



JAMDA

journal homepage: www.jamda.com

Original Study-Brief Report

A 3-Item Screening Scale for Caregiver Burden in Dementia Caregiving: Scale Development and Score Mapping to the 22-Item Zarit Burden Interview



Tau Ming Liew MRCPsych(UK)^{a,b,c,*}, Philip Yap MRCP(UK)^{d,e}

^a Department of Geriatric Psychiatry, Institute of Mental Health, Singapore

^b Psychotherapy Service, Institute of Mental Health, Singapore

^c Saw Swee Hock School of Public Health, National University of Singapore, Singapore

^d Department of Geriatric Medicine, Khoo Teck Puat Hospital, Singapore

^e Geriatric Education and Research Institute, Singapore

A B S T R A C T

Keywords:

Caregiver burden
dementia
screening
Zarit Burden Interview
score mapping

Objectives: Brief screening scales for caregiver burden are much needed in routine dementia services to efficiently identify caregivers of persons with dementia (PWD) for further intervention. Although the 22-item Zarit Burden Interview (ZBI) is often used, its available screening versions have not performed as well as the full version in distinguishing significant burden. We developed a brief screening scale that is valid and comparable to ZBI in distinguishing caregiver burden.

Design and setting: Baseline data of an ongoing cohort study.

Participants: Family careivers of community-dwelling PWD ($n = 394$).

Measures: Participants completed questionnaires containing ZBI and other caregiving scales. Initially, we split the study samples into 2—the derivation sample ($n = 215$) was used to develop a brief scale that best distinguishes significant burden (using the best-subset approach with 10-fold cross-validation), whereas the validation sample ($n = 179$) verified its actual performance in distinguishing significant burden. We then evaluated the derived scale in its internal consistency reliability, factorial validity, known group validity, and construct validity, and mapped the scores between the brief scale and ZBI using the equipercenile equating method.

Results: We derived a 3-item scale which had comparable performance to ZBI in distinguishing significant burden (area under the receiver operating characteristic curve 0.86, 95% confidence interval 0.81–0.92). It had a single dimension in exploratory factor analysis and maintained good psychometric properties similar to those of ZBI. It also explained 77.8% of the variability in ZBI, and had scores that could be mapped to ZBI with reasonable precision.

Conclusions and Implications: We have derived a highly accessible tool to screen for caregiver burden, which can have a wider health system effect of expanding the reach of caregiver-focused interventions to services involved in the care of PWD. Notably, this screening tool was developed using rigorous methods and demonstrated comparability to ZBI in its validity, reliability, and total scores.

© 2018 The Author(s). Published by Elsevier Inc. on behalf of AMDA – The Society for Post-Acute and Long-Term Care Medicine. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Globally, the cost of dementia is estimated at US\$818 billion,¹ and it far exceeds that of cancer and heart disease.^{2,3} The largest portion of the cost (40.4%) is attributed to family caregiving,¹ which is

understandable because persons with dementia (PWD) become increasingly dependent on their family members as they lose the ability to care for themselves. The costs are only expected to rise

This research was supported by the Singapore Ministry of Health's National Medical Research Council under the Centre Grant Program (grant no. NMRC/CG/004/2013). It also received pilot funding from the National University of Singapore. Separately, the first author (T.M.L.) was supported by research grants under the Singapore Ministry of Health's National Medical Research Council (NMRC) (grant no. NMRC/Fellowship/0030/2016 and NMRC/CSSSP/0014/2017). The funding sources had no involvement in any part of the project.

The authors declare no conflicts of interest.

* Address correspondence to Tau Ming Liew, MRCPsych(UK), Department of Geriatric Psychiatry, Institute of Mental Health, 10 Buangkok View, Singapore 539747.

E-mail address: tau_ming_liew@imh.com.sg (T.M. Liew).

<https://doi.org/10.1016/j.jamda.2018.11.005>

1525–8610/© 2018 The Author(s). Published by Elsevier Inc. on behalf of AMDA – The Society for Post-Acute and Long-Term Care Medicine. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

further as the number of PWD is projected to triple from 46.8 million in 2015 to 131.5 million in 2050.¹ This has led to the call by the World Health Organization⁴ to stress support for family caregivers who typically constitute the cornerstone of dementia care. Moreover, caring for PWD can be stressful⁵ and is not uncommonly associated with physical and emotional burden.^{5–7} Caregiver burden, in turn, has been shown to predict less desirable outcomes such as premature nursing home placement⁸ and mortality in PWD.⁹ Although many interventions are available to address caregiver burden and improve caregiving outcomes,^{6,10} a critical issue that remains is the ability to identify caregivers with high burden in clinical care so that these caregivers can gain access to the necessary interventions.

In the published literature, the 22-item Zarit Burden Interview (ZBI)¹¹ is among the most widely used scale to identify caregiver burden.¹² However, despite being well evidenced in research settings, the scale is relatively lengthy and can increase the burden of administration to the caregivers. Several shorter versions of ZBI have been developed in an attempt to address this shortcoming.¹³ Unfortunately, the selection of scale items in most of these shorter variants had been done without reference to a gold standard or external criterion.¹³ Consequently, as shown in a recent study,¹³ many of them can have more questionable performance when benchmarked against an external criterion. In particular, the available short screening versions of the ZBI (the 1- and 4-item variants)^{14,15} showed poorer area under the receiver operating characteristic curve (AUROC) than the full version of ZBI in identifying caregivers with significant depression.¹³ This may raise concerns about the validity of these screening scales in distinguishing the caregiver burden, considering that burden and depression are commonly conceptualized as 2 ends of the same spectrum (under the diathesis stress model),¹⁶ with depression being the manifestation of high and significant burden.

