HHS Public Access

Author manuscript

Res Gerontol Nurs. Author manuscript; available in PMC 2022 August 15.

Published in final edited form as:

Res Gerontol Nurs. 2022 ; 15(3): 124–130. doi:10.3928/19404921-20220428-02.

Examining the Clinical Utility of the Brief Interview for Mental Status

Chih-Ying Li, PhD, OTR,
Department of Occupational Therapy, University of Texas Medical Branch, Galveston, Texas

Paul Arthur, PhD, OTR,
Department of Occupational Therapy, University of Southern Indiana, Evansville, Indiana

Brian Downer, PhD,
Department of Nutrition, Metabolism and Rehabilitation Sciences, School of Health Professions, University of Texas Medical Branch, Galveston, Texas

Craig A. Velozo, PhD, OTR,
Division of Occupational Therapy, College of Health Professions, Medical University of South Carolina, Charleston, South Carolina

Yong-Fang Kuo, PhD,
Office of Biostatistics, Preventive Medicine and Population Health, University of Texas Medical Branch, Galveston, Texas

Huey-Ming Tzeng, PhD, RN, FAAN,
School of Nursing, University of Texas Medical Branch, Galveston, Texas

Kenneth J. Ottenbacher, PhD, OTR
Department of Nutrition, Metabolism and Rehabilitation Sciences, School of Health Professions, University of Texas Medical Branch, Galveston, Texas

Abstract

The Brief Interview for Mental Status (BIMS) is used to identify delirium and needed supports in patients living in skilled nursing facilities (SNFs) and long-term care facilities (LTCFs). We analyzed 3,537,404 patients discharged from acute hospitals to SNFs or LTCFs with factor and Rasch analyses to examine the clinical utility of the BIMS. More than 40% of the sample had maximum scores, indicating a ceiling effect. “Repetition of three words” was the easiest and the only misfit item (Outfit = 3.14). The ability of the BIMS to distinguish individuals into two cognitive levels (with person strata of 1.48) was limited. Although the BIMS is a widely used screening tool for cognitive impairment, we found it lacked sensitivity for approximately one half of patients admitted to SNFs/LTCFs. Our results suggest the BIMS should be interpreted with caution, particularly for patients with mild cognitive impairment.

Address correspondence to Chih-Ying Li, PhD, OTR, Department of Occupational Therapy, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-1142; chili@utmb.edu.

Disclosure: The authors have disclosed no potential conflicts of interest, financial or otherwise.

Skilled nursing facilities (SNFs) and long-term care facilities (LTCFs) are the most frequently used post-acute sites for older adults following hospital stays (U.S. Department of Health and Human Services, 1995). Cognitive impairment is common among older adults admitted to SNFs and LTCFs (Björk et al., 2016). Changes in cognitive status can be due to dementia, medication interactions (Saraf et al., 2016), or alternative conditions contributing to the hospital stay (Ehlenbach et al., 2018). It is important for nursing home professionals to accurately assess and interpret the cognitive status of older adults over the course of their stay. Using only hospital discharge medical records may result in incomplete or inaccurate information on health status and cognitive function (King et al., 2013).

BACKGROUND

In the United States, the Centers for Medicare & Medicaid Services (CMS) uses the Minimum Data Set (MDS) as the mandatory assessment to document the health, function, and care processes for patients in SNFs and LTCFs. Version 3.0 of the MDS includes several performance-based assessments (Saliba et al., 2012). The MDS includes a short cognitive screening tool, the Brief Interview for Mental Status (BIMS), designed to measure basic cognitive function, including repetition of words, temporal orientation, and word recall (Saliba et al., 2012). Earlier researchers have found it easier to administer and interpret BIMS results after excluding two judgment items in the pilot testing version (Chodosh et al., 2008). Mansbach et al. (2014) reported that the Brief Cognitive Assessment Tool (BCAT) includes broader memory and executive items, which may be too challenging for LTC residents.

The BIMS appears to have acceptable accuracy in identifying patients in SNFs with severe cognitive impairment (Saliba et al., 2012). However, Mansbach et al. (2014) reported that the BIMS does not differentiate between patients with normal cognition and those with mild cognitive impairment, or between mild-moderate dementia, implying insufficient evidence regarding the clinical utility of the BIMS for nursing home residents.

