
1	Has dyspnoea during rest within last two days
2	Coughs, has increased saliva and respiration frequency above 20/min
3	Has irregular pulse increased to more than 20/min within last two days
4	Has temperature above 38.5
5	Is substantially dehydrated
6	Skin has rash, wounds, is red or itchy
7	Has reduced appetite and food intake
8	Is not able to eat
9	Has pain and discomfort in mouth
10	Is incontinent for urine, stings when urinates
11	Has fresh blood in stool
12	Has increased needs to full-time care within last two days
13	Has fallen two times during previous week
14	Has symptoms of partial paralysis
15	Has changes in sight, hearing, speech and comprehension
16	Has newly occurring chest pain
17	Has lost interest in keeping home in order, sleeps in chair instead of bed
18	Has short attention span and delusions
19	Is more tired during the day

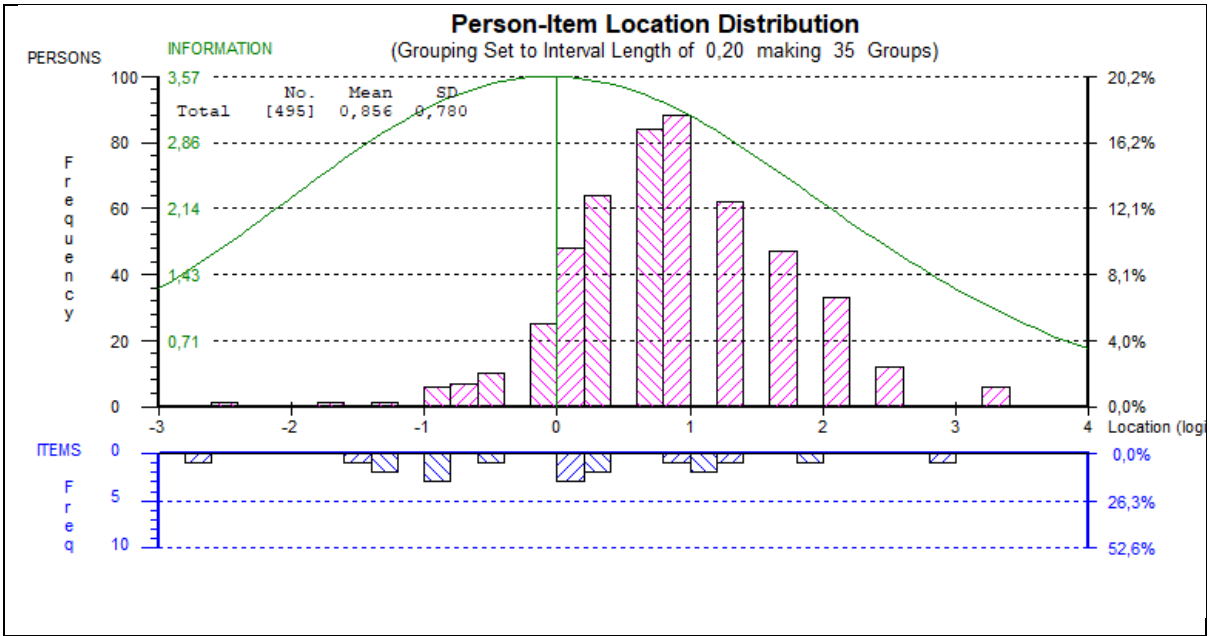

FIGURE 1 RNs Person-Item location distributions with top histogram (pink bars) sample distribution of nurses' abilities ranging from low to high abilities and bottom histogram (blue bars) distribution of location of Ms. Olsen item task difficulty for RNs ranging from easy to more challenging items. Green line shows information curve. The mean person location is 0.86 and person and items are about equally distributed, indicating a slightly positive targeting but with gaps.

Table 2

The Ms. Olsen test for RNs: Item locations, uncertainties, fit-statistics and pre-classification of item difficulty to be easy, middle or hard

