



Validation of two short versions of the Zarit Burden Interview in the palliative care setting: a questionnaire to assess the burden of informal caregivers

Martina B. Kühnel¹ · Christina Ramsenthaler¹ · Claudia Bausewein¹ · Martin Fegg¹ · Farina Hodiamont¹

Received: 5 June 2019 / Accepted: 29 December 2019

© The Author(s) 2020

Abstract

Purpose Several validated outcome measures, among them the Zarit Burden Interview (ZBI), are valid for measuring caregiver burden in advanced cancer and dementia. However, they have not been validated for a wider palliative care (PC) setting with non-cancer disease. The purpose was to validate ZBI-1 (ultra-short version and proxy rating) and ZBI-7 short versions for PC.

Methods In a prospective, cross-sectional study with informal caregivers of patients in inpatient (PC unit, hospital palliative support team) and outpatient (home care team) PC settings of a large university hospital, content validity and acceptability of the ZBI and its structural validity (via confirmatory factor analysis (CFA) and Rasch analysis) were tested. Reliability assessment used internal consistency and inter-rater reliability and construct validity used known-group comparisons and a priori hypotheses on correlations with Brief Symptom Inventory, Short Form-12, and Distress Thermometer.

Results Eighty-four participants (63.1% women; mean age 59.8, SD 14.4) were included. Structural validity assessment confirmed the unidimensional structure of ZBI-7 both in CFA and Rasch analysis. The item on overall burden was the best item for the ultra-short version ZBI-1. Higher burden was recorded for women and those with poorer physical health. Internal consistency was good (Cronbach's $\alpha = 0.83$). Inter-rater reliability was moderate as proxy ratings estimated caregivers' burden higher than self-ratings (average measures $ICC = 0.51$; $CI = 0.23\text{--}.69$; $p = 0.001$).

Conclusion The ZBI-7 is a valid instrument for measuring caregiver burden in PC. The ultra-short ZBI-1 can be used as a quick and proxy assessment, with the caveat of overestimating burden.

Keywords Caregivers · Caregiver burden · Palliative care · Validation studies · Zarit burden interview · Psychometrics

Introduction

According to the WHO definition, palliative care (PC) addresses the needs of patients and offers a support system to help the family cope during the patients' illness and in bereavement [1]. Not only family members but also friends or neighbours can be involved in taking care of a patient, and as

long as their support is not financially rewarded, they can be defined as informal caregivers [2].

Informal caregivers can become "patients" themselves, as their psychological morbidity is substantially higher compared with the general population [3]. There is a close relationship between the patient's perceived burden and that of the caregiver [4, 5], often leading to higher caregiver burden in the later stages of the patient's illness and a corresponding increase in need for physical and emotional support for caregivers [6, 7].

There is a growing number of intervention programmes [8, 9] aiming at caregiver outcomes, such as reducing caregivers' burden, improving caregivers' coping, or their quality of life. However, quantifying the impact of interventions is impossible without validated outcome measures for caregivers. A systematic review by Michels et al. showed that the majority of studies measuring informal caregiver outcomes in PC use

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00520-019-05288-w>) contains supplementary material, which is available to authorized users.

✉ Martina B. Kühnel
martina.kuehnel@med.uni-muenchen.de

¹ Department of Palliative Medicine, Munich University Hospital, LMU Munich, Munich, Germany

carer-specific measures, primarily measures of caregiver burden [10]. According to Michels et al. the Zarit Burden Interview (ZBI) [11] is one of the two most frequently used measures of burden [10], the other one being the caregiver reaction assessment (CRA) [12]. The ZBI, originally comprising 22 items [11, 13, 14], has several short forms including between four and twelve items, and the overall burden is assessed by the total score of all items, with a higher score representing greater caregiver burden [15–17]. Higginson et al. validated ZBI short versions in advanced conditions with caregivers of patients with advanced cancer, dementia, and acquired brain injury (ABI) [18]. The authors recommended using ZBI-6 and ZBI-7 (ZBI-6 plus ZBI-1) in the PC setting as they showed good validity, internal consistency, and discriminatory performance. Additionally, it was reported that the ZBI-1 might be suitable for screening [18].

However, although the ZBI is well-known and used, a formal validation of ZBI short versions in the PC setting using psychometric testing and Rasch analysis, and complementing the results Higginson et al. [18], is still lacking. Furthermore, the German ZBI 22-item version was validated by Braun et al. [19] for female caregivers of dementia patients, but not validated in a German PC setting yet.

The palliative care context differs from dementia due to often rapidly progressing diseases, and caregiving at the end of life causes the greatest caregiver burden. [20]

Therefore, the aim of this study was (1) to test the ZBI-7, the ZBI-6, and the ZBI-1 short versions for content validity, structural validity, construct validity, and reliability in the PC setting; (2) to confirm findings using Rasch analysis; (3) to evaluate the suitability of ZBI-1 as a proxy assessment for staff members; and (4) to evaluate the suitability of ZBI-1 item as an ultra-short instrument for quick assessment based on validity, reliability, and Rasch analysis.

