

RESEARCH ARTICLE

Inter-rating reliability of the Swiss easy-read integrated palliative care outcome scale for people with dementia

Frank Spichiger^{1,2*}, Thomas Volken³, Philip Larkin^{1,4}, André Anton Meichtry⁵, Andrea Koppitz²

1 UNIL, Institute of Higher Education and Research in Healthcare, Lausanne, Switzerland, **2** HES-So, School of Health Sciences Fribourg, Switzerland, **3** ZHAW, School of Health Sciences, Winterthur, Switzerland, **4** Palliative and Supportive Care Service, Lausanne University Hospital, Lausanne, Switzerland, **5** School of Health Professionals, Bern University of Applied Sciences, Bern, Switzerland

* frank.spichiger@hefr.ch



OPEN ACCESS

Citation: Spichiger F, Volken T, Larkin P, Meichtry AA, Koppitz A (2023) Inter-rating reliability of the Swiss easy-read integrated palliative care outcome scale for people with dementia. PLoS ONE 18(8): e0286557. <https://doi.org/10.1371/journal.pone.0286557>

Editor: Mitsunori Miyashita, Tohoku University: Tohoku Daigaku, JAPAN

Received: August 15, 2022

Accepted: May 10, 2023

Published: August 2, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0286557>

Copyright: © 2023 Spichiger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The full dataset and R-Code to reproduce the data are available at: Frank Spichiger, & Andrea Koppitz. (2023). Inter-rating reliability of the Swiss Easy-Read Integrated

Abstract

Background

The Integrated Palliative Care Outcome Scale for People with Dementia is a promising instrument for nursing home quality improvement and research in dementia care. It enables frontline staff in nursing homes to understand and rate the needs and concerns of people with dementia. We recently adapted the measure to include easy language for users from various educational backgrounds.

Objectives

In this study, we examine the inter-rating reliability of the Integrated Palliative Care Outcome Scale for People with Dementia for frontline staff in nursing homes.

Methods

In this secondary analysis of an experimental study, 317 frontline staff members in 23 Swiss nursing homes assessed 240 people with dementia from a convenience sample. Reliability for individual items was computed using Fleiss Kappa. Because of the nested nature of the primary data, a generalisability and dependability study was performed for an experimental IPOS-Dem sum score.

Results

The individual Integrated Palliative Care Outcome Scale for People with Dementia items showed kappa values between .38 (95% CI .3–.48) and .15 (95% CI .08–.22). For the experimental IPOS-Dem sum score, a dependability index of .57 was found. The different ratings and time between ratings explain less than 2% of the variance in the sum score. The different nursing homes make up 12% and the people with dementia make up 43% of the sum score variance. The dependability study indicates that an experimental IPOS-Dem sum score could be acceptable for research by averaging two ratings.

Palliative Care Outcome Scale for People with Dementia [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8036812>.

Funding: This work was supported by the Swiss Academy of Medical Sciences (<https://www.samw.ch/de.html>), the Gottlieb and Julia Bangert Foundation (grant number PC 20/17), and the Swiss Academy for Socratic Care (<https://www.maeeutik-schweiz.ch/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: AK, FS, PL and TV are the translators/developers of the Swiss easy-read IPOS-Dem. The Swiss easy-read IPOS-Dem is a secondary outcome measure in a trial where AK is the Principal Investigator. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Conclusion

Limited research has been conducted on the measurement error and reliability of patient-centred outcome measures for people with dementia who are living in nursing homes. The Swiss Easy-Read IPOS-Dem is a promising instrument but requires further improvement to be reliable for research or decision making. Future studies may look at its measurement properties for different rater populations or at different stages of dementia. Furthermore, there is a need to establish the construct validity and internal consistency of the easy-read IPOS-Dem.

Background

Dementia is a name given to a group of progressive cognitive diseases [1]. People with dementia may develop impaired functioning, memory, cognition and performance of activities of daily living [1]. According to Sleeman et al. [2], people with moderate to severe dementia face the prospect of health-related suffering. Evidence indicates that people with dementia have inadequate access to the palliative care required for their complex symptoms [2–5]. The complexity of caring for people with dementia arises from their multidimensional symptoms that influence their health; these symptoms also limit accurate prognostic assertions, palliation and treatment [1, 6–8]. In addition, the quality of life and care of people with dementia are also frequently impacted by compromised verbal communication [5, 9–11]. A structured, systematic symptoms assessment process that fosters communication among people with dementia, their family members and frontline staff may help identify symptoms, enable family members to gain insights into caring for people with dementia and improve therapy regimes [12–15]

In Switzerland, people with dementia live in nursing homes for an average of two years and often have multiple comorbidities [16], along with the main diagnosis of moderate to advanced dementia. Swiss nursing homes' usual care follows routinely used assessment instruments [17], namely the Resident Assessment Instrument (RAI-NH), 'Bewohner/-innen-Einstufungs-und-Abrechnungssystem' (BESA). Evaluations using these standardised instruments routinely occur only every six months. Frontline staff in Swiss nursing homes may not have the optimal skills to meet all the care needs of people with Dementia nor are there enough qualified staff [18]. Moreover, limited resources are available for frontline staffs support in Swiss nursing homes, resulting in a lack of systematic use of expertise, assessment instruments and evidence in everyday dementia care [19].