	Item	Location	2SE	FitResid	ChiSq	DF	Prob	PC loading	Classification
Is incontinent of urine, stings when urinates	10	-2.67	0.48	-0.88	10.78	7	0.15	Negative	Easy
Has symptoms of partial paralysis	14	-1.45	0.30	-0.74	7.33	7	0.40	Positive	Easy
Coughs, has increased saliva and respiration frequency above 20/min	2	-1.34	0.29	-1.07	3.66	7	0.82	Negative	Easy
Has fallen two times during previous week	13	-1.21	0.27	-1.10	8.12	7	0.32	Positive	Middle
Is not able to eat	8	-0.96	0.25	-1.01	2.95	7	0.89	Negative	Easy
Has temperature above 38.5	4	-0.85	0.25	-0.87	12.51	7	0.08	Negative	Easy
Has newly occurring chest pain	16	-0.83	0.24	-0.53	5.85	7	0.56	Positive	Easy
Has increased needs to full-time care within last two days	12	-0.57	0.23	-0.37	5.61	7	0.59	Positive	Middle
Has reduced appetite and food intake	7	0.10	0.20	0.74	8.25	7	0.31	Negative	Easy
Has changes in sight, hearing, speech and comprehension	15	0.18	0.20	0.59	17.38	7	0.02	Positive	Easy
Has short attention span and delusions	18	0.18	0.20	-0.25	4,81	7	0.68	Positive	Middle
Has dyspnoea during rest within last two days	1	0.20	0.20	1.49	6.54	7	0.48	Negative	Middle
Is substantially dehydrated	5	0.25	0.20	1.10	2.30	7	0.94	Negative	Easy
Has fresh blood in stool	11	0.92	0.19	1.93	5.29	7	0.62	Negative	Middle
Has pain and discomfort in mouth	9	1.10	0.19	0.73	3.30	7	0.86	Negative	Easy
Has lost interest in keeping home in order, sleeps in chair instead of bed	17	1.13	0.19	-1.04	7.37	7	0.39	Positive	Middle
Skin has rash, wounds, is red or itchy	6	1.20	0.19	0.97	5.58	7	0.59	Negative	Middle
Is more tired during the day	19	1.82	0.21	-0.77	5.47	7	0.60	Positive	Middle
Has irregular pulse increased to more than 20/min within last two days	3	2.81	0.26	0.40	8.14	7	0.32	Negative	Middle

Table 3

The Ms. Olsen test for ANs: Item locations, uncertainties, fit-statistics and pre-classification of item difficulty to be easy, middle or hard

	Item	Location	2SE	FitResid	ChiSq	DF	Prob	PC loading	Classification
Has symptoms of partial paralysis	14	-1.49	0.31	0.94	13.15	9	0.16	Positive	Easy
Has fresh blood in stool	11	-1.31	0.30	-0.12	3.02	9	0.96	Negative	Hard
Has short attention span and delusions	18	-0.86	0.28	0.54	7.15	9	0.62	Negative	Hard
Has newly occurring chest pain	16	-0.71	0.27	2.03	13.24	9	0.15	Positive	Middle
Has reduced appetite and food intake	7	-0.65	0.27	0.84	7.17	9	0.62	Negative	Middle
Has dyspnoea during rest within last two days	1	-0.52	0.27	1.17	15.66	9	0.07	Negative	Hard
Has changes in sight, hearing, speech and comprehension	15	-0.27	0.26	2.32	15.97	9	0.07	Positive	Hard
Is incontinent of urine, stings when urinates	10	0.02	0.26	1.25	12.28	9	0.20	Positive	Easy
Has pain and discomfort in mouth	9	0.10	0.26	0.28	7.34	9	0.60	Positive	Easy
Has temperature above 38.5	4	0.18	0.26	-1.47	12.09	9	0.21	Positive	Easy
Skin has rash, wounds, is red or itchy	6	0.20	0.26	-1.90	24.04	9	0.00	Positive	Middle
Has fallen two times during previous week	13	0.23	0.26	-0.31	6.58	9	0.68	Negative	Hard
Has lost interest in keeping home in order, sleeps in chair instead of bed	17	0.51	0.27	-0.95	11.72	9	0.23	Negative	Hard
Is not able to eat	8	0.57	0.27	-0.06	5.36	9	0.80	Negative	Hard
Coughs, has increased saliva and respiration frequency above 20/min	2	0.60	0.27	0.56	5.08	9	0.83	Positive	Easy
Is more tired during the day	19	0.65	0.27	-0.13	10.91	9	0.28	Negative	Hard
Has increased needs to full-time care within last two days	12	0.78	0.28	0.85	5.54	9	0.78	Negative	Hard
Has irregular pulse increased to more than 20/min within last two days	3	0.79	0.28	-0.27	21.53	9	0.01	Positive	Hard
Is substantially dehydrated	5	1.17	0.30	0.49	8.58	9	0.48	Positive	Hard