Understanding the item-level psychometric properties of the BIMS is the first step to identify where it is performing inadequately and how it may be improved. The purpose of the current study was to examine the strengths and weaknesses of the BIMS by evaluating its items and structure to determine if it is a reliable and useful clinical assessment to identify cognitive status for older adults admitted to SNFs and LTCFs.

METHOD

Sample and Settings

The current study was approved by the Institutional Review Board at the University of Texas Medical Branch and under an established Data Use Agreement with the CMS. We analyzed national MDS 3.0 Medicare data of older adults admitted to SNFs or LTCFs within 3 days after hospital discharge from January 1, 2013 to December 31, 2014. We selected the first SNF or LTCF stay and included patients who were evaluated with the MDS assessment within 2 weeks after admission. The final analytic sample was 3,537,404 patients after

excluding those unable to attempt or complete the BIMS and those who missed any BIMS items (Table A, available in the online version of this article). We recognize that SNFs and LTCFs are different regarding the services provided, operational environment, and payment models (Mansbach et al., 2014). However, >90% of SNFs are dually certified as LTCFs (Medicare Payment Advisory Commission, 2020) and the BIMS is commonly assessed at both settings.

Measures

The BIMS measures orientation (year, month, day), learning (repeat three words), and memory (recall three words). The range of scores for the BIMS is 0 to 15, with higher scores indicating greater cognition. Two items have a rating scale of 0 to 3 (repeat words and remember year), four items have a rating scale of 0 to 2 (remember month, recall the first/second/third word), and one item has a rating scale of 0 to 1 (remember day) (Saliba et al., 2012).

Statistical Analysis

We analyzed patient demographics using SAS 9.4 and conducted confirmatory factor analysis (CFA) to examine factor structure of the BIMS using Mplus 7.1. Item-level psychometrics were analyzed with Rasch analysis using Winsteps 3.92.1. CFA was used to validate the hypothesized one-factor model of the BIMS. We presumed all items in the BIMS measured the same latent trait (basic cognition). After one-factor model was validated, we used Rasch Partial Credit Model to examine unidimensionality of the scale. We used grouping command (Isgroups) in Winsteps to examine rating scale structures.

We used four model fit indices to interpret factor analysis results: normed chi-square (≤5 for good fit) (MacCallum et al., 1996), comparative fit index (CFI) (>0.95 for good fit) (Hu & Bentler, 1999), Tucker-Lewis index (TLI) (>0.95 for good fit) (Hu & Bentler, 1999), and root mean square error of approximation (RMSEA) (<0.08 for good fit) (MacCallum et al., 1996). Normed chi-square is calculated by dividing chi-square with degree of freedom to adjust for the effect of sample size. The CFI/TLI indicates agreement between the hypothesized model to a null model (with no factor). RMSEA represents agreement between expected and actual error values in a matrix after adjusting for sample size.

The Rasch residual principal component analysis (PCA) examined the underlying dimensionality of the scale by identifying if any meaningful residual remained after extracting primary Rasch/unidimensional dimension (Linacre, 2010). If the Rasch dimension explains >40% variance of the data and the eigenvalue of the first contrast was ≤2.0, then a scale is assumed to be unidimensional (Linacre, 2010). Theoretically, factor analysis results and dimensionality testing results should be consistent; we conducted both to cross-validate scale dimensionality of the BIMS.

We examined clinical utility by reporting person-item fit, person-item map, person strata, and ceiling/floor effects of the BIMS. We used the modified sample size–adjusted range of Infit and Outfit mean square (Infit MNSQ, Outfit MNSQ) of 0.9 to 1.1 as fit statistics criteria based on Smith's (2007) approach to account for the error variance for the clinical assessment with the value of standardized fit statistics (ZSTD) within ±2 (Bond & Fox,