The Integrated Palliative Care Outcome Scale for People with Dementia (IPOS-Dem) is a tool used to inform assessments. The IPOS-Dem is multidimensional; using a person-centred approach, it asks about the most important symptoms and concerns of people with dementia. Using this instrument, frontline staff and family members can identify and address symptoms and concerns [15]. Being attentive to symptoms and concerns is considered a core process in dementia care [20]. The IPOS-Dem may also improve screening, communication, care quality and outcomes in routine care [15]. The IPOS-Dem and its family of tools are informed by empirical qualitative and quantitative work among various populations with palliative care needs [21, 22], and all versions can be downloaded at <https://pos-pal.org/>.

Thus far, no reliability data have been published for the IPOS-Dem [15, 23, 24]. Ellis-Smith et al. reported on feasibility, mechanisms of action and content validity after analysing focus group and semistructured interview data using directed content analysis [15, 22].

The original IPOS for general palliative care populations, from which the IPOS-Dem is derived, showed inter-rater reliability for 11 of 17 items, ranging from $\kappa_w = .4$ to $\kappa_w = .82$. Several items—including ‘Having had enough information’, ‘Having had practical matters addressed’, ‘Sharing feelings with family or friends’, ‘Drowsiness’, ‘Inner peace’ and ‘Dry or sore mouth’—repeatedly stood out in analyses, with the κ_w ranging between .02 and .29 [22].

The rater population—frontline staff working with people with dementia—is primarily made up of nurses with secondary vocational training degrees or without formal training but with several years of employment and clinical exposure [18, 25]. In Swiss nursing homes, less than one-fifth of the staff working with people with dementia are registered nurses; therefore, we included interns, healthcare assistants and nurses with secondary vocational training.

We developed a Swiss easy-read version of the IPOS-Dem [26] to use in the IPOS-Dem project, which has a stepped-wedge controlled randomised trial (SW-CRT) design [27]. Compared with its predecessor, the easy-read IPOS-Dem is more understandable and adapted to the skill-grade mix and competence of frontline staff in nursing homes [26]. The translation and adaptation to IPOS-Dem is described in detail in another study [26]. Here, we present the inter-rating reliability, generalisability and decision study for the easy-read IPOS-Dem, as assessed by frontline staff. Aspects of the validity of the IPOS-Dem will be reported separately to follow Kottner et al.’s [28] Guidelines for Reporting Reliability and Agreement Studies (GRRAS).

Methods

This is a secondary analysis of a multicentre experimental study with a total of 15 time-shifted assessment periods. For the analysis presented in the present study, data from the baseline measurement period were used. The sample size was determined by power calculations for the overarching SW-CRT, in which the IPOS-Dem was applied. The psychometric analysis of IPOS-Dem was preplanned during the SW-CRT preparation. For this SW-CRT, we aimed to enrol 220 people with dementia living in 22 nursing homes [27] between September 2020 and October 2021. Regarding the raters, we aimed to enrol 20 frontline staff members per nursing home, resulting in a rater population of 440 people. The sample of people with dementia was determined by the nursing homes and based on the agreement of people with dementia to participate (i.e., a convenience sample). The raters were also assigned according to convenience; therefore, no comparison among different levels of training or experience was undertaken. The detailed recruitment process is described in the SW-CRT protocol cited above.

Ethical approval and consent to participate

The study was approved by the Research Ethics Committee of the canton of Zurich, Switzerland (BASEC-ID: 2019–01847) and was conducted in line with the principles of the Helsinki Declaration [29]. The overarching trial was registered with DRKS00022339. All participants and/or their respective attorneys signed written informed consent for participation and (as outlined in the PLOS consent form) publication. All raters have signed written informed consent for participation and (as outlined in the PLOS consent form) publication.

Population

People with dementia. People with dementia were included if they (a) were not hospitalised at baseline and, therefore, were physically present in the nursing home at the

commencement of the study, (b1) had a diagnosis of vascular dementia or Alzheimer's disease or (b2) had minimum data sets (MDS) data indicating symptoms of dementia.

Frontline staff. Frontline staff members were invited to participate if they (a) were at least 18 years old, (b) had a tenure of at least 3 months in the nursing home, (c) worked at least 20% of the full-time equivalent, provided continuing care to people with dementia (d) and were able to communicate in German.