Methods

Design

This is a prospective, cross-sectional validation study. Psychometric properties are reported according to the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) guidelines [21, 22] and the quality criteria for measurement properties of health status questionnaires by Terwee et al. [23]. The Ethics Committee of the Ludwig-Maximilians University Munich approved the study (REC-No 772–16).

Setting and population

The study was conducted in the Department for Palliative Medicine at Munich University Hospital. Informal caregivers

of patients treated by the hospital support team, and the home care team were consecutively recruited.

In the inpatient PC unit, questionnaires were included in the pre-intervention assessment of a randomized controlled trial evaluating an intervention for informal caregivers ([Clinical.Trials.gov](https://clinicaltrials.gov) registration NCT02325167). The combination of the two studies was approved by the university's ethics committee.

Inclusion criteria were being an informal caregiver of a palliative patient, a minimum age of 18 years, proficiency in written and spoken German, and the ability to give written informed consent. Caregivers with poor general condition, caregivers of patients who had been admitted to PC the same day, or who were imminently dying were excluded. Eligibility for inclusion was assessed by a staff member.

All participating caregivers and patients provided written informed consent. Consent of a legal guardian was sought for those patients unable to give consent.

Data collection

Data were collected between February 2017 and February 2018 using self-assessed questionnaires. Demographic data included age, sex, ethnicity, religion, highest academic qualification, profession, and marital status. Information on type of relation to the patient, role in caring for the patient, and the living status was collected for caregivers. Patient data were collected from medical notes and included age, date of PC admission, symptom burden at day of admission (via routinely collected Integrated PC Outcome Scale [24]), diagnosis, and date of discharge or death.

A member of the attending PC team assessed caregiver burden as a proxy using the ZBI-1 for inter-rater agreement. Staff members were asked for written informed consent and the following demographic data: age, sex, profession, work setting, and years of experience in PC.

Caregivers, who appeared highly burdened personally or in the assessments, were offered additional supportive talks by the multidisciplinary team.

Measurement instruments

Zarit Burden Interview-7 (ZBI-7): 7-item version of the original 22-item version measuring caregivers' physical and psychosocial burden on a five-point Likert scale [13, 18]. From the German translation by Braun et al. [19], we chose the seven items (see Table 1) recommended for use in PC by Higginson et al. [18]

Brief Symptom Inventory (BSI): measuring psychological distress and psychiatric disorders with 53 items on a five-point rating scale [25].

Distress thermometer: one-item measure with a 0–10 scale ranging from “No distress” to “Extreme distress” [26].

Short form 12:12-item version of the Short Form Health Survey measuring subjective health status on three- and five-point Likert scales [27].

Analysis

Descriptive statistics were used to describe sample distribution and distribution of responses. Missing data were imputed using expectation-maximization technique as data was missing completely at random, as indicated by the non-significant χ^2 statistic in Little’s MCAR test performed in SPSS [28].

Sample size

Two sample size calculations were conducted to power the study for detecting moderate reliability scores and to allow the detection of medium differences between known subgroups regarding the extent of burden ($d=0.3$ at a power of 80% and at a significance level of 95%). Sample size estimates ranged from 64 to 144 participants, with a minimum of 90 participants needed to detect known-group differences.

Validity analysis

Content validity and acceptability comprised the analysis of ceiling and floor effects, indicated if more than 15% of responses are in the highest or lowest category [23], as well as assessment of acceptability by analysis of missing items and user comments.

Structural validity: confirmatory factor analysis (CFA) were run to confirm that all items load on one latent factor, excluding the existence of subscales [29]. CFA was run with maximum likelihood estimation as it is robust to minor deviations from normality and accounts for missing data [30, 31]. Evaluation of model fit was based on fit indices and on the χ^2/df -ratio rather than on χ^2 , as the latter reacts sensitively to sample size [32]. A χ^2/df -ratio between 2 and 3 was regarded as indicative of acceptable data-model fit [32, 33]. Fit indices of CFI/TLI ≥ 0.90 were regarded acceptable, and root mean square error of approximation (RMSEA) < 0.08 was regarded as showing good fit [31].

Construct validity: we tested a priori hypotheses on scale-to-scale correlations with other measures, assuming that high correlations imply high convergent validity and suggest that the two scales measure similar concepts [34]. BSI, Distress Thermometer, and SF-12 were chosen, as they are well-known and established measurement instruments and, while not explicitly validated for caregivers in palliative care, have all been used in studies on this population [35, 36].

Twelve a priori hypotheses were formulated on ZBI-1, ZBI6, and ZBI-7—each correlating significantly with the BSI subscales depression and Global Severity Index, with the Distress Thermometer and the SF-12 subscale Mental Health Composite. Moderate correlations (0.4–0.7) were assumed, as all measures represent different aspects of burden-related caregiver outcomes. The family-wise alpha error rate was Bonferroni-corrected to a value of $0.05/12 = 0.004$.