To address the challenges, we sought (as the *primary aim*) to develop a brief burden scale with the fewest items possible, and evaluate its performance in distinguishing caregiver burden. Additionally, we had 2 *secondary aims* to demonstrate that this new brief scale has (1) properties of reliability and validity that are similar to those expected of ZBI and (2) scores that can be accurately mapped to the total scores of ZBI to demonstrate comparability.

Methods

Participants and Procedures

This study was based on the baseline data of an ongoing cohort study, where we consecutively sampled caregivers who accompanied the PWD to the dementia services of 2 tertiary hospitals in Singapore (Institute of Mental Health and Khoo Teck Puat Hospital). Our inclusion criteria comprised (1) spouses or children of PWD, (2) caring for PWD who is residing in the community, and (3) age ≥ 21 years. At the point of recruitment, the participants completed on site a set of self-administered questionnaires which included ZBI and a depression scale [Center for Epidemiologic Studies–Depression Scale (CES-D)].¹⁷ Participants from one of the recruitment sites Khoo Teck Puat Hospital also completed an additional scale assessing caregiving gains [Gain in Alzheimer care INstrument (GAIN)].

A total of 394 participants were recruited—215 (54.6%) from Hospital 1 and 179 (45.4%) from Hospital 2—with a total response rate of 88%. Ethics approval was granted by the Domain Specific Review Board of Singapore.

Measures

ZBI is a 22-item scale that assesses the perceived burden experienced by caregivers of older persons.¹¹ The items are self-administered by the caregivers on 5-point Likert-type scales and

summed to generate a total score ranging from 0 to 88. CES-D comprises 20 items that measure the frequency of depressive symptoms over the past week using 4-point Likert-type scales. The total score ranges from 0 to 60, with scores ≥ 16 suggestive of significant depression.^{17,18} CES-D has been shown to contain 4 domains—Depressed Affect, Somatic Symptoms, Positive Affect, and Interpersonal Problems.¹⁷ GAIN is a 10-item scale to measure caregiving gains in dementia. It is assessed on 5-point Likert-type scales and has a score range of 0 to 40. Higher scores denote higher gains. The ZBI,^{13,19} CES-D,²⁰ and GAIN²¹ have been validated in Singapore.

The severity of dementia was assessed based on the descriptors of the 3 dementia severities in the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R).²² Participants chose the description that best matched the PWD: still capable of independent living (mild stage), needs some assistance with daily living (moderate stage), or needs round-the-clock supervision (severe stage). This brief measure was previously shown to have adequate agreement with Clinical Dementia Rating Scale (kappa 0.56–0.6),^{23,24} which has been well validated and widely used to stage dementia.²⁵ The presence of severe behavioral problem was indirectly measured through the need for admission to a geriatric psychiatry ward, indicating behavioral problems that were too severe to be managed in the community setting.

Statistical Analyses

For the primary aim (scale development), we split the study samples into 2 (*derivation sample* and *validation sample*)—the *derivation sample* (based on participants from Institute of Mental Health, $n = 215$) was used to develop a brief scale that can best distinguish significant burden, whereas the *validation sample* (based on participants from Khoo Teck Puat Hospital, $n = 179$) was used to evaluate the actual performance of this brief scale in distinguishing significant burden.

In the *derivation sample* ($n = 215$), we employed the best-subset approach²⁶ with 10-fold cross-validation to select the scale items in ZBI that can best distinguish significant burden. Significant burden was indirectly identified by the presence of significant depression in caregivers (CES-D ≥ 16),^{17,18} which indicated the definitive need for further intervention. The best-subset approach is a computationally intensive method of variable selection.²⁷ It uses logistic regression to exhaustively evaluate all possible combinations of the 22 items from ZBI, and narrows down to a list of top models that have the lowest prediction errors. It then selects the best model using 10-fold cross-validation; this is done by randomly dividing the sample into 10 folds of equal size, cross-validating the prediction error within the 10 folds, and selecting the least complex model that is within 1 standard error of the best model (commonly described as the “1 standard error” rule which ensures that the selected model remains stable and can be consistently replicated even in other independent samples).²⁷ The selected model would then constitute the new, brief scale [henceforth referred to as the ZBI–Screening version (ZBI-S)]. In the *validation sample* ($n = 179$), we evaluated the actual performance of ZBI-S in distinguishing significant burden, by computing the AUROC. In general, an AUROC of >0.8 is considered excellent performance.²⁸

As part of the secondary aims, we evaluated whether ZBI-S maintained psychometric properties that were expected of a caregiver burden scale. Specifically, we assessed its psychometric properties in the full sample ($n = 394$) with respect to internal consistency reliability, factorial validity, known group validity, and construct validity. Internal consistency reliability was assessed with Cronbach alpha and McDonald omega,²⁹ with values of ≥ 0.70 indicating the minimally acceptable reliability.³⁰ Factorial validity was assessed with exploratory factor analysis using maximum-likelihood estimation methods and oblique rotation (oblimin), with the number of factors in exploratory factor analysis identified using Horn parallel analysis.³¹

Known group validity was assessed by comparing the mean scores based on variables that have been known to defer in the levels of caregiver burden, including primary caregiving role, coresidence with the PWD, severity of dementia, and presence of behavioral problems in the PWD.³² Construct validity was assessed using the Pearson correlation coefficient (r), with the hypotheses regarding construct validity further described in [Appendix 1](#).