2013; Linacre, 2010). Only high Infit/Outfit MNSQ values indicated a problematic outlier. Point measure correlation represents the correlation between item measures and person measures, with a value between ±0.3 indicating a weak correlation (Linacre, 1998). We also calculated the person strata with the person separation index (Gp), using the formula: ([4Gp + 1]/3) to represent the levels of person ability distinguished by all items (Bond & Fox, 2013; Linacre, 2012). Person separation reliability was used to indicate the levels of person reliability across all items, with a value <0.75 indicating low reliability, 0.75 to 0.95 indicating moderate reliability, and >0.95 indicating good reliability (Bond & Fox, 2013). The item-person map places item difficulty and person ability in the same continuum using a standardized unit of logit. A ceiling effect indicates items are too easy and a floor effect indicates items are too difficult for the study sample. We defined ceiling/floor effect as existing if >15% of the sample had the maximum/minimal scores (Li et al., 2018).

To examine the clinical utility of the BIMS as it relates to different racial groups and sexes, we conducted differential item functioning (DIF) analysis after item-level psychometric analyses using Winsteps Rasch-Welch (logistic regression) *t* test (Holland & Thayer, 1986). The magnitude of DIF was defined as slight to moderate if the absolute value of DIF size was ≥0.43 logits at significance level $p > 0.05$, and moderate to large if the absolute value of DIF size was ≥0.64 logits at significance level $p > 0.05$ (Zwick et al., 1999).

RESULTS

Demographics

Mean age of patients was 78.5 years (*SD* = 11 years) and the majority were White (81.7%), female (62.5%), and widowed (40.1%). Approximately 87% of the sample had a risk of pressure ulcer and approximately 37% had a fall history 30 days prior to SNF or LTCF admission (Table 1).

Stage 1: Testing Hypothesized One-Factor Model Using Factor Analysis

CFA showed that the one-factor model of BIMS met three of four indices. CFI (0.99)/TLI (0.98) both met the criterion >0.95, and RMSEA (0.079) marginally met the criterion <0.08. However, normed chi-square did not meet a model fit criterion (>5). In large sample sizes, there is a greater likelihood of yielding statistically significant results. Because all other fit statistics consistently met good model fit criteria (Shi et al., 2019), we consider the factor analysis results supporting one-factor structure assumption for the BIMS.

Stage 2: Rasch Principal Component Analysis of BIMS

Rasch PCA identified that the Rasch dimension explained a total 60.8% variance in the data (>40%). The eigenvalue of first contrast was 1.62 (<2.0), indicating no meaningful residuals were left after extracting the Rasch dimension, implying unidimensionality (a singular underlying construct) in the BIMS (Table B, available in the online version of this article).

Stage 3: Clinical Utility of BIMS Item-Level Psychometrics

The rating scale structures of the BIMS met all rating scale diagnostic criteria except one: rating scale of 1 in the rating scale group of 0 to 3 (4-point scale) showed Outfit value >2.0, suggesting potential revision of the scale (i.e., combining categories) (Linacre, 1998) (Table 2). However, all scales showed expected score disorder based on the given category definitions. “Repetition of three words” was the easiest item to answer and the only misfit item (Outfit = 3.14) based on the defined criteria (0.9 to 1.1) (Table 3). ZSTD were unusually high for all, with values of ±9.9 due to a large sample size. *Person strata,* the ability a scale can statistically discriminate different levels for respondents, was 1.48 based on the person separation index of 0.86, indicating the BIMS can hardly distinguish patients’ cognitive levels into two groups. *Person reliability,* the level of reproducibility for the person estimates (relative person measure location) and equivalent to Cronbach’s alpha, was 0.43, indicating the person estimates may not consistently be reliable when using the BIMS with this population. Of the sample, 41.9% had the maximum score, indicating a ceiling effect (Figure A, available in the online version of this article).

Differential Item Functioning

No DIF item was identified across sex (male vs. female) and race (White vs. racial and ethnic minority) groups (all DIF size <0.43 logit). The DIF sizes of BIMS items for sex (male vs. female) ranged from −0.28 (remember year) to 0.13 (repeat words). Males tended to do better than females in remembering year, whereas females tended to do better than males in repeating words. The DIF sizes of BIMS items for race (White vs. racial and ethnic minority) ranged from −0.22 (recall the third word) to 0.26 (repeat words). Participants who were White tended to do better than participants who were of racial and ethnic minorities in recall of the third word, whereas participants who were of racial and ethnic minorities tended to do better than participants who were White in repeat words. However, those above-mentioned items were not considered DIF items based on DIF criteria.