Construct validity was also determined through known-groups comparisons [34]. Eight hypotheses were formulated. We hypothesised that burden would be higher for female caregivers, due to studies suggesting sex differences [37, 38].

Table 1 Short versions, item wording, and distribution of responses ($n = 84$)

Short versions				Responses (%)					
ZBI 7	ZBI 6	ZBI 1	Items	0	1	2	3	4	Miss
Y	Y		1. Do you feel you do not have enough time for yourself?	17.9	21.4	41.7	14.3	4.8	0
Y	Y		2. Do you feel stressed between caring and meeting other responsibilities?	19.0	26.2	28.6	20.2	6.0	0
Y	Y		3. Do you feel your relative affects your relationship with others in a negative way?	56.0	21.4	16.7	3.6	1.2	1.2
Y	Y		4. Do you feel strained when are around your relative?	34.5	27.4	27.4	9.5	1.2	0
Y	Y		5. Do you feel your health has suffered because of your involvement with your relative?	32.1	20.2	35.7	8.3	3.6	0
Y	Y		6. Do you feel you have lost control of your life since your relative’s illness?	48.8	17.9	20.2	9.5	2.4	1.2
Y		Y	7. Overall, how burdened do you feel in caring for your relative?	6.0	17.9	27.7	44.6	3.6	1.2

Y: yes

Italic: responses $> 15\%$; indicating floor-effects

Miss: percentage of missing items

Possible responses to item 1–6: 0, never; 1, rarely; 2, sometimes; 3, quite frequently; 4, nearly always

Possible responses to item 7: 0, not at all; 1, a little; 2, moderately; 3, quite a bit; 4, extremely

Furthermore, a block of hypotheses referred to (a) the relationship between caregivers and patient. It was hypothesised that burden would be higher for (i) parents or partners as losing a child conflicts with life cycle expectations, and losing a partner is ranked as one of the most stressful life events [39]; (ii) those living with the patient as studies suggest that there are more negative consequences for caregivers when caregiving in-house [40]; (iii) those giving physical care to the patient, and (iv) those who had power of attorney or legal guardianship for the patient as we suspected a relationship to burden, since caregivers are neither trained nurses nor legal guardians.

A second block of hypotheses referred to (b) caregivers who felt physically strained which can impact on caregivers' distress [38]. It was hypothesised that ZBI outcomes would be higher for (i) those who scored high on the SF-12 Physical Health Composite (via median split); those who due to physical health in the past 4 weeks (SF-12) (ii) had accomplished less or (iii) had been limited in work or activities. Parametric tests were used for all comparisons, complemented by non-parametric tests to account for non-normal distribution (*t*-test and Kruskal-Wallis H-test for hypothesis (a, i); *t*-test and Mann-Whitney-U tests for all other hypotheses). Hypotheses were tested using non-imputed data to avoid misleading results due to imputation.

The first block of known-group comparisons was tested to a Bonferroni-corrected alpha of $0.05/8 = 0.006$; and the last block to a corrected alpha of $0.05/3 = 0.017$.

Reliability analysis

Internal consistency was assessed as an aspect of reliability [41]. Cronbach's $\alpha = 0.7\text{--}0.9$ indicated internal consistency without item redundancy.

Inter-rater reliability between self-rating of burden and proxy rating by a staff member was examined with the Intraclass Correlation Coefficients (ICC) [21] and a two-way mixed model of the type consistency [42]. ICC < 0.5 indicated poor reliability, ICC of $0.5\text{--}0.75$ moderate, $0.75\text{--}0.90$ good, and ICC > 0.90 excellent reliability [42].

Rasch analysis

Rasch analysis complemented the validity analyses and tested items for use as an ultra-short version. The Rasch measurement model tests validity of unidimensional measures. It assumes that the response to a ZBI item is determined by the level of burden a person experiences (person fit) and the level of burden that the item represents (item fit). The Partial Credit Model was used which does not require equidistant categories and is suitable for ordinal-level data. ZBI-7 and ZBI-6 were compared with each other, and for ZBI-1 the self-rating data was compared with the one-item proxy rating by staff members.

Best-performing item candidates for the ultra-short version ZBI-1 were determined by item fit residuals ($< \text{and} > 2.5$), a summary mean item and person fit close to 0 (with SD = 1), ordered Likert response scale weightings for individual answer categories for each item, and the overall floor and ceiling effect for item parameters to person parameters. Overall model fit was assessed using the X^2 -test [43, 44].

CFA was run using IBM SPSS Amos 25 [45]. Rasch analysis was conducted using RUMM 2030 [46]. For all other analyses, SPSS version 25 was used [47]. A *p* value of < 0.05 was considered significant.