As part of the other secondary aims, we also evaluated whether the scores of ZBI-S can be accurately mapped to those of ZBI. Using the full sample ($n = 394$), we mapped ZBI-S scores to the original 22-item ZBI using the equipercentile equating method with log-linear smoothing.³³ Equipercentile equating provides equivalent scores from one scale to another on the basis of their corresponding percentile rankings.³⁴ This method does not require any assumption on the score distribution, and can ensure that the mapped scores always fall within the range of the intended scale. Log-linear smoothing was applied to avoid an irregular distribution of the scores. The 95% confidence intervals (CIs) of the mapped scores were computed using 1000 bootstrap samples.

The reliability analyses²⁹ and score mapping³³ were performed in R (version 3.5.1). The other analyses were conducted in Stata (version 14).

Results

Demographic information of the 394 participants is presented in [Appendix 2](#). The participants had a mean age of 53.0 years (SD 10.7), with the majority being Ethnic 1 (86.6%), children caregivers (86.3%), and primary caregivers (70.8%).

In the *derivation sample* ($n = 215$), the exhaustive search method identified a list of top models as presented in [Appendix 3](#). As shown in [Figure 1](#), the 10-fold cross-validation then selected the 3-item model as the most parsimonious model based on the established 1-standard error rule.²⁷

In the *validation sample* ($n = 179$), the selected 3-item scale (ZBI-S) had excellent performance in distinguishing significant burden, with an AUROC of 0.86 (95% CI 0.80–0.91), and a sensitivity of 0.86 and a specificity of 0.73 at the optimal cut-off score of ≥ 4 ([Appendix 4](#)). Notably, the AUROC of ZBI-S was not significantly different from that of the 22-item ZBI ($P = .710$) ([Table 1](#)). In contrast, the previously known 1- and 4-item screening versions of ZBI^{14,15} had significantly

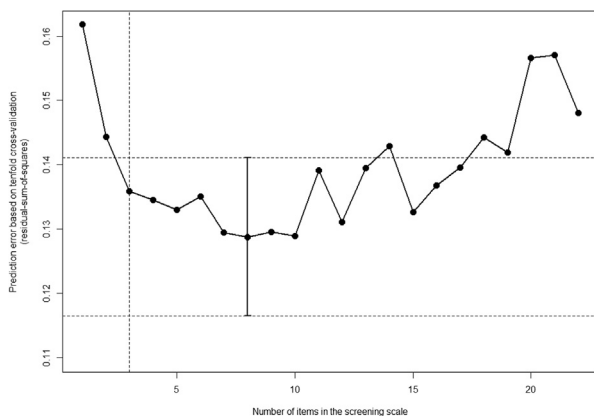


Fig. 1. Model selection with 10-fold cross-validation in the derivation sample ($n = 215$). Models with lower prediction error are considered better. Among which, the model with 8 items had the lowest values and was considered the best model. However, based on the “1 standard error” rule, the model with 3 items was selected to constitute the new screening scale because it was the least complex model that still fell within 1 standard error of the best model (indicated by the 2 horizontal dotted lines in the plot).

Table 1
Performance of the Screening Versions of ZBI in Identifying Significant Burden* in the Validation Sample ($n = 179$), and a Comparison With the Performance of the Original 22-Item Version

ZBI Variants	AUROC (95% CI)	Compared With Original 22-Item†		Optimal Cut-off	Sensitivity	Specificity	PPV	NPV
		P	Adj. P					
Original 22-item		—	—	≥ 34	0.86	0.75	0.73	0.86
Three-item ZBI-S (derived from this study)	0.86 (0.81–0.92)	.710	1.000	≥ 4	0.86	0.73	0.71	0.87
Four-item ¹⁵	0.82 (0.76–0.88)	.008	.024	≥ 7	0.82	0.71	0.70	0.84
One-item ¹⁴	0.78 (0.71–0.84)	.001	.004	≥ 2	0.83	0.61	0.63	0.83

Adj., Bonferroni adjusted; NPV, negative predictive value; PPV, positive predictive value.

*Significant burden was indirectly identified by the presence of significant depression (Center for Epidemiologic Studies-Depression score ≥ 16) in caregivers, which indicated the definitive need for further interventions.

†P values represent the statistical significance of the difference in AUROC between a ZBI variant and the original 22-item version.

‡The optimal cut-off score is based on a balance between sensitivity and specificity, with a preference for slightly higher sensitivity to reduce the false negative rates in screening scales.

worse AUROC than the original ZBI ($P = .008$ and $.001$ respectively) when they were evaluated in our sample (Table 1). The 3 screening versions of ZBI (1-, 4-, and 3-item variants) are separately shown in Appendix 5 for reference purposes.

We then evaluated the psychometric properties of ZBI-S in the full sample ($n = 394$). Despite being much briefer, the 3-item ZBI-S had an acceptable internal consistency reliability ($\alpha = 0.78$, 95% CI 0.74–0.82; $\omega = 0.80$, 95% CI 0.76–0.83). In exploratory factor analysis, ZBI-S demonstrated only 1 dimension among its scale items (the scree plot and the factor loading are shown in Appendix 6 and 7, respectively). The results for known group and construct validities were consistent with the characteristics expected of a caregiver burden scale (Appendices 8–12).