DISCUSSION

The current study examined the clinical utility of the BIMS in SNF and LTCF settings by examining its item-level psychometric properties. Our findings indicate that the BIMS is primarily a unidimensional assessment and may be appropriate to measure basic cognitive function for patients in SNFs or LTCFs. However, the BIMS had limited ability to distinguish different levels of cognitive functions (i.e., cognitively intact or impaired), raising concern for its sole use as a cognitive screening tool in SNFs/LTCFs. We also found the BIMS was too easy for most older adults admitted to a SNF or LTCF. We recommend adding more challenging cognitive items to the BIMS that can capture executive functioning or problem-solving ability to improve its clinical usefulness.

It is important to raise awareness to use accurate cognitive assessments to better serve residents in SNFs and LTCFs. This need is further underscored by the current global pandemic impacting SNF and LTCF care. Our finding, along with other published studies (Chodosh et al., 2008; Downer et al., 2017; Mansbach et al., 2014), suggest that SNFs and LTCFs may consider a two-step procedure (conducting a screening test first then following

with a more comprehensive cognitive assessment) for patients with mild impaired cognitive status to better tailor their care needs (Downer et al., 2017; Saliba et al., 2012; Schumacker & Lomax, 2004). Mace et al. (2016) found that the BCAT-SF had better sensitivity than the BIMS in differentiating cognitive levels. As the BIMS is mandatory in the CMS Medicare system, it is of marked benefit to produce discriminative information to properly screen cognitive status for patients in SNFs and LTCFs. The inability to identify cognitive challenges in a timely manner limits a gateway for preventive care, prompting societal and ethical concerns, as unmet health care needs among long-term care residents may increase. Further study of the clinical utility of the BIMS for patients with mild cognitive status is warranted to reduce the clinical and societal challenges of its use.

Approximately 16% of long-stay nursing home residents (McHorney & Tarlov, 1995) and 9% of SNF patients (Downer et al., 2017) are unable to attempt or complete the BIMS. If a patient cannot be understood or give verbal or written responses, the BIMS cannot be completed. Approximately 20% of patients in SNFs who are unable to attempt or complete the BIMS have a diagnosis of Alzheimer's disease or related dementia (Downer et al., 2017). This finding highlights the fact that patients with Alzheimer's disease or related dementia may require different cognitive screening procedures. The clinical utility of BIMS for individuals with Alzheimer's disease or related dementia also needs further evaluation.

Ceiling effects are a common limitation of cognitive screening (Proust-Lima et al., 2007). Clinicians and staff in SNFs and LTCFs need to be aware of how a ceiling effect of the assessment may affect patient care and outcome evaluations, as well as how other factors may affect cognitive performance during assessment, particularly for patients recently discharged from acute care. For instance, symptoms of dehydration may manifest cognitive changes (Morley, 2014). The BIMS has limited ability to detect subtle declines in cognition for patients who are at the ceiling (e.g., those with mild cognitive deficits). Thus, changes in cognition may not be detected until there is more pronounced cognitive decline and the optimal period to intervene may be missed. The simplicity of the BIMS rating scale makes it easy to administer, but with the cost of compromising sensitivity and discriminative properties. Clinicians in SNFs and LTCFs need to use additional information to better capture patients' cognitive status, including patients' ability to understand instructions, make decisions, and communicate ideas before determining patients' cognitive abilities and needs. Future research may examine how the BIMS performs in specific patient populations routinely discharged to SNFs and LTCFs. Our sensitive analyses found no DIF item across male versus female and White versus racial and ethnic minority groups, suggesting no difference of BIMS responses by sex and race. This finding may partially be due to the BIMS being designed and developed for the purpose of implementation in nationally representative samples (Saliba et al., 2012). There is no current literature suggesting a difference of BIMS responses in different sex or racial groups.