Results

Acceptability

Overall, 123 informal caregivers participated. Acceptability was assessed after 39 participants had completed the questionnaires. In open-response text fields, problems with the German translation of "care" were noted. Two participants commented "I don't nurse" and "No nursing" and 2.6–7.7% of items were missing. We therefore decided to change the wording of the German translation and employed the revised version on a sample of 84 participants. Percentage of missing items dropped to 0–1.2%, and overall, the revised version showed better characteristics than the first version. All following analyses in this study were conducted with data of the revised German version only ($n = 84$).

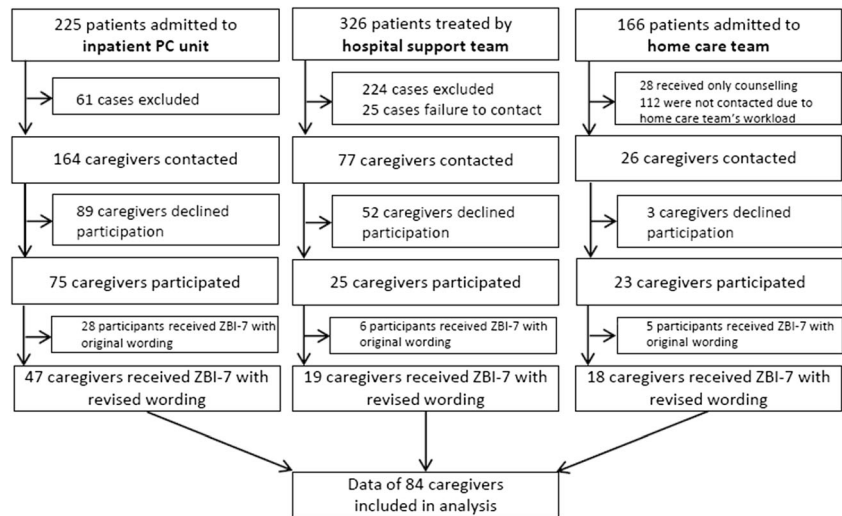
Characteristics of participants

Data of 84 participants who received the revised ZBI-7 were included in the analyses. Figure 1 shows the participant flow of the three settings. Most participants were female (63.1%; see Table 2); the mean age was 59.8 years (standard deviation (SD) 14.4). Approximately, one third of the participants held a university degree (32.1%), and the majority were married (76.2%). Participants were mostly partners (including wives or husbands) (53.6%) or children (32.1%) of the patients. Cancer was the prevailing diagnosis of the patients (79.8%). For characteristics of participating staff members see electronic supplementary material 1.

Structural validity

Scores were non-normally distributed for items 3, 6, and 7, with skewness of 1.215 (standard error (SE) = 0.264), 0.842 (SE = 0.264), and -0.582 (SE = 0.264), respectively; the latter left-skewed, all others right-skewed. Floor effects were observed for all items except item 7 on overall burden (see Table 1).

Fig. 1 Flow-chart of participants



The CFA analyses showed a good to moderate fit of a unidimensional model, meaning that all items in the ZBI short versions measure one construct, caregiver burden, only. Fit indices were good (CFI = 0.938, TLI = 0.907, standardized RMR = 0.0643), and RMSEA was moderate (RMSEA = 0.100, 90% confidence interval (CI) = 0.033–0.161). The χ^2/df -ratio was 1.84, also indicating a good fit to a unidimensional model. Overall, the fit indices and other measures (absence of Heywood cases, meaning negative variances or implausible values for variances and factor loadings) of fit confirm a unidimensional model of caregiver burden and the potential to shorten the ZBI further. All factor loadings were above 0.30, indicating good yet variable ability of individual items in the ZBI to measure the underlying construct of caregiver burden. Factor loadings varied between 0.41 for item 3 and 0.81 for item 7 on overall burden. Item 7 loaded highest onto the latent variable “burden” and showed the highest level of explained variance (see Table 3). ZBI-6 showed lower factor loadings and explained variance, as it lacks the overall item 7. The following results are therefore reported for ZBI-1 and ZBI-7 only.

Convergent validity

Correlations between the ZBI-1 and ZBI-7 scales and individual Zarit items with the Distress Thermometer, the SF-12 Mental Health subscale, the BSI global scale, and BSI depression subscale were analysed. Of the 12 a priori hypotheses nine, 75%, had hypothesised the correct direction of correlations (Bonferroni-corrected alpha level of 0.004; see Table 4).

Known-group comparisons

Caregiver burden measured with ZBI-7 was significantly higher for female caregivers. The results for the outcome ZBI-1 did not reach statistical significance, based on the

Bonferroni-corrected alpha level of 0.006 (ZBI-1 $t = 2.32$, $p = 0.023$; ZBI-7 $t = 2.96$, $p = 0.004$). No hypothesis in block (a) regarding relationship between carers and patient was significant.