The 3-item ZBI-S explained 77.8% of variance in ZBI (based on the results from R-squared). As shown in Table 2, its scores can be mapped to those of ZBI, with a precision of approximately ± 3 in their 95% CI.

Discussion

This study developed a 3-item screening scale (ZBI-S) for caregiver burden in dementia caregiving, using a rigorous method of item selection (through exhaustive search of all possible combinations of items), and following the well-established processes of derivation, cross-validation, and independent validation. Unlike the previous screening versions of ZBI, the new ZBI-S was developed with reference to an external criterion and, consequently, demonstrated excellent performance in distinguishing significant burden (similar to that of ZBI, and comparatively better than the previously known 1- or 4-item variant of ZBI). Notably, ZBI-S also inherited the key properties of the original ZBI (the best-known scale to date for caregiver burden),¹² having properties of reliability and validity that are consistent with those of ZBI, explaining most of the variance in ZBI, and having scores that could be mapped back to ZBI with reasonable precision.

As demonstrated in this study, the 3-item ZBI-S can be as useful as the original 22-item ZBI (and better than the previous screening versions of the ZBI) in identifying caregivers with significant burden who might benefit from further intervention. The demonstrable score mapping offers the brief scale as a viable alternative when the original ZBI cannot be feasibly administered, as well as affords comparability across clinical sites that administer the 2 different versions. With only 3 items, ZBI-S can make screening of caregiver burden more accessible to clinical practitioners. It may thus have a wider health system effect of promoting caregiver-focused evaluations in clinical services that are involved in the care of PWD (including those in primary care and social care settings) and potentially expand the reach of caregiver-focused interventions beyond specialized dementia services.

Although there can be other alternative methods to derive a brief scale, the best-subset approach that we adopted has the strength of producing the most efficient brief scale, which has the least number of items yet remains comparable to ZBI. Although brief scales are not uncommonly fraught with deficiencies in psychometric properties,³⁰ our approach (to exhaustively search for the best subset in ZBI) ensures that ZBI-S inherits the characteristics of the original scale and minimizes the possibility of altering the psychometric properties of the brief scale in relation to the original scale.

Several limitations should be considered. First, the participants were recruited only from tertiary dementia services. However, they should largely still represent those in the community, because most of the PWD in Singapore receive their dementia care from tertiary centers, and the 2 recruitment centers in this study are the only 2 dementia services that serve the population in the North-East of Singapore. Second, the proportion of spousal caregivers in this study was relatively lower than that of children caregivers. However, this probably is not due to sampling bias considering that our proportion of spousal caregivers (13.7%) is not dissimilar to the 16.0% reported in a

Table 2
Equivalent ZBI Scores for a Given ZBI-S Score

ZBI-S	Equivalent ZBI (95% CI)
0	10 (7–12)
1	17 (14–20)
2	22 (20–25)
3	28 (25–31)
4	34 (31–36)
5	39 (36–42)
6	45 (42–48)
7	51 (47–55)
8	57 (52–61)
9	62 (58–67)
10	68 (62–73)
11	73 (67–80)
12	80 (75–86)

separate study based on a nationally representative sample.³⁵ Third, one may argue that the external criterion in this study (CES-D) may not reflect a definitive diagnosis of depression to indicate the caregivers' need for further intervention. However, a recent meta-analysis had demonstrated the excellent performance of CES-D ≥ 16 in identifying clinical depression, with a pooled AUROC of 0.87.¹⁸

Conclusions and Implications

In conclusion, this study has procured a highly accessible tool to screen for caregiver burden, which can have a wider health system effect of expanding the reach of caregiver-focused interventions to clinical and social services that are involved in the care of PWD. Notably, this 3-item tool was developed using rigorous methods and demonstrated comparability to ZBI in its validity, reliability, and total scores.

References

1. Prince MJ, Wimo A, Guerchet MM, et al. World Alzheimer Report 2015—The Global Impact of Dementia. London: Alzheimer's Disease International; 2015.
2. Abbott A. Dementia: A problem for our age. *Nature* 2011;475:S2.
3. Kelley AS, McGarry K, Gorges R, Skinner JS. The burden of health care costs in the last 5 years of life. *Ann Intern Med* 2015;163:729–736.
4. World Health Organization, Alzheimer's Disease International. Dementia: A Public Health Priority. London: WHO; 2012.
5. Mannion E. Alzheimer's disease: The psychological and physical effects of the caregiver's role. Part 1. *Nurs Older People* 2008;20:27–32.
6. Gilhooly KJ, Gilhooly MLM, Sullivan MP, et al. A meta-review of stress, coping and interventions in dementia and dementia caregiving. *BMC Geriatr* 2016;16:106.
7. Vitaliano PP, Zhang J, Scanlan JM. Is caregiving hazardous to one's physical health? A meta-analysis. *Psychol Bull* 2003;129:946–972.
8. Toot S, Swinson T, Devine M, et al. Causes of nursing home placement for older people with dementia: A systematic review and meta-analysis. *Int Psychogeriatr* 2016;29:195–208.
9. Lwi SJ, Ford BQ, Casey JJ, et al. Poor caregiver mental health predicts mortality of patients with neurodegenerative disease. *Proc Natl Acad Sci U S A* 2017;114:7319–7324.
10. Liew TM, Lee CS. Reappraising the efficacy and acceptability of multicomponent interventions for caregiver depression in dementia: The utility of network meta-analysis. *Gerontologist*; 2018 May 31 [Epub ahead of print].
11. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: Correlates of feelings of burden. *Gerontologist* 1980;20:649–655.
12. Whalen KJ, Buchholz SW. The reliability, validity and feasibility of tools used to screen for caregiver burden: A systematic review. *JBI Libr Syst Rev* 2009;7:1373–1430.
13. Yu J, Yap P, Liew TM. The optimal short version of the Zarit Burden Interview for dementia caregivers: diagnostic utility and externally validated cutoffs. *Aging Mental Health*; 2018:1–5.
14. Higginson IJ, Gao W, Jackson D, et al. Short-form Zarit Caregiver Burden Interviews were valid in advanced conditions. *J Clin Epidemiol* 2010;63:535–542.
15. Bedard M, Molloy DW, Squire L, et al. The Zarit Burden Interview: A new short version and screening version. *Gerontologist* 2001;41:652–657.
16. Patten SB. Major depression epidemiology from a diathesis-stress conceptualization. *BMC Psychiatry* 2013;13:19.