It is important to accurately assess patients' cognitive status at SNFs and LTCFs. However, the importance of using psychometrically sound tools is often overlooked in clinical practice. In particular, providers should recognize the limitations of the BIMS to effectively interpret the results for the best provision of care in SNFs and LTCFs. For instance, patients obtaining perfect scores may still encounter mild cognitive challenges related to daily

function and require additional assistance. The recent pandemic highlights that patients in SNFs and LTCFs are the most vulnerable in terms of receiving adequate care and services. Our findings indicate the importance of and need for efficient and effective evaluations at SNFs and LTCFs and the importance of sensitivity of the tool to capture cognitive decline. Using a well-established cognitive tool can allow for meaningful prevention, evaluation, and treatment for patients in SNFs and LTCFs. Our findings also suggest the importance for providers to advocate for patients with mild cognitive impairment and the need to select a more specific cognitive assessment to capture cognitive decline. Ultimately, we believe more research and development is needed to build more suitable tools for patients of SNFs and LTCFs, such as a tool with challenging questions that is easy to administer but can also discriminate between cognitive levels.

LIMITATIONS

The current analysis only included older adults enrolled in the Medicare fee-for-service program and are not generalizable to older adults with different insurance plans, such as Health Maintenance Organization plans, Accountable Care Organizations, or Medicare Advantage. The Winsteps Rasch-Welch *t* test only allows DIF analysis for the dichotomous variables, thus the DIF results could not be directly applied to Black/African American and Hispanic/Latino participants. Our findings provide baseline information regarding BIMS administration and its “real life” use for reporting patient outcomes. We recognize that due to a very large sample, the likelihood of misfitting response patterns or unexpected residuals from person abilities may potentially impact reliability indices. Although removing these samples from the analysis may potentially improve reliability statistics, we reported results as collected to reflect “real” clinical utility and psychometrics of this assessment.

CONCLUSION

Our findings indicate the unidimensionality of the BIMS, implying that clinicians and researchers may use the BIMS as a measure of basic cognitive function for patients in SNFs and LTCFs. However, the BIMS demonstrated a limited ability to identify difference in cognitive levels. Our findings suggest clinicians should interpret BIMS results carefully. We also suggest using supplementary sources of patients’ health data to cross-validate reliability of the BIMS results. Continued research is necessary to identify more challenging but easily administered items to enhance the clinical usefulness of the BIMS. With those efforts, frontline practitioners and administrators may use an improved cognitive screening as part of a multidomain assessment to determine meaningful patient cognitive health and function at SNFs and LTCFs.

Acknowledgment:

The authors thank Sarah Toombs Smith, PhD, a board-certified Editor in the Life Sciences (bels.org) at the Sealy Center on Aging, University of Texas Medical Branch, for her assistance in reviewing and editing the manuscript prior to submission. She received no compensation for this effort beyond her university salary.

Funding:

This study was funded, in part, by the National Institutes of Health (P2CHD065702, K01AG058789, K01HD101589).

REFERENCES

Björk S, Juthberg C, Lindkvist M, Wimo A, Sandman PO, Winblad B, & Edvardsson D (2016). Exploring the prevalence and variance of cognitive impairment, pain, neuropsychiatric symptoms and ADL dependency among persons living in nursing homes: A cross-sectional study. BMC Geriatrics, 16(1), 154. 10.n86/s12877-016-0328-9 [PubMed: 27549203]

Bond TG, & Fox CM (2013). Applying the Rasch model: Fundamental measurement in the human sciences. Psychology Press. 10.4324/9781410614575

Chodosh J, Edelen MO, Buchanan JL, Yosef JA, Ouslander JG, Berlowitz DR, Streim JE, & Saliba D (2008). Nursing home assessment of cognitive impairment: Development and testing of a brief instrument of mental status. Journal of the American Geriatrics Society, 56(11), 2069–2075. 10.1111/j.1532-5415.2008.01944.x [PubMed: 19016941]

Downer B, Thomas KS, Mor V, Goodwin JS, & Ottenbacher KJ (2017). Cognitive status of older adults on admission to a skilled nursing facility according to a hospital discharge diagnosis of dementia. Journal of the American Medical Directors Association, 18(8), 726–728. 10.1016/j.jamda.2017.04.021 [PubMed: 28623153]

Ehlenbach WJ, Gilmore-Bykovskyi A, Repplinger MD, Westergaard RP, Jacobs EA, Kind AJH, & Smith M (2018). Sepsis survivors admitted to skilled nursing facilities: Cognitive impairment, ADL dependence, and survival. Critical Care Medicine, 46(1), 37–44. 10.1097/CCM.0000000000002755 [PubMed: 28991827]

Holland PW, & Thayer DT (1986) Differential item performance and the Mantel Haenszel procedure (Technical Report No. 86-69). Educational Testing Service.