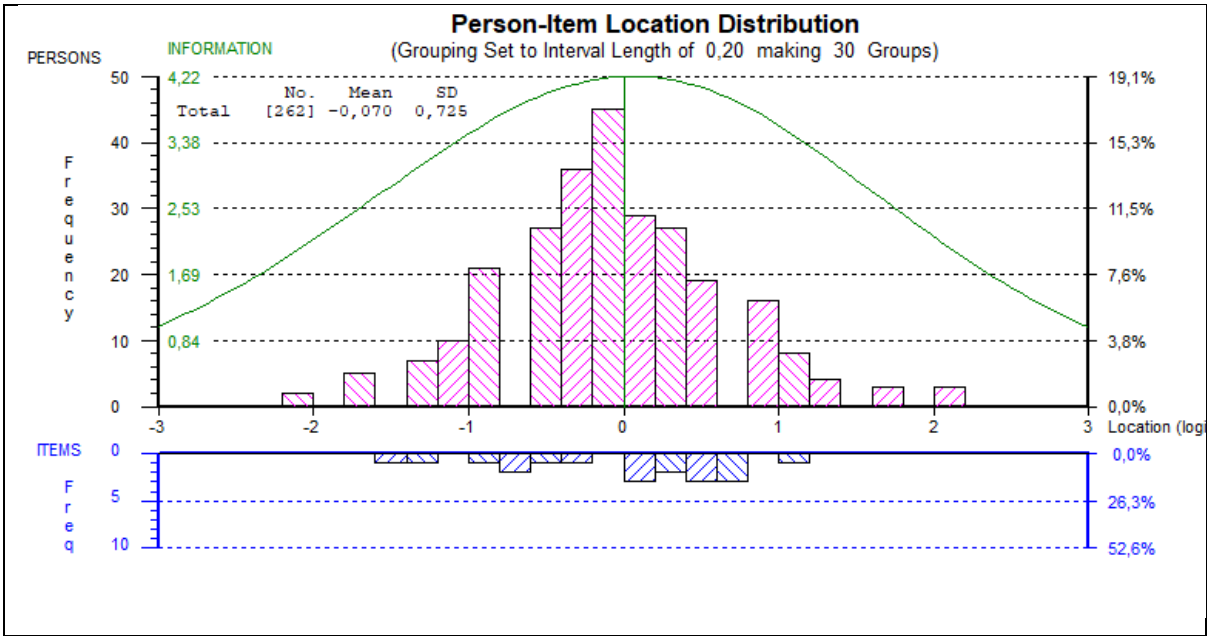

FIGURE 2 ANs Person-Item location distributions with top histogram (pink bars) sample distribution of nursing assistants' abilities ranging from low to high abilities and bottom histogram (blue bars) distribution of location of Ms. Olsen item task difficulty for nursing assistants ranging from easy to more challenging items. Green line shows information curve. The mean person location is –0.07 and person and items are about equally distributed, indicating well-targeted person-item.

Appendix 1. The score sheet for test items of the Ms. Olsen test.

Item no.	Item wording	Correct response for RNs	Correct response for ANs
1	Has dyspnoea during rest within last two days	5*	4 or 5
2	Coughs, has increased saliva and respiration frequency above 20/min	4 or 5	4
3	Has irregular pulse increased to more than 20/min within last two days	4	4
4	Has temperature above 38.5	4 or 5	4
5	Is substantially dehydrated	4 or 5	4
6	Skin has rash, wounds, is red or itchy	4	4
7	Has reduced appetite and food intake	3 or 4	3 or 4
8	Is not able to eat	4 or 5	4
9	Has pain and discomfort in mouth	4	4
10	Is incontinent of urine, stings when urinates	4 or 5	4
11	Has fresh blood in stool	5	4 or 5
12	Has increased needs to full-time care within last two days	4 or 5	4
13	Has fallen two times during previous week	4 or 5	4
14	Has symptoms of partial paralysis	6	4 or 6
15	Has changes in sight, hearing, speech and comprehension	4	4
16	Has newly occurring chest pain	6	4 or 6

17	Has lost interest in keeping home in order, sleeps in chair instead of bed	6	4 or 6
18	Has short attention span and delusions	4	4
19	Is more tired during the day	5	4 or 5

* The response categories were: 1. *No action required*, 2. *Observe again the following day*, 3. *Consult with an RN*, 4. *Nursing-related measure required immediately*, 5. *Have patient assessed by physician*, and 6. *Requires acute help in hospital*.

Appendix 2. Transformation tables.

Transformation from raw scores to logits for RNs, the defensible cut off was 0.62, indicated by double line.

Raw score	Location	2SE
2	-2.52	1.53
4	-1.65	1.23
5	-1.30	1.16
6	-0.99	1.12
7	-0.70	1.08
8	-0.42	1.07
9	-0.15	1.06
10	0.12	1.06
11	0.40	1.07
12	0.68	1.09
13	0.97	1.12
14	1.29	1.17
15	1.64	1.25
16	2.05	1.36
17	2.55	1.55
18	3.24	1.92

Transformation from raw scores to logits for ANs, the defensible cut off was 0.01, indicated by double line.

Raw score	Location	2SE
2	-2.17	1.45
3	-1.74	1.27
4	-1.39	1.16
5	-1.09	1.09
6	-0.82	1.04
7	-0.57	1.01
8	-0.33	0.99
9	-0.10	0.98
10	0.13	0.97
11	0.36	0.98
12	0.59	1.00
13	0.84	1.03
14	1.10	1.07
15	1.39	1.14
16	1.72	1.25
17	2.14	1.42

Appendix 2 STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2-3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	1&4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4-5
Bias	9	Describe any efforts to address potential sources of bias	8
Study size	10	Explain how the study size was arrived at	4-5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4-5
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6-8
		(*b*) Describe any methods used to examine subgroups and interactions	6-8
		(*c*) Explain how missing data were addressed	6-8
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5-7
		(b) Give reasons for non-participation at each stage	5-7
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	
		(*b*) Report category boundaries when continuous variables were categorized	8
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8-10
Discussion			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Title page

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.