17. Radloff LS. The CES-D scale—A self-report depression scale for research in the general population. *Appl Psych Meas* 1977;1(3):385–401.
18. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): A systematic review with meta-analysis. *PloS One* 2016;11:e0155431.
19. Seng BK, Luo N, Ng WY, et al. Validity and reliability of the Zarit Burden Interview in assessing caregiving burden. *Ann Acad Med Singapore* 2010;39:758–763.
20. Stahl D, Sum CF, Lum SS, et al. Screening for depressive symptoms: validation of the Center for Epidemiologic Studies Depression scale (CES-D) in a multiethnic group of patients with diabetes in Singapore. *Diabetes Care* 2008;31:1118–1119.
21. Yap P, Luo N, Ng WY, et al. Gain in Alzheimer care INstrument—A new scale to measure caregiving gains in dementia. *Am J Geriatr Psychiatr* 2010;18:68–76.
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R*. Washington, DC: American Psychiatric Association; 1987.
23. Forsell Y, Fratiglioni L, Grut M, et al. Clinical staging of dementia in a population survey: comparison of DSM-III-R and the Washington University Clinical Dementia Rating Scale. *Acta Psychiatr Scand* 1992;86:49–54.
24. Juva K, Sulkava R, Erkinjuntti T, et al. Staging the severity of dementia: Comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales. *Acta Neurol Scand* 1994;90:293–298.
25. Rikkert MG, Tona KD, Janssen L, et al. Validity, reliability, and feasibility of clinical staging scales in dementia: A systematic review. *Am J Alzheimers Dis Other Dement* 2011;26:357–365.
26. McLeod I, Xu C. *bestglm: Best subset GLM*. 2010. Available at: <https://cran.r-project.org/package=bestglm>. Accessed August 1, 2018.
27. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; 2009.
28. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley; 2013.
29. Peters G-J. The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *Eur Health Psychol* 2014;16:56–69.
30. Fayers PM. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*. 2nd ed. Chichester, England: John Wiley & Sons; 2007.
31. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179–185.
32. Chiao CY, Wu HS, Hsiao CY. Caregiver burden for informal caregivers of patients with dementia: A systematic review. *Int Nurs Rev* 2015;62:340–350.
33. Albano AD. *equate: An R package for observed-score linking and equating*. *J Stat Softw* 2016;74:36.
34. Kolen MJ, Brennan RL. *Test Equating: Methods and Practices*. New York, NY: Springer Science & Business Media; 2013.
35. Malhotra R, Chan A, Malhotra C, Ostbye T. Validity and reliability of the Caregiver Reaction Assessment scale among primary informal caregivers for older persons in Singapore. *Aging Ment Health* 2012;16:1004–1015.

Appendix 1. Our Predefined Hypotheses Regarding the Construct Validity of the New ZBI-S (Zarit Burden Interview–Screening Version)

Construct validity was assessed using the Pearson correlation coefficient (r), with correlation coefficient of >0.50 considered strong while values ≤ 0.50 are considered weak or moderate.¹ We expected ZBI-S to demonstrate the following 4 characteristics consistent of a caregiver burden scale:

1. It should correlate strongly ($r > 0.50$) with constructs related to caregiver depression, such as with the Depressed Affect subscale of the CES-D and the Somatic Symptoms subscale of the

CES-D. This is because burden and depression has been viewed as 2 related constructs within the same spectrum,² with more severe burden manifesting as depression.