Data collection

Each participating nursing home was assigned a clinical champion, that is, a full-time on-site employee who oversaw recruiting, data collection and the general study coordination with the study team, as outlined in the overarching SW-CRT protocol [27]. At baseline, the clinical champions entered the demographic and clinical details of the people with dementia, as derived from their nursing homes' MDS [30, 31], into our research electronic data capture (REDCap) data management system [32]. A survey developed for the frontline staff was completed by them directly following a training session. The participating staff had 120 minutes of on-site introductory training, and they attempted to complete an assessment for a chosen case using the IPOS-Dem.

Frontline staff were explicitly informed during the training—through an informed consent discussion and written material—that inter-rating agreement was being assessed at baseline. For the reliability study, staff independently assessed people with dementia during the baseline period of 30 days. There were no data captured on which of the staff members submitted the IPOS-Dem to the clinical champion. The clinical champion, however, assured that two independent staff members assessed IPOS-Dem independently during baseline. Staff independently rated and completed the instruments for people with dementia between August 2020 and January 2022. Staff were never blinded to clinical information about the people with dementia and completed the paper version of the IPOS-Dem. The data were subsequently entered into REDCap [32], browser-based software that could give continuous feedback to the clinical champion entering the data (e.g., erroneous or missing data). Automated tests run by REDCap also checked the data for plausibility and completeness.

Study measures. The Swiss easy-read version of the IPOS-Dem consists of 27 items related to physical, psychological, spiritual and practical concerns [26]. While mostly taking a self-proxy perspective [33], it asks three types of questions. After an introduction, there are three open questions about main issues during the last week the person with dementia had from the person with dementia's, the frontline staff's and the family member perspective. Following the textboxes, the user is asked to rate a 19-item list of symptoms and concerns regarding how much the symptoms and concerns impacted the person with dementia during the last week, in their opinion. These items are scored on a 5-point scale ranging from 0 (not at all) to 4 (very severe), with each point having its own descriptor. The symptom list continues with eight more questions, switching to a proxy-proxy perspective by asking how frequently a situation occurred. These items are scored on a 5-point scale ranging from 0 (not at all) to 4 (always), with each point again having its own descriptor. IPOS-Dem closes with three scoreable 'wild card' symptom fields. The IPOS-Dem was completed independently by frontline staff at the baseline of a cluster-randomised trial. The clinical champions oversaw frontline staff members' independent completion of two assessments per person with dementia at baseline. In previous studies [15], it took frontline staff on average between 4 and 12 minutes to complete IPOS-Dem, depending on their experience with the instrument.

People with dementia's sociodemographic information was captured by the clinical champion at baseline, as derived from the nursing home minimum datasets and charts at the time

point. The minimum datasets in Swiss nursing homes we referred to are a translation of the RAI-NH [30] or BESA [31]. The extracted chart and minimum dataset data were gender, marital status, nursing home, dementia type (if diagnosed) and dementia severity (if diagnosed).

Analysis

For each rating, an experimental IPOS-Dem sum score was calculated by adding the individual item responses of the 27 standard items. The scores are added with list-wise deletions of missing and 'do not know' responses. To inform the analyses of inter-rating reliability, we calculated information on the duration between the two IPOS-Dem assessments at baseline and developed an experimental sum score. The sum score was computed per assessment, with the list-wise exclusion of missing or 'do not know' ratings. The answer option 'do not know' was handled as missing. If not stated otherwise, missing data were excluded pairwise from the item-wise analyses. Sociodemographic and clinical data were analysed for the frontline staff, as well as the people with dementia using frequencies, proportions, ranges and distributions, both per nursing home and in total, with the tidyverse package 1.3.2 for R 4.1.2 [34, 35]. The IPOS-Dem item scores were described in a similar manner.

Item-wise analysis of inter-rating reliability. Fleiss' kappa is an extension of Cohen's kappa and can be used for more than two raters [36]; it considers the proportion of agreement beyond chance that would be expected if all ratings had been randomly scored. Fleiss' kappa ranges from 0 to 1, with values closer to 1 indicating higher inter-rater reliability. The coefficient (κ) is computed by the proportions of expected (\bar{P}_e) and observed (\bar{P}) agreements between ratings: $\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$. To complement the reporting, the percentage of agreement per item was also calculated and is presented in tables.

Generalisability study. Generalisability theory allows for the estimation of reliability for various combinations of raters in complex study designs [37]. Our design was based on 460 observations, with four additional factors: 230 people with dementia; 24 different durations between two assessments; 23 clusters and two ratings. This was a nested design, where some factors were nested within levels from other factors. Therefore, the ratings were nested within durations between the two assessments and nursing homes. Furthermore, people with dementia are nested within ratings and nursing homes. The reliability of the experimental IPOS-Dem sum scores is expressed by generalisability coefficients. Like an intraclass correlation coefficient, the generalisability coefficients indicate the reliability of a scale. By estimating variance components, the generalisability coefficients can be calculated. The variance components are estimated using a restricted maximum likelihood approach.