Hu LT, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. 10.1080/10705519909540118

King BJ, Gilmore-Bykovskyi AL, Roiland RA, Polnaszek BE, Bowers BJ, & Kind AJ (2013). The consequences of poor communication during transitions from hospital to skilled nursing facility: A qualitative study. Journal of the American Geriatrics Society, 61(7), 1095–1102. 10.1111/jgs.12328 [PubMed: 23731003]

Li CY, Romero S, Bonilha HS, Simpson KN, Simpson AN, Hong I, & Velozo CA (2018). Linking existing instruments to develop an activity of daily living item bank. Evaluation & the Health Professions, 41(1), 25–43. 10.1177/0163278716676873 [PubMed: 27856680]

Linacre JM (1998). Table 13.1 Item statistics in measure order. Rasch Measurement Forum. http://www.winsteps.com/winman/index.htm?correlations.htm

Linacre JM (2010). Predicting responses from Rasch measures. Journal of Applied Measurement, 11(1), 1–10. [PubMed: 20351444]

MacCallum RC, Browne MW, & Sugawara HM (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1(2), 130–149. 10.1037/1082-989X.1.2.130

Mace RA, Mansbach WE, & Clark KM (2016). Rapid cognitive assessment of nursing home residents: A comparison of the Brief Interview for Mental Status (BIMS) and Brief Cognitive Assessment Tool-Short Form (BCAT-SF). Research in Gerontological Nursing, 9(1), 35–44. 10.3928/19404921-20150522-05 [PubMed: 26020578]

Mansbach WE, Mace RA, & Clark KM (2014). Differentiating levels of cognitive functioning: A comparison of the Brief Interview for Mental Status (BIMS) and the Brief Cognitive Assessment Tool (BCAT) in a nursing home sample. Aging & Mental Health, 18(7), 921–928. 10.1080/13607863.2014.899971 [PubMed: 24679128]

McHorney CA, & Tarlov AR (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? Quality of Life Research, 4(4), 293–307. 10.1007/BF01593882 [PubMed: 7550178]

Medicare Payment Advisory Commission. (2020). Chapter 8: Skilled nursing facility services. http://www.medpac.gov/docs/default-source/reports/mar20_medpac_ch8_sec.pdf?sfvrsn=0

Morley JE (2014). Mild cognitive impairment: A treatable condition. Journal of the American Medical Directors Association, 15(1), 1–5. 10.1016/j.jamda.2013.11.001 [PubMed: 24359697]

Proust-Lima C, Amieva H, Dartigues JF, & Jacqmin-Gadda H (2007). Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. American Journal of Epidemiology, 165(3), 344–350. 10.1093/aje/kwk017 [PubMed: 17105962]

Saliba D, Buchanan J, Edelen MO, Streim J, Ouslander J, Berlowitz D, & Chodosh J (2012). MDS 3.0: Brief interview for mental status. Journal of the American Medical Directors Association, 13(7), 611–617. 10.1016/j.jamda.2012.06.004 [PubMed: 22796362]

Saraf AA, Petersen AW, Simmons SF, Schnelle JF, Bell SP, Kripalani S, Myers AP, Mixon AS, Long EA, Jacobsen JM, & Vasilevskis EE (2016). Medications associated with geriatric syndromes and their prevalence in older hospitalized adults discharged to skilled nursing facilities. Journal of Hospital Medicine, 11(10), 694–700. 10.1002/jhm.2614 [PubMed: 27255830]

Schumacker RE, & Lomax RG (2004). A beginner's guide to structural equation modeling. Psychology Press. 10.4324/9781410610904

Shi D, Lee T, & Maydeu-Olivares A (2019). Understanding the model size effect on SEM fit indices. Educational and Psychological Measurement, 79(2), 310–334. 10.1177/0013164418783530 [PubMed: 30911195]

Smith RM (2007). Using item mean squares to evaluate fit to the Rasch model. In Bond TG & Fox CM (Eds.), Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.). Lawrence Erlbaum Associates.