2. It should correlate less strongly ($r \leq 0.50$) with the Positive Affect and Interpersonal Problems subscales of the CES-D because caregiver burden is expected to differ from constructs such as positive feelings, or the feeling that others are being critical.
3. It should correlate less strongly ($r \leq 0.50$) with Gain in Alzheimer care INstrument (GAIN). This is because GAIN, which measures positive outcomes in caregiving, is a different construct from caregiver burden and has only been shown to correlate weakly with ZBI.³

Appendix 2

Demographic Information of the Caregivers and the Persons With Dementia They Cared For (n = 394)

Variable	Overall Sample (n = 394)	Derivation sample (n = 215)	Validation sample (n = 179)
Variables related to caregivers			
Age, mean (SD)	53.0 (10.7)	52.5 (10.8)	53.6 (10.5)
Female gender, n (%)	236 (59.9)	121 (56.3)	115 (64.3)
Ethnicity, n (%)			
Chinese	341 (86.6)	180 (83.7)	161 (89.9)
Malay	25 (6.3)	18 (8.4)	7 (3.9)
Indian/Eurasian/Others	28 (7.1)	17 (7.9)	11 (6.2)
Marital status, n (%)			
Married	271 (68.8)	153 (71.2)	118 (65.9)
Single	94 (23.9)	45 (20.9)	49 (27.4)
Widowed/divorced/separated	29 (7.3)	17 (7.9)	12 (6.7)
Employment status, n (%)			
Not working	123 (31.2)	64 (29.8)	59 (33.0)
Working part-time	52 (13.2)	32 (14.9)	20 (11.2)
Working full-time	219 (55.6)	119 (55.4)	100 (55.9)
Educational attainment, n (%)			
Primary or below	41 (10.4)	29 (13.5)	12 (6.7)
Secondary	228 (57.9)	130 (60.5)	98 (54.8)
Tertiary	125 (31.7)	56 (26.1)	69 (38.6)
Relationship with PWD, n (%)			
Child	340 (86.3)	185 (86.1)	155 (86.6)
Spouse	54 (13.7)	30 (14.0)	24 (13.4)
Coresidence with PWD, n (%)	264 (67.0)	137 (63.7)	127 (71.0)
Duration of caregiving, y, mean (SD)	6.8 (6.7)	6.3 (6.1)	7.4 (7.4)
Frequency of caregiving, n (%)			
Daily, for at least 4 h a day	211 (53.6)	115 (53.5)	96 (53.6)
Daily, but less than 4 h a day	79 (20.0)	44 (20.5)	35 (19.6)
At least once a week	84 (21.3)	45 (20.9)	39 (21.8)
Less than once a week	20 (5.1)	11 (5.1)	9 (5.0)
Primary caregiving role, n (%)	279 (70.8)	158 (73.5)	121 (67.6)
ZBI score, mean (SD)	34.8 (16.8)	36.5 (17.0)	32.9 (16.3)
CES-D score, mean (SD)	15.7 (11.0)	17.5 (11.4)	13.5 (10.2)
Variables related to PWD			
Age, mean (SD)	79.5 (8.2)	78.7 (8.4)	80.5 (7.8)
Female gender, n (%)	278 (70.6)	156 (72.6)	122 (68.2)
Duration of dementia diagnosis, y, mean (SD)	4.5 (3.5)	4.3 (3.4)	4.6 (3.5)
Severity of dementia, n (%)			
Mild	62 (15.7)	25 (11.6)	37 (20.7)
Moderate	163 (41.4)	87 (40.5)	76 (42.5)
Severe	169 (42.9)	103 (47.9)	66 (36.9)
Severe behavioral problems, n (%)	22 (5.6)	22 (10.2)	0 (0.0)

CES-D, Center for Epidemiologic Studies–Depression Scale; PWD, persons with dementia; SD, standard deviation; ZBI, Zarit Burden Interview.

Appendix 3

Scale Items in the Original ZBI, and the Top Models (Which Best Predict Significant Burden) as Identified by the Best-Subset Approach

Items in the Original ZBI	The Identified Top Models (Arranged by the Number of Items in the New Scale)																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Your relative asks for more help than he/she needs?															✓	✓	✓	✓	✓	✓	✓	✓
2. You don't have enough time for yourself?						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3. Stressed between caring and meeting other responsibilities?												✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4. Embarrassed over your relative's behaviors?										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5. Angry when around your relative?																					✓	✓
6. Your relative affects your relationship with others in a negative way?									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
7. Are you afraid what the future holds for your relative?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
8. Your relative is dependent on you?						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
9. Strained when you are around your relative?					✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10. Do you feel your health has suffered because of your involvement with your relative?		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
11. You don't have as much privacy as you would like, because of your relative?											✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
12. Your social life has suffered because you are caring for your relative?											✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
13. Uncomfortable about having friends over because of your relative?																				✓	✓	✓
14. Your relative seems to expect you to take care of him/her, as if you were the only one he/she could depend on?																						✓
15. You don't have enough money to care for your relative, in addition to the rest of your expenses?														✓	✓	✓	✓	✓	✓	✓	✓	✓
16. You will be unable to take care of your relative much longer?														✓	✓	✓	✓	✓	✓	✓	✓	✓
17. Do you feel you have lost control of your life since your relative's illness?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
18. You could just leave the care of your relative to someone else?																		✓	✓	✓	✓	✓
19. Uncertain about what to do about relative?								✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
20. You should be doing more for your relative?								✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
21. You could do a better job in caring for your relative?																✓	✓	✓	✓	✓	✓	✓
22. Overall, how burdened do you feel in caring for your relative?				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

ZBI, Zarit Burden Interview.

Appendix 4

Performance of ZBI-S in Identifying Significant Burden* in the Validation Sample
(n = 179)

Cut-off Score	Sensitivity	Specificity
≥0	1.00	0.00
≥1	0.98	0.18
≥2	0.97	0.27
≥3	0.94	0.47
≥4 [†]	0.86[‡]	0.73[‡]
≥5	0.75	0.84
≥6	0.60	0.89
≥7	0.42	0.94
≥8	0.31	0.97
≥9	0.15	0.99
≥10	0.03	1.00
AUROC (95% CI)	0.86 (0.80–0.91) [‡]	

AUROC, area under the receiver operating characteristics curve; CI, confidence interval, ZBI-S, Zarit Burden Interview—Screening version.