In block (b), one of the three hypotheses concerning caregivers who felt physically strained was significant (b ii): Caregiver burden was significantly higher measured with ZBI-7 for those who had indicated on SF-12 that they had accomplished less in the past 4 weeks due to their physical health. The results for the outcome ZBI-1 did not reach statistical significance, based on the Bonferroni-corrected alpha level of 0.0017 (ZBI-1 $t = 2.01$, $p = 0.048$; ZBI-7 $t = 3.32$, $p = 0.001$). Comparisons were also run using non-parametric tests, yielding the same pattern of significant and non-significant results.

Reliability

Cronbach’s α for the ZBI-7 scale was 0.83 and was reduced with removal of any item. Item 7 on overall burden (ZBI-1) correlated highest with the whole ZBI-7 scale ($r = 0.73$) and if deleted reduced Cronbach’s α most (ZBI-6, Cronbach’s $\alpha = 0.78$).

ICC was significant for the 1-item ratings by staff members and informal caregivers. Agreement, however, was moderate for average measures (ICC = 0.51; CI = 0.23–0.69; $p = 0.001$). ICCs for the 1-item ratings of staff members and caregivers’ ZBI-7 self-rating were not significant ($p = 0.211$; single measures, ICC = 0.09; CI = -0.13–0.31; average measures ICC = 0.17; CI = -0.31–0.47).

Rasch analysis

All three models (ZBI-7, ZBI-6, and ZBI-1) showed good model fit. Mean of ZBI-7 item difficulty was 0.00 (SD = 0.63). Item 3 “affecting relationships” measured the highest

Table 2 Descriptive characteristics of participants ($n = 84$)

Characteristic	<i>n</i>	Percentage
Setting		
Palliative care unit	47	56.0%
Hospital support team	19	22.6%
Home care team	18	21.4%
Age	59.8 ^a	14.4 ^b
Sex		
Female	53	63.1%
Religion		
Catholic	36	42.9%
None	31	36.9%
Protestant	17	20.2%
Other		
Nationality		
German	82	97.6%
Other	2	2.4%
Education		
University degree	27	32.1%
Upper secondary	12	14.3%
Intermediate secondary	30	35.7%
Lower secondary	13	15.5%
Missing	2	2.4%
Marital status		
Married	64	76.2%
In relationship	13	15.5%
Single	4	4.8%
Divorced/separated	2	2.4%
Widowed	1	1.2%
Relationship with patient (“Patient is my...”)		
Wife/husband	43	51.2%
Partner	2	2.4%
Mother/father	27	32.1%
Sister/brother	7	8.3%
Grandmother/grandfather	1	1.2%
Friend	2	2.4%
Other	2	2.4%
Procuration for patient		
Procuration or legal guardianship	72	85.7%
No procuration or guardianship	12	14.3%
Diagnosis of patient		
Digestive tract cancer	17	20.2%
Genito-urinary cancer	12	14.3%
Brain cancer	8	9.5%
Lung cancer	7	8.3%
Gynaecological cancer	4	4.8%
Breast cancer	3	3.6%
Hemic cancer	3	3.6%
Other cancer	13	15.5%
Neurological disease	8	9.5%
Cardiovascular disease	2	2.4%
Other disease	7	8.3%

^a Mean^b SD**Table 3** Factor loadings of confirmatory factor analyses with EM-imputed data of ZBI-7 and ZBI-6

ZBI Item	ZBI-7		ZBI-6	
	λ	SMC	λ	SMC
1	0.64	0.41	0.66	0.43
2	0.77	0.60	0.79	0.62
3	0.49	0.24	0.51	0.26
4	0.51	0.26	0.48	0.23
5	0.76	0.58	0.75	0.56
6	0.51	0.26	0.50	0.25
7	0.81	0.65	-	-

SMC, squared multiple correlation

 λ standardized regression weight (factor loading)

levels of burden, while item 7 “overall burden” measured the lowest levels. There was no major deviation from the Rasch model as no item showed residuals of ± 2.5 and all χ^2 measures were non-significant (Bonferroni-corrected, $p < 0.001$, see electronic supplementary material 2).

The person-item threshold distribution showed a slight mismatch of item and person parameters (see electronic supplementary material 3). Items measured the medium to higher levels of burden. Person parameters (amount of burden as reported by caregivers), however, showed lower to medium values. For ZBI-1, the distribution of scores indicated lower person parameters for caregivers, indicating lower burden, than was observed for staff members’ proxy ratings. Item characteristic curves showed that items 5 “health suffered” and 7 “overall burden” marginally over-discriminated by differentiating well between caregivers with high or low burden. Interval-scale assumption via category probability curves yielded items 2 “meeting responsibilities,” and 4 “feeling strained” as most evenly distributed items. Moreover, item 7 “overall burden,” the designated item of the ZBI-1 ultra-short version, showed comparatively good fit to the Rasch model.

The fit of the self-rated caregiver version (location = 1.172, SE = 0.136, fit residual = -0.006) was better than the fit of the staff version (location = -1.172, SE = 0.153, fit residual = 0.715).