U.S. Department of Health and Human Services. (1995). Overview of long-term care in five nations: Australia, Canada, the Netherlands, Norway, and the United States. https://aspe.hhs.gov/reports/overview-long-term-care-five-nations-australia-canada-netherlands-norways-united-states-0

Zwick R, Thayer DT, & Lewis C (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. Journal of Educational Measurement, 36(1), 1–28. 10.1111/j.1745-3984.1999.tb00543.x

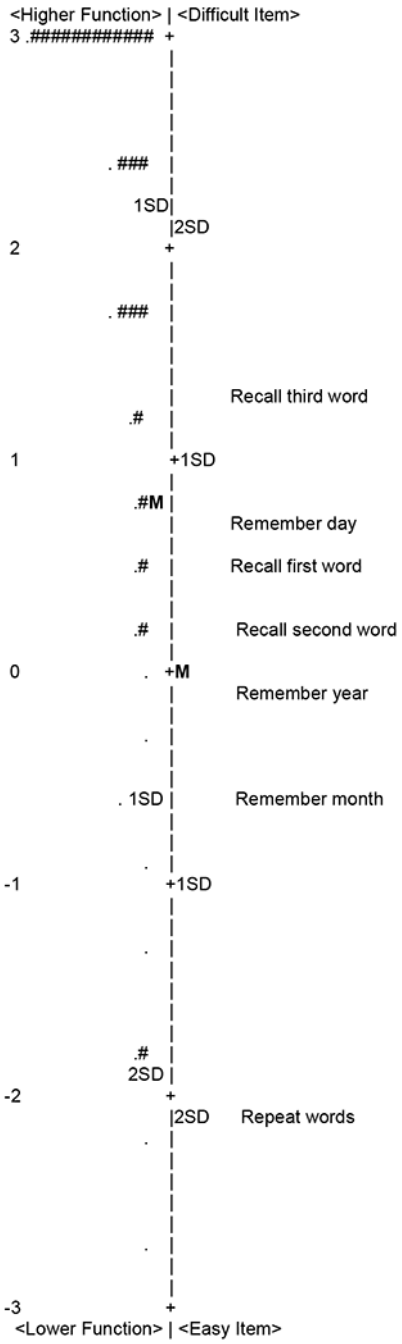

Figure A.

Item-person map of person ability and item difficulty of the Brief Interview for Mental Status.

Note. **M** represents the ‘mean’ (or the average) of the person ability (left) and item difficulty (right) after logarithm transformation. The scale of −3 to 3 represents Rasch logit with the mean item difficulty set as zero in the Rasch model. **1SD** represents one standard deviation from the mean of person ability/item difficulty. **2SD** represents two standard deviations from

the mean of person ability/item difficulty. Each dot represents 1 to 114083 patients. Each "#" represents 114084 patients.

TABLE 1

Demographics of Study Sample (N = 3,537,404)

Variable	*n* (%)
Sex	
Female	2,210,780 (62.5)
Male	1,326,468 (37.5)
Missing	156 (<0.01)
Marital status	
Widowed	1,416,700 (40.1)
Married	1,227,890 (34.7)
Divorced	382,164 (10.8)
Never married	368,958 (10.4)
Separated	35,989 (1.0)
Missing	105,703 (3)
Race	
White	2,888,884 (81.7)
Black	317,594 (9)
Hispanic	140,248 (4)
Asian	49,093 (1.4)
American Indian or Alaskan Native	10,789 (0.3)
Native Hawaiian or Pacific Islander	9,423 (0.2)
Missing	121,373 (3.4)
Pressure ulcer risk	
No	3,081,838 (87.1)
Yes	452,152 (12.8)
Missing	3,414 (0.1)
Fall incidence[a]	
No	1,909,712 (57.8)
Yes	1,234,292 (37.4)
Missing	393,400 (11.1)
	Mean (*SD*) (Range)
Age (years)	78.5 (11) (65 to 119)
BIMS score[b]	12.2 (3.7) (1 to 15)

Note. BIMS = Brief Interview for Mental Status.