The optimal cut-off score is highlighted in bold.

*Significant burden was indirectly identified by the presence of significant depression (Center for Epidemiologic Studies–Depression score ≥16) in caregivers, which indicated the definitive need for further interventions.

[†]The optimal cut-off score based on a balance between sensitivity and specificity, with a preference for slightly higher sensitivity to reduce the false-negative rates in the screening scale.

[‡] $P < .001$.

Appendix 5. The Scale Items in the 3 Screening Versions of Zarit Burden Interview (ZBI)

The numbers to the left of the scale items correspond to the item numbers in the original ZBI. Each item is rated on a 5-point Likert scale based on how often a caregiver experiences a specific feeling when providing care (0 = never; 1 = rarely; 2 = sometimes; 3 = quite frequently; 4 = nearly always).

One-item variant of ZBI (Higginson et al⁴)

22. Overall, how burdened do you feel in caring for your relative?

Four-item variant of ZBI (Bédard et al⁵)

2. Do you feel that because of the time you spend with your relative that you don't have enough time for yourself?

3. Do you feel stressed between caring for your relative and trying to meet other responsibilities for your family or work?

9. Do you feel strained when you are around your relative?

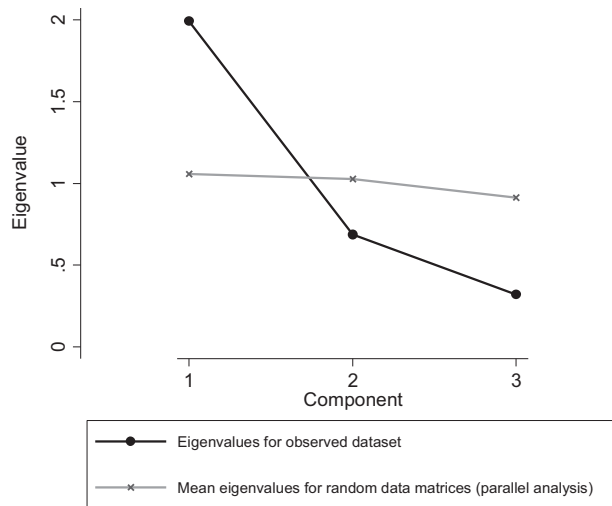
19. Do you feel uncertain about what to do about your relative?

Three-item ZBI–Screening version (derived from this study)

7. Are you afraid what the future holds for your relative?

10. Do you feel your health has suffered because of your involvement with your relative?

17. Do you feel you have lost control of your life since your relative's illness?



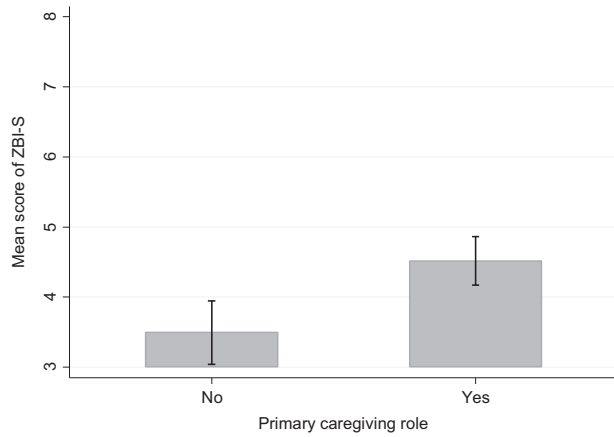
Appendix 6. Scree Plot from Exploratory Factor Analysis. Horn's parallel analysis suggests one factor in the new scale, as the first factor had eigenvalues which was greater than the mean eigenvalues generated from random data matrices*.

*Horn's parallel analysis⁶ has been recognized as a more accurate method to determine the number of factors in exploratory factor analysis.⁷⁻⁹ It is implemented by generating a large number of data matrices from random data and retaining factors that are greater than the mean eigenvalue generated from random data matrices.⁷

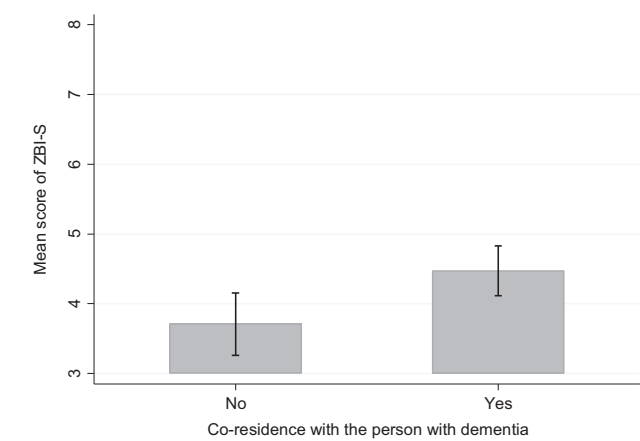
Appendix 7
Exploratory Factor Analysis of the ZBI-S and Factor Loadings of the Scale Items, Using Maximum-Likelihood Estimation and Oblique Rotation (oblimin)