Discussion

Our aim for this study was to close the gap of a formal validation of the ZBI short versions in the PC setting. Additionally, the acceptability of the German ZBI was improved by the change of wording (report in preparation).

Concerning convergent validity, scale-to-scale correlations were significant but moderate, as expected, due to the comparison instruments measuring different aspects of burden-related

Table 4 A priori hypotheses and results for construct validity using spearman correlation coefficients of the ZBI with SF-12 and BSI ($n = 84$)

Hypothesis	Scales correlated	ρ	p
Moderate ^a convergent validity expected:			
Between ZBI scales and BSI (sub-)scale	<i>ZBI-1 + BSI Global Severity Index</i>	0.41	0.000
	<i>ZBI-6 + BSI Global Severity Index</i>	0.53	0.000
	<i>ZBI-7 + BSI Global Severity Index</i>	0.53	0.000
Between ZBI scales and the Distress Thermometer	ZBI-1 + BSI depression	0.36	0.001
	<i>ZBI-6 + BSI depression</i>	0.45	0.000
	<i>ZBI-7 + BSI depression</i>	0.45	0.000
Between ZBI scales and SF-12 subscale Mental Health Composite	<i>ZBI-1 + Distress Thermometer</i>	0.51	0.000
	ZBI-6 + Distress Thermometer	0.35	0.001
	ZBI-7 + Distress Thermometer	0.39	0.000
Between ZBI scales and SF-12 subscale Mental Health Composite	<i>ZBI-1 + SF12 Mental Health Composite</i>	- 0.40	0.000
	<i>ZBI-6 + SF12 Mental Health Composite</i>	- 0.48	0.000
	<i>ZBI-7 + SF12 Mental Health Composite</i>	- 0.49	0.000

^a Expected correlations: ρ (0.4)–(0.7)

Italics: correlations that were consistent with hypotheses

caregiver outcomes. Two of the eight hypotheses formulated on known groups were significant. As suggested by other studies [37, 38], burden was higher for female caregivers, and for those with poor physical health, which also concurs with other findings [38]. Unlike expected, caregiver burden was not higher for those who were partners or parents, who lived with the patient, physically nursed, or acted as legal guardian.

Our results on reliability for ZBI-7 (Cronbach's α 0.83) were only minimally higher than in Higginson et al.'s validation (α 0.82). [18]

Analysis of structural validity using CFA and Rasch analysis confirmed the unidimensional structure of the ZBI, allowing for use of the overall score as outcome measure. ZBI-7 showed advantages over ZBI-6 in factor loadings, explained variance, and internal consistency as the additional item 7 on overall burden proved to be the best item and the best choice as the ultra-short version ZBI-1.

Our results concerning ZBI-1 differ from Higginson et al.'s validation study where ZBI-1 for cancer caregivers showed the lowest discriminative ability and the lowest correlation with the 22-item version. Higginson et al. obtained 91% sensitivity and 53% specificity for ZBI-1, meaning that ZBI-1 oversensitively rated most caregivers as burdened [18]. In our study, ZBI-1 showed good fit with the Rasch model, which means that it discriminated very well between high and low burden and only when used as a proxy rating by staff members overestimated caregiver burden.

Using ZBI-1 as a proxy rating, staff members rated caregivers' level of burden higher than in caregivers' self-ratings, resulting in mediocre inter-rater reliability. Social desirability could have led to lower self-ratings, as caregivers might have presented themselves as more stable to prevent their ability to care being questioned. A potential consequence of personnel's

higher evaluation of burden could be the provision of support to caregivers who would not have asked for support themselves.

Rasch analysis and analysis of content validity suggested that items were constructed to measure higher levels of burden but caregivers reported lower levels. This may suggest a comparatively poor fit between sample and measure, resulting in false negative ratings of burden. However, participation bias could explain floor effects as participating caregivers possibly felt less burdened than those who decided to decline study participation. Dura and Kiecolt-Glaser reported a similar account of caregiver participation bias [48]. Additionally, caregivers included in this study were recruited from three specialized PC settings, which could have resulted in them being less burdened than caregivers who receive less professional support. Similarly, Higginson et al. reported lower levels of burden for advanced cancer caregivers, who had been recruited solely from specialized support facilities, while caregivers of patients with dementia and ABI showed higher levels of burden and had been recruited from diverse settings [18].

A strength of this study is that it is the first validation study of ZBI short versions that focusses on the PC setting alone. Participants were recruited in all three relevant PC settings. Additionally, this validation study was conducted with methods based on classical test theory and with Rasch analysis, which comprises aspects of item-response theory. Reliability of the ZBI-7 was higher than in previous studies and relative reliability was tested using inter-rater agreement. While the ZBI is well-known and used, our study closes the gap of a formal validation in the PC setting.