[a]30-day prior to skilled nursing facility/long-term care facility admission.

[b]Score ranges from 0 to 15, with higher scores indicating greater cognition.

TABLE 2

Rating Scale Diagnostics and Person Separation Index of the Brief Interview for Mental Status

Category Level	Observed Count, *n* (%)	Observed Average	Sample Expect	Infit MNSQ	Outfit MNSQ	Andrich Threshold
Rating scale 0 to 3						
0	527,514 (7)	−0.81	−0.82	1.04	1.73	NA
1	124,736 (2)	0.17	0.05	1.09	3.47[a]	1.07
2	229,436 (3)	1.24	0.99	1.08	1.96	−0.11
3	6,193,122 (88)	2.44	2.46	0.99	0.96	−0.96
Rating scale 0 to 2						
0	2,327,182 (16)	−1.11	−1.12	0.98	0.89	NA
1	2,070,442 (15)	0.25	0.31	0.95	0.84	−0.27
2	9,751,992 (69)	1.46	1.45	0.97	0.99	0.27
Rating scale 0 to 1						
0	992,502 (28)	0.07	0.04	1.04	1.08	NA
1	2,544,902 (72)	1.46	1.47	0.99	0.93	NA
Person separation = 0.86						
Person strata = (0.86×4+1)/3 = 1.48						
Person reliability = 0.43						
Item reliability = 0.91						

Note. MNSQ = mean square; NA = not applicable.

[a]Rating scale of 1 in the rating scale group of 0 to 3 showed outfit value >2.0.

TABLE 3

Item Difficulty in Hierarchical Order of the Brief Interview for Mental Status (BIMS)

BIMS Item	Score		Infit		Outfit		Point Measure Correlation
	Measure	SE	MNSQ	ZSTD	MNSQ	ZSTD	
Recall the third word	1.26	0.00	1.01	9.9[a]	1.02	9.9	0.79
Remember day	0.74	0.00	1.03	9.9	1.01	8.3	0.67
Recall the first word	0.54	0.00	0.98	−9.9	0.90	−9.9	0.77
Recall the second word	0.24	0.01	0.87[b]	−9.9	0.84[b]	−9.9	0.76
Remember year	−0.07	0.00	1.04	9.9	−0.81[b]	−9.9	0.72
Remember month	−0.58	0.00	1.05	9.9	0.83[b]	−9.9	0.68
Repeat words	−2.13	0.00	0.94	−9.9	3.14[c]	9.9	0.35

Note. SE = standard error; MNSQ = mean square; ZSTD = standardized fit statistics.

[a]ZSTD only had the value of ±9.9 due to a large sample size.

[b]Misfit item based on the defined criteria (a higher fit statistics value and beyond the range of 0.9 to 1.1).

[c]Items that are overfitting the model.

Table A

Cohort Selection Procedures

	N	% of prior step
1. All Minimum Data Set 3.0 assessments between 1/1/2013 and 12/31/2014	36,591,088	
2. Select patients who did not have missing assessment code for the Brief Interview for Mental Status	23,685,778	64.7
3. Select the 1st assessment of the Brief Interview for Mental Status for each patient	4,887,154	20.6
4. Select the assessment which occurred within 14 days of the entry date at Skilled Nursing Facility and long-term nursing home care.	3,830,304	78.4
5. Select patients with completed Brief Interview for Mental Status data	3,537,404	92.3

Res Gerontol Nurs. Author manuscript; available in PMC 2022 August 15.

Table B

Rasch Principal Component Analysis of Rasch Residuals

	Standardized Residual Variance	
	Eigenvalue	**Percent**
Total Raw Variance in Observations	17.87	100.0
Raw Variance Explained by Measures	10.87	60.8
Raw Variance Explained by Persons	6.89	38.6
Raw Variance Explained by Items	3.98	22.3
Raw Unexplained Variance (Total)	7.00	39.2
Unexplained Variance in First Contrast	1.62	9.1