The variance components were estimated with the experimental IPOS-Dem sum score as the outcome variable and each of the factors (person with dementia, rating, cluster and time between assessments) as a random effect. Reliability was then quantified, with the universe score being the expected IPOS-Dem sum score of a person with dementia over the facets of generalisation for rating but fixed for clusters and time between measurements. The index of dependability (Φ) of a single measurement is the ratio of a person with dementias' score variance to the observed score variance: $\Phi = \frac{\sigma_p^2 + \sigma_c^2 + \sigma_r^2}{\sigma_p^2 + \sigma_c^2 + \sigma_r^2 + \sigma_e^2}$. In this model, the index is computed with a formula for consistency rather than agreement. A consistency model was chosen because IPOS-Dem is considered complex and multidimensional; this was also done to adjust for chance agreement. Model fitting and variance component estimation were performed with the lmer package [38] in R [35] 4.1.2.

Additional analysis and criteria for interpretation. The dependability index Φ represents inter-rating reliability for one assessment sum score for a randomly chosen time and

cluster. To compute the reliability of the mean measure of k measurements, we undertook a decision study. This means that the error variance components are divided by k to quantify the reliability of an average sum score over k repetitions. This decision study can help determine how many repetitions (i.e., ratings) would be required to reach an acceptable dependability Φ . For our analysis, this was performed for $k = 1, 2, 3$, to six repetitions.

For the interpretation of the results, different interpretation criteria were used. The item floor and ceiling effects were interpreted according to the criteria proposed by McHorn and Tarlov [39]. Their defined threshold for such an effect to occur was 15%, that is, the proportion of the sample rated with the lowest (floor) or highest (ceiling) possible score possible. The κ was interpreted according to Fleiss' [40] classification. Fleiss' classification for the interpretation sets only two cut-off values; kappa values below .40 are deemed 'poor', kappa values between .40 and .75 should be considered 'fair to good', and all kappa values above .75 'are deemed excellent' [40]. The G- and D-Study index values can range from 0 to 1 and are interpreted according to Nunnally's proposed criteria [41]. Nunnally [41] described coefficients at .7 as 'modest' and sufficient for early stages of research for instrument development.

Results

Observations

We analysed data from 257 people who were recruited from 23 nursing homes. On average, frontline staff completed the two IPOS-Dem measures for the inter-rating reliability analysis at baseline of the SW-CRT within 6.1 days (standard deviation [SD] = 7.4). The majority completed both observations within the first week, while some took up to 30 days to complete the repeated assessments. The heterogeneity in the time between the two assessments per nursing home is illustrated in [S1 Table](#).

Sample characteristics

[Table 1](#) shows the sociodemographic and clinical details of the people with dementia. Because the data were derived from a multicentre trial, we refer the reader to [S1 Table](#) for an illustration of the heterogeneity between the nursing homes.

As expected, 79% of the frontline staff were involved in various nursing roles, as shown in [Table 2](#). Interns, therapists, chaplains and others made up 15% of the raters. The mean tenure was 6.5 years. (Please see [S1 Table](#), which illustrates the heterogeneity between the nursing homes.)

Item characteristics

The item characteristics for the baseline data are presented in [Table 3](#). At baseline, we were able to match between 139 and 239 ratings per item per person with dementia. The items 'Nausea', 'Shortness of breath' and 'Vomiting' showed substantial floor effects, with more than 80% of the answers concentrating on a rating of 0. For the items 'Family anxious or worried', 'Inner peace' and 'Lost interest', frontline staff chose 'Don't know' in more than 29% of the assessments. Additional item characteristics are provided in [S2 Table](#).

Inter-rating reliability. In terms of Fleiss' kappa, the values varied between .39 and .15, as shown in [Table 4](#). The proportions of exact agreement varied between 39% and 89.5%.

Generalisability and decision study for an experimental sum score

We computed matched IPOS-Dem sum scores for 230 people with dementia; further statistics are shown in [Table 5](#) below. The maximum possible sum score was 108, which was not reached in our sample.

Table 1. Sociodemographic and clinical details of people with dementia.

Variable	N (%)	Mean (SD)	Min–Max (Median)
People with Dementia	257 (100%)		
Gender			
<i>Female</i>	180 (70%)		
<i>Male</i>	77 (30%)		
Age		86 (7.29)	56–102 (86)
Marital Status			
<i>Single</i>	21 (8%)		
<i>Married</i>	70 (27%)		
<i>Divorced</i>	30 (12%)		
<i>Widowed</i>	136 (53%)		
Area of Residence			
<i>Intermediate</i>	176 (68%)		
<i>Rural</i>	43 (17%)		
<i>Urban</i>	38 (15%)		
Dementia			
<i>Alzheimer's</i>	83 (32%)		
<i>Vascular</i>	22 (9%)		
<i>Other</i>	106 (41%)		
<i>Not formally diagnosed</i>	46 (18%)		
Severity			
<i>Mild</i>	6 (2%)		
<i>Moderate</i>	81 (32%)		
<i>Advanced</i>	86 (34%)		
<i>Not applicable</i>	84 (32%)		

<https://doi.org/10.1371/journal.pone.0286557.t001>

We fitted a linear mixed model to the sum score with person, rating, cluster and occasion as random intercepts.

Based on the variance components shown in Table 6 we computed $\Phi = 0.58$ for a single rating on a random day in a random cluster. In addition, we computed Φ for a mean of k ratings ($k = 1, 2, 3, \dots, 6$) to identify an acceptable lower bound of reliability for the sum score as shown in Table 7.

Our dependability study indicates that an acceptable sum score above the .7 could be obtained by averaging the sum scores from two ratings.

Discussion

The present study aimed to assess the reliability of the newly developed, easy-read IPOS-Dem when used by frontline staff in nursing homes. We computed the generalisability coefficient from two ratings of an experimental sum score and the individual Fleiss' kappa for each item. The κ of the items was between .38 (95% CI .3–.48) and .15 (95% CI .08–.22), indicating 'poor' agreement ($\kappa < .4$) when interpreted with Fleiss [40] criteria. An experimental IPOS-Dem sum score was used to enable the computation of a reliability coefficient under the generalisability framework. The findings of these analyses show a G-coefficient of .58. Our decision study shows that, by averaging two ratings, acceptable reliability for research could be obtained. The generalisability study also showed that the differences between participating nursing homes could explain 12% of the variance in the sum IPOS-Dem scores. Only small

Table 2. Sociodemographic details of frontline staff (i.e., raters).

Variable	N (%)	Mean (SD)	Min–Max (Median)
Staff	311 (100%)		
Age	304 (98%)	43 (13.6)	18–70 (45)
Gender			
<i>Female</i>	277 (89%)		
<i>Male</i>	34 (11%)		
Tenure (years)		6.6 (6.6)	0–32 (5)
Occupation			
<i>Registered nurse</i>	108 (35%)		
<i>Nursing associate professionals</i>	58 (19%)		
<i>Health care assistants</i>	96 (31%)		
<i>Registered nurse (intern)</i>	9 (3%)		
<i>Nursing associate professionals (intern)</i>	19 (6%)		
<i>Intern</i>	1 (< 1%)		
<i>Other^a</i>	17 (5%)		
<i>Missing</i>	3 (< 1%)		
Education			
<i>Tertiary</i>	121 (39%)		
<i>Upper secondary</i>	137 (44%)		
<i>Lower secondary</i>	23 (7%)		
<i>Other</i>	28 (9%)		
<i>Missing</i>	2 (< 1%)		

^a ‘Other’ included: housekeeping staff, chaplains, volunteers and social workers

<https://doi.org/10.1371/journal.pone.0286557.t002>

fractions of the variance were explained by ratings or time between assessments alone. The high proportion of IPOS-Dem sum score variance (41%) explained by residual variance may indicate interactions and measurement errors that must be investigated in future studies. Furthermore, without further investigation into the validity of the IPOS-Dem, the construction of a sum score remains experimental.

Limitations and strengths

We were able to obtain data from a considerable sample of people with dementia and involve frontline staff with different backgrounds, experiences and education in the primary study. This is the first study to evaluate the psychometric properties of the IPOS-Dem in a larger sample.

The present study has several limitations that we want to highlight. First, we were not able to ensure blinding of the raters regarding prior findings, clinical information and the accepted reference standard measurements like the RAI MDS. Second, there is no consensus in the literature on the stability of the IPOS-Dem ratings, as well as the symptoms and concerns of people with dementia in general. Because routine measurement is undertaken every six months, the relatively research-inexperienced setting and the design of the overarching SW-CRT, we considered one month suitable. We could have determined the sample size based on acceptable CIs (i.e. $\pm 0.1/\pm 0.2$) for ICCs reported in previous IPOS studies presented above [42, 43]. With 256 people with dementia, however, we exceeded the typical recommended number of participants in reliability studies often based on rule of thumb ($n = 50$) [42]; the 95% CIs around the Fleiss kappa are provided in Table 5.

Table 3. Easy-read IPOS-Dem item characteristics.

Item	Mean Score	Score (SD)	None (%)	Some (%)	Moderate (%)	Severe (%)	Very Severe (%)	Don't Know (%)	N Matched Cases
Pain ^a	1.3	1.1	26.7	35.1	25.1	10	3.1	10.6	220
Shortness of breath ^a	0.2	0.6	83.5	10.8	4.5	0.8	0.4	5.7	232
Weakness ^a	1.5	1.1	20.9	33.3	29.8	11.2	4.9	4.9	234
Nausea ^a	0.2	0.6	83.9	10.7	3.8	1	0.6	11.4	218
Vomiting ^a	0.1	0.4	93.8	3.9	1.2	0.8	0.2	7.3	228
Poor appetite ^a	0.8	1	53.7	24.7	14.7	4.5	2.4	6.5	230
Constipation ^a	0.7	0.9	56.9	25.6	13.5	3	1.1	14.6	210
Sore or dry mouth ^a	0.4	0.9	75.9	12.4	8.3	1.5	2	18.7	200
Drowsiness ^a	1.5	1.1	21	31.4	28.8	12	6.7	6.5	230
Poor mobility ^a	1.2	1.4	43.6	22.2	12.5	11.3	10.3	4.5	235
Sleeping problems ^a	0.8	1	51.9	26.2	12.8	7.5	1.7	11.4	218
Diarrhoea ^a	0.3	0.7	78.4	14.3	5	1.5	0.8	11.8	217
Dental problems ^a	0.6	1	70.2	14.5	8.7	3.2	3.4	14.2	211
Swallowing problems ^a	0.5	1	74.5	13.1	6.3	2.9	3.3	6.5	230
Skin breakdown ^a	1	1.1	44.8	27.1	16.6	8.9	2.6	4.1	236
Difficulty communicating ^a	1.6	1.4	31.7	19	22.2	14.1	13.1	4.1	236
Hallucinations and/or delusions ^a	0.8	1.1	61	16.5	11.2	7.9	3.3	19.9	197
Agitation ^a	1.7	1.3	22.6	21.8	27.7	16	11.8	2.8	239
Wandering ^a	1.3	1.4	45.3	15.1	17.1	12.2	10.2	6.5	230
Anxious or worried ^a	1.8	1.1	15.9	20.3	37.8	21.5	4.4	3.7	237
Family anxious or worried ^a	1.6	1.3	28.8	22.6	24.8	11.8	12	43.5	139
Felt depressed ^a	1.5	1	18	29	36.4	14.8	1.9	13	214
Lost interest ^a	1.1	1.2	42.7	22.3	19.5	12.3	3.2	31.3	169
Inner peace ^a	1.5	0.9	9.7	48.2	25.8	13.4	3	29.3	174
Able to interact ^a	1.5	1.3	29.5	24.9	19.9	18.9	6.8	3.3	238
Irritable or aggressive ^a	1.3	1	27.6	26	33.2	12.1	1.2	3.7	237
Practical matters ^a	1.4	1.1	24.5	34.3	24.9	10.6	5.7	14.6	210

Item characteristics for baseline data. Items are ordered as they occur in the easy-read IPOS-Dem.

^aItems with floor effect (more than 15% of answers in lowest category).

<https://doi.org/10.1371/journal.pone.0286557.t003>

The assignment of assessors to people with dementia was delegated to clinical champions, and the assessors' skills and grades were not linked to their respective ratings. The sample of people with dementia was rather heterogeneous, with a fifth lacking a formal diagnosis and different stages of reported severity. The lack of a severity assignment in a third of the sample deterred us from analysing the subsamples of the population and may also have contributed to the observed reliability. To control for the lack of assessment, the use of dementia staging instruments like FAST [44] at the baseline of research projects is highly recommended instead of relying on routine data. These shortcomings of the reported design may contribute to a major part of the unexplained variability in the sum scores.

Comparison with other instruments for people with dementia

QUALIDEM [45] was developed for observation-based quality of life assessment in people with dementia living in nursing homes. Ettema et al. [45] developed a scale for rating by nursing assistants, placing their scale within a similar scope as the IPOS-Dem. In their study, 68

Table 4. Item-wise reliability coefficients and proportions of agreement.

Item	Kappa	CI Lower Bound	CI Upper Bound	Don't Know	Agreement	Adjacent	Two Scores Apart	Three Scores Apart	Four Scores Apart
Pain ^a	0.33	0.25	0.41	10.6	50.7	38.5	9.5	1.4	0
Shortness of breath ^a	0.35	0.25	0.45	5.7	80.6	15.1	3.4	0.4	0.4
Weakness ^a	0.15	0.08	0.22	4.9	37.2	50.9	9.8	1.7	0.4
Nausea ^a	0.39	0.28	0.49	11.4	80.9	14.5	3.2	0.9	0.5
Vomiting ^a	0.21	0.11	0.31	7.3	89.5	6.1	2.2	1.8	0.4
Poor appetite ^a	0.25	0.17	0.33	6.5	52.8	33.3	10.4	2.6	0.9
Constipation ^a	0.28	0.19	0.37	14.6	56.2	29.5	10.5	3.8	0
Sore or dry mouth ^a	0.3	0.21	0.4	18.7	72.8	14.9	8.9	2	1.5
Drowsiness ^a	0.22	0.15	0.29	6.5	40.7	45.9	10.8	2.2	0.4
Poor mobility ^a	0.29	0.22	0.37	4.5	49.4	32.3	13.2	4.3	0.9
Sleeping Problems ^a	0.28	0.2	0.37	11.4	54.3	35.6	9.1	0.5	0.5
Diarrhoea ^a	0.34	0.24	0.44	11.8	74.3	20.2	2.8	2.3	0.5
Dental Problems ^a	0.39	0.3	0.48	14.2	71.8	19.7	5.6	2.3	0.5
Swallowing problems ^a	0.31	0.23	0.4	6.5	71.3	18.3	7	0.9	2.6
Skin breakdown ^a	0.28	0.2	0.35	4.1	50	32.2	12.7	4.2	0.8
Difficulty communicating ^a	0.31	0.24	0.37	4.1	46.2	30.1	17.8	5.1	0.8
Hallucinations and/or delusions ^a	0.34	0.25	0.42	19.9	61.6	18.7	13.8	3.9	2
Agitation ^a	0.26	0.19	0.32	2.8	41.8	37.2	17.6	2.9	0.4
Wandering ^a	0.33	0.26	0.41	6.5	52.4	24.2	16.5	5.6	1.3
Anxious or worried ^a	0.21	0.14	0.28	3.7	40.9	46	11.4	1.3	0.4
Family anxious or worried ^a	0.24	0.16	0.33	43.5	41.3	29.7	20	5.8	3.2
Felt Depressed ^a	0.24	0.16	0.31	13	44	42.1	11.1	2.3	0.5
Lost interest ^a	0.2	0.11	0.29	31.3	44.1	30.5	16.9	7.9	0.6
Inner peace ^a	0.17	0.08	0.26	29.3	45.1	46.7	7.1	0.5	0.5
Able to interact ^a	0.27	0.2	0.34	3.3	44.1	37	13.9	4.6	0.4
Irritable or aggressive ^a	0.26	0.18	0.34	3.7	46	42.6	10.1	1.3	0
Practical matters ^a	0.18	0.1	0.25	14.6	39	41.3	14.1	4.7	0.9

Fleiss' kappa (κ) from two matched independent frontline staff assessments, including 95% confidence intervals (Cis) and proportions of agreement per IPOS-Dem. Items are ordered as they occur in the easy-read IPOS-Dem.

^a Items with floor effect (more than 15% of answers in lowest category).

<https://doi.org/10.1371/journal.pone.0286557.t004>

raters assessed 238 people with very severe dementia. Ettema et al. subsequently calculated an overall reliability coefficient between .55 and .79. With later improvements in the German translation of QUALIDEM, reliability coefficients for individual items were improved. This was achieved by increasing the number of response options from four to seven and by the

Table 5. Characteristics of the IPOS-Dem sum scores for both ratings.

Statistics	1 st assessment sum score	2 nd assessment sum score
Number of matched cases	230	230
Mean (SD)	25.3 (13.0)	28 (13.7)
Median (IQR)	25 (17.75)	27 (17.75)
Range (min–max)	0–73	0–78

<https://doi.org/10.1371/journal.pone.0286557.t005>

Table 6. Variance components with respective proportions.

Component	Absolute variance component	% variance component
Person σ_p^2	79.28	42.71
Time between ratings σ_T^2	3.25	1.75
Rating σ_R^2	3.36	1.81
Cluster σ_C^2	22.99	12.39
Residuals σ^2	76.73	41.34

<https://doi.org/10.1371/journal.pone.0286557.t006>

Table 7. Dependability coefficients for multiple ratings.

Number of ratings (k)	Coefficients
1	0.579
2	0.733
3	0.805
4	0.846
5	0.873
6	0.892

<https://doi.org/10.1371/journal.pone.0286557.t007>

development of a detailed guide booklet [46, 47]. Dichter et al.'s German QUALIDEM study involved 36 people with advanced dementia who were rated by four caregivers with the revised QUALIDEM. In Dichter's QUALIDEM paper, only 6 out of 18 items showed floor or ceiling effects, although the authors opted to define floor effects by mean scores, with kappa values between .31 and .62. The items with the lowest reliability coefficients in the study were from the affect and social subscales. Similarly, some of the items that had low reliability in our study (i.e., 'Felt depressed' or 'Anxious or Worried').

Dichter et al. concluded that the QUALIDEM subscales generally showed sufficient reliability (between .64 and .91). However, in their related work, Dichter et al. [48] highlighted the lack of reliability investigations for instrument translations specific to the dementia population. The current Swiss guideline for dementia care in nursing homes [49] does not include any recommendations for instruments that can be used with all frontline staff members (e.g., Health care assistants, nursing associate professionals and interns).

Other popular instruments used for research on people with dementia are the Quality of Dying Instruments End-of-Life in Dementia Comfort Assessment in Dying (EOLD-CAD) and the Quality of Dying in Long-Term Care (QOD-LTC) [50–52]. However, the EOLD-CAD's reliability coefficient was moderate (0.59) and fair for the QOD-LTC (0.28) [50].

A review of instruments tested in long-term care settings by Ellis-Smith et al. [14] showed that different symptom-specific measures had reliability coefficients ranging between .76 and .73 for pain, .47 and .66 for measures of oral health and .20 for the single identified depression scale. In accordance with Dichter et al. and Kupeli et al. [48, 53], Ellis-Smith et al. highlighted that the evaluation of psychometric properties for many instruments is lacking. The findings regarding measurement properties identified above is in line with Soest-Poortvliet et al. [54], who looked at instruments evaluating end-of-life care and dying in long-term care residents. Their review of different instruments showed reliability coefficients between .25 and .59. These and our findings imply the difficulty [55] and complexity [48, 56] of evaluating patient outcomes in people with dementia.

Implications

Clinical practice. With the evidence reported in the present, the Swiss Easy-Read IPOS-Dem cannot be recommended for routine use in clinical practice or decision making. Further research into its psychometric properties needs to be conducted. To improve the reliability of the IPOS-Dem, additional actions targeting rating and observation procedures could be proposed. For example, a handbook could complement raters' training; this has already proven to be successful in developing other measures for this population [57, 58]. However, the underlying philosophy of user-friendly symptoms and concerns assessment permeates the IPOS family of measures [22]. An advantage of using the easy-language IPOS-Dem is its accessibility to frontline staff and family members in clinical practice without extensive training or a reading exercise in a handbook. This strength of the IPOS-Dem was theorised as mitigating setting-specific barriers to the effective implementation of palliative and person-centred care, such as high staff turnover, low incentives for professional staff development and the supersaturation of methods and instruments for geriatric care.

Research. With the evidence reported here, the Swiss Easy-Read IPOS-Dem experimental sum score might be used in research when averaged over two ratings. Because of these limitations, we caution against generalising our findings to other populations, settings and configurations of rater populations. Furthermore, the structural validity and validity of the sum score must be investigated first. Future studies investigating the reliability of the easy-read IPOS-Dem may avoid specific sources of variation in the ratings. There are a few options by means of restrictions in the design of such psychometric studies. A classical fully crossed design to determine test-retest and interrater reliability could be realised. First, researchers could restrict the rater population regarding qualifications and clinical exposure in a future study. Second, rigid assessment scheduling could be imposed on the day, the time between assessments and other factors. To date, there has been no guidance on the frequency at which routine assessments of symptoms and concerns in people with dementia should be conducted; therefore, we had no guiding frequency for imposing limitations on the scheduling of assessments or rater-subject assignments. Further improvements and changes regarding implementation and development will be derived from the experience of our colleagues at the United Kingdom Outcomes Assessment and Complexity Collaborative [59] and findings from the Australian Palliative Aged Care Outcomes Collaborative [12].

Conclusion

Comprehensive studies on the reliability of multidimensional instruments for people with dementia living in nursing homes have been infrequent. Especially in translated measures, reviews have not reported many publications on this measurement property. Generally, the reliability coefficients of most instruments to rate individual symptoms, quality of care or health-related quality of life in people with dementia hover below acceptable thresholds for clinical decision making and research. Some of the easy-read IPOS-Dem items have shown comparably poor coefficients. The experimental IPOS-Dem sum score may be reliable if averaged over two ratings. However, its validity needs to be investigated first. The present study has provided comprehensive information on the statistical parameters of measurement properties in the Swiss easy-read IPOS-Dem for its intended rater population. Our research shows that further development is needed to improve the easy-read IPOS-Dem to the point that the results can be considered reliable for research on caring quality and clinical decision making.

Supporting information

S1 Table. Cluster-wise sociodemographic statistics. This file contains tabular data for each cluster in a long format.

(HTML)

S2 Table. Item characteristics. This file shows additional item characteristics for the easy-read IPOS-Dem and complements [Table 3](#).

(HTML)

Acknowledgments

We would like to thank the frontline staff who were involved in this study. Furthermore, we wish to thank the clinical champions who participated: A. Beqiri, R. Benz, M. Bonaconsa, A. Brunner, A. Conti, M. Deflorin, D. Deubelbeiss, L. Ebener, S. Egger, E. Eichinger, D. Elmer, A. Ermler, M. Fuhrer, C. Grichting, M. Havarneanu, H. Hettich, E. Hoffmann, E. Imgrueth, R. Juchli, I. Juric, K. Knöpfli, S. Kuonen, F. Laich, H. Meiser, N. Mergime, B. Michel, C