Limitations include rather low participant numbers in the home care setting due to low home care team staffing situation and high workload. Therefore, initially only few caregivers

had been contacted in this setting, and reasons for exclusion were not recorded consecutively. Inclusion decisions were hence recorded by a member of the study team. Additionally, it must be noted that the recruitment of the biggest part of caregivers was combined with an intervention study, to both preserve resources and spare caregivers, but the approach might have influenced caregivers' self-ratings. This study provides good validity for ZBI-1 as a proxy rating and potential as an ultra-short instrument, but because of lacking resources further analyses, e.g., of sensitivity or specificity, were not possible. Sample size was slightly smaller than the minimum of 90 participants needed to detect known-group differences, and subgroup comparison was infeasible due to unequal proportion of settings. However, results were obtained by combining methods of classical test theory and Rasch analysis and can therefore be regarded as robust.

In conclusion, this study complements earlier results of Higginson et al. [18]. ZBI-1 and ZBI-7 were shown to be valid in the PC context. ZBI-1 shows promising indication for use as an ultra-short instrument for caregiver burden while ZBI-7 could be used for more comprehensive measurement of caregiver burden, for example, when quantifying the impact of interventions aimed at caregivers in clinical trials and evaluation studies.

Acknowledgments The authors thank all informal caregivers and staff members who participated in our study.

Funding information Open Access funding provided by Projekt DEAL. This study was supported by the Verein zur Förderung von Wissenschaft und Forschung an der Medizinischen Fakultät der Ludwig-Maximilians-Universität.

Compliance with ethical standards

Ethical approval precludes the data being provided to researchers who have not signed the appropriate confidentiality agreement. These restrictions are as per the Ethics Committee of Ludwig-Maximilians University Munich which approved the study (No. 772-16). In accordance with ethical approval, all results are in aggregated form to maintain confidentiality and privacy. Data are held at the Klinik und Poliklinik für Palliativmedizin, Klinikum der Universität München, Ludwig-Maximilians University, Munich, Germany.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Health Organization. Definition of palliative care. <http://www.who.int/cancer/palliative/definition/en/>. Accessed April 7 2019
2. Payne S (2010) EAPC task force on family carers white paper on improving support for family carers in palliative care: part 1. *Eur J Palliat Care* 17(5):238–245
3. Grande G, Rowland C, van den Berg B, Hanratty B (2018) Psychological morbidity and general health among family caregivers during end-of-life cancer care: a retrospective census survey. *Palliat Med* 32(10):1605–1614
4. Pitceathly C, Maguire P (2003) The psychological impact of cancer on patients' partners and other key relatives: a review. *Eur J Cancer* 39(11):1517–1524
5. Hodges LJ, Humphris GM, Macfarlane G (2005) A meta-analytic investigation of the relationship between the psychological distress of cancer patients and their carers. *Soc Sci Med* 60(1):1–12
6. Kurtz ME, Given B, Kurtz JC, Given CW (1994) The interaction of age, symptoms, and survival status on physical and mental health of patients with cancer and their families. *Cancer* 74(7 Suppl):2071–2078
7. Hashemi M, Irajpour A, Taleghani F (2018) Caregivers needing care: the unmet needs of the family caregivers of end-of-life cancer patients. *Support Care Cancer* 26(3):759–766. <https://doi.org/10.1007/s00520-017-3886-2>
8. Chi N-C, Demiris G, Lewis FM, Walker AJ, Langer SL (2016) Behavioral and educational interventions to support family caregivers in end-of-life care: a systematic review. *Am J Hosp Palliat Care* 33(9):894–908
9. Harding R, List S, Epiphaniou E, Jones H (2012) How can informal caregivers in cancer and palliative care be supported? An updated systematic literature review of interventions and their effectiveness. *Palliat Med* 26(1):7–22
10. Michels CT, Boulton M, Adams A, Wee B, Peters M (2016) Psychometric properties of carer-reported outcome measures in palliative care: a systematic review. *Palliat Med* 30(1):23–44
11. Zarit SH, Reever KE, Bach-Peterson J (1980) Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist* 20(6):649–655
12. Hudson PL, Hayman-White K (2006) Measuring the psychosocial characteristics of family caregivers of palliative care patients: psychometric properties of nine self-report instruments. *J Pain Symptom Manag* 31(3):215–228
13. Zarit SH, Orr NK, Zarit JM (1985) The hidden victims of Alzheimer's disease: families under stress. New York University Press, New York
14. Conde-Sala JL, Turro-Garriga O, Calvo-Perxas L, Vilalta-Franch J, Lopez-Pousa S, Garre-Olmo J (2014) Three-year trajectories of caregiver burden in Alzheimer's disease. *J Alzheimers Dis* 42(2):623–633
15. Bedard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M (2001) The Zarit Burden Interview: a new short version and screening version. *Gerontologist* 41(5):652–657
16. Gort A, March J, Gómez X, Mazarico S, Ballesté J (2005) Short Zarit scale in palliative care. *Med Clin (Barc)* 124(17):651–653
17. Arai Y, Tamiya N, Yano E (2003) The short version of the Japanese version of the Zarit Caregiver Burden Interview (J-ZBI_8): its reliability and validity. *Nihon Ronen Igakkai zasshi Japanese journal of geriatrics* 40(5):497–503
18. Higginson IJ, Gao W, Jackson D, Murray J, Harding R (2010) Short-form Zarit Caregiver Burden Interviews were valid in advanced conditions. *J Clin Epidemiol* 63(5):535–542
19. Braun M, Scholz U, Homung R, Martin M (2010) Caregiver burden with dementia patients. A validation study of the German language