Items in the New ZBI-S	Factor Loading
7. Are you afraid what the future holds for your relative?	0.48
10. Your health has suffered because of your involvement with your relative?	0.85
17. You have lost control of your life since your relative's illness?	0.80

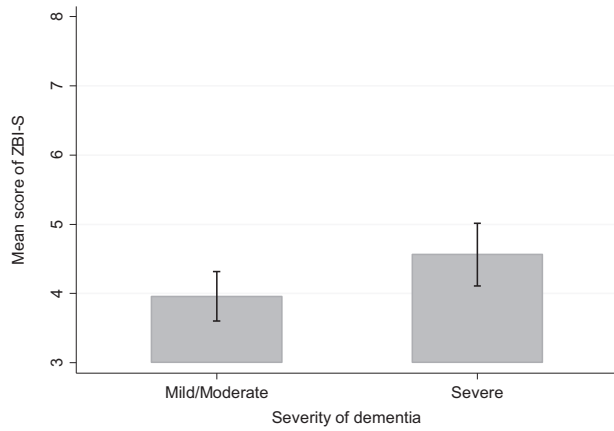
ZBI-S, Zarit Burden Interview—Screening version.



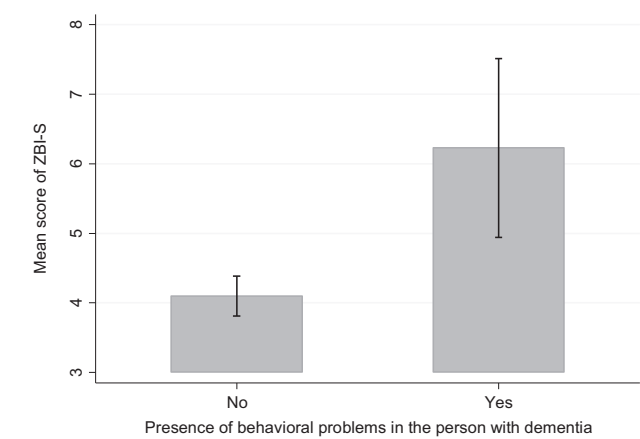
Appendix 8. Comparison of the Mean Scores of Zarit Burden Interview–Screening version (ZBI-S) Based on Primary Caregiving Role, in the Assessment of Known Group Validity. The mean scores were significantly different between the 2 groups ($P = .001$ in 2-sample t test). The vertical lines of error bar indicate the 95% confidence interval of the scores.



Appendix 9. Comparison of the Mean Scores of Zarit Burden Interview–Screening Version (ZBI-S) Based on Coresidence With the Person With Dementia, in the Assessment of Known Group Validity. The mean scores were significantly different between the 2 groups ($P = .012$ in 2-sample t test). The vertical lines of error bar indicate the 95% confidence interval of the scores.



Appendix 10. Comparison of the Mean Scores of Zarit Burden Interview–Screening version (ZBI-S) Based on Severity of Dementia, in the Assessment of Known Group Validity. The mean scores were significantly different between the 2 groups ($P = .038$ in 2-sample t test). The vertical lines of error bar indicate the 95% confidence interval of the scores.



Appendix 11. Comparison of the Mean Scores of Zarit Burden Interview–Screening Version (ZBI-S) Based on the Presence of Behavioral Problems in the Person With Dementia, in the Assessment of Known Group Validity. The mean scores were significantly different between the 2 groups ($P < .001$ in 2-sample t test). The vertical lines of error bar indicate the 95% confidence interval of the scores.

Appendix 12

Construct Validity as Shown by the Correlation Among Various Scales, Using Pearson Correlation Coefficient

	ZBI-S*
CES-D scale	
Depressed Affect subscale	0.72 [†]
Somatic Symptoms subscale	0.71 [†]
Interpersonal Problems subscale	0.46 [‡]
Positive Affect subscale	−0.38 [‡]
GAIN scale	−0.19 [‡]

CES-D, Center for Epidemiologic Studies Depression Scale; GAIN, Gain in Alzheimer care INstrument; ZBI, Zarit Burden Interview; ZBI-S, Zarit Burden Interview–Screening version.

*P values (after Bonferroni correction) are <.001 for all estimates, except for the correlation between ZBI-S and GAIN ($P = .296$).

[†]ZBI-S correlated strongly (correlation coefficient >0.50) with these scales as initially hypothesized.

[‡]ZBI-S correlated less strongly (correlation coefficient ≤0.50) with these subscales as initially hypothesized.

- Bedard M, Molloy DW, Squire L, et al. The Zarit burden Interview: A new short version and screening version. *Gerontologist* 2001;41:652–657.
- Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179–185.
- Courtney MGR. Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Pract Assess Res Eval* 2015;18:1–14.
- Gaskin CJ, Happell B. On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *Int J Nurs Stud* 2014;51:511–521.
- Dinno A. Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behav Res* 2009;44:362–388.

References

- Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia, PA: Lippincott-Raven; 1996.
- Patten SB. Major depression epidemiology from a diathesis-stress conceptualization. *BMC Psychiatry* 2013;13:19.
- Yap P, Luo N, Ng WY, et al. Gain in Alzheimer care INstrument—a new scale to measure caregiving gains in dementia. *Am J Geriatr Psychiatry* 2010;18:68–76.
- Higginson IJ, Gao W, Jackson D, et al. Short-form Zarit caregiver burden Interviews were valid in advanced conditions. *J Clin Epidemiol* 2010;63:535–542.