- version of the Zarit Burden Interview. *Z Gerontol Geriatr* 43(2): 111–119
20. Williams AM, Wang L, Kitchen P (2014) Differential impacts of care-giving across three caregiver groups in Canada: end-of-life care, long-term care and short-term care. *Health Soc Care Comm* 22(2):187–196
 21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19(4):539–549
 22. Terwee CB, Prinsen CA, Chiarotto A, Westerman M, Patrick DL, Alonso J, Bouter LM, De Vet HC, Mokkink LB (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 27(5):1159–1170
 23. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
 24. Schildmann EK, Groeneveld EI, Denzel J, Brown A, Bernhardt F, Bailey K, Guo P, Ramsenthaler C, Lovell N, Higginson IJ (2016) Discovering the hidden benefits of cognitive interviewing in two languages: the first phase of a validation study of the integrated palliative care outcome scale. *Palliat Med* 30(6):599–610
 25. Derogatis LR, Melisaratos N (1983) The brief symptom inventory: an introductory report. *Psychol Med* 13(3):595–605
 26. Mehnert A, Müller D, Lehmann C, Koch U (2006) [The German version of the NCCN distress thermometer: validation of a screening instrument for assessment of psychosocial distress in cancer patients]. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* (54):213–223
 27. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M (1998) Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA project. *International quality of life assessment. J Clin Epidemiol* 51(11): 1171–1178
 28. Little RJ, Rubin DB (2002) *Statistical analysis with missing data*, 2nd edn. Wiley, Hoboken, NJ
 29. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 63(7):737–745
 30. Byrne BM (2010) *Structural equation modeling with AMOS: basic concepts, applications, and programming*. Routledge
 31. Brown TA (2010) *Confirmatory factor analysis for applied research*. Guilford Publications
 32. Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Meth Psychol Res* 8(2):23–74
 33. Ullman J (2001) Structural equation modeling. In: Tabachnick BG, Fidell LS (eds) *Using multivariate statistics*, 4th edn. Allyn & Bacon, Needham Heights, MA, pp 653–771
 34. Fayers PM, Machin D (2015) Scores and measurements: validity, reliability, sensitivity. In: Fayers PM, Machin D (eds) *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons, Chichester, pp 89–124
 35. Hudson PL, Trauer T, Graham S, Grande G, Ewing G, Payne S, Stajduhar KI, Thomas K (2010) A systematic review of instruments related to family caregivers of palliative care patients. *Palliat Med* 24(7):656–668
 36. Grande GE, Austin L, Ewing G, O’Leary N, Roberts C (2017) Assessing the impact of a Carer Support Needs Assessment Tool (CSNAT) intervention in palliative home care: a stepped wedge cluster trial. *BMJ Support Palliat Care* 7(3):326–334
 37. Pillemer S, Davis J, Tremont G (2018) Gender effects on components of burden and depression among dementia caregivers. *Aging Ment Health* 22(9):1162–1167. <https://doi.org/10.1080/13607863.2017.1337718>
 38. Dumont S, Turgeon J, Allard P, Gagnon P, Charbonneau C, Vezina L (2006) Caring for a loved one with advanced cancer: determinants of psychological distress in family caregivers. *J Palliat Med* 9(4):912–921
 39. Osterweis M, Solomon F, Green M (1984) Reactions to particular types of bereavement. In: *Bereavement: reactions, consequences, and care*. National Academies Press (US),
 40. Kaschowit J, Brandt M (2017) Health effects of informal caregiving across Europe: a longitudinal approach. *Soc Sci Med* 173:72–80
 41. Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 78(1):98
 42. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428
 43. Rasch G (1960) *Probabilistic models for some intelligence and attainment tests*. The Danish Institute of Educational Research, Copenhagen
 44. Tennant A, Conaghan PG (2007) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res (Hoboken)* 57(8):1358–1362
 45. Arbuckle JL (2017) *Amos*. 25.0 edn. IBM SPSS, Chicago
 46. RUMM 2030 (2010). RUMM Laboratory University of Western Australia,
 47. IBM SPSS Statistics for Windows (2017). 25.0 edn. IBM Corp., Armonk, NY
 48. Dura JR, Kiecolt-Glaser JK (1990) Sample bias in caregiving research. *J Gerontol* 45(5):P200–P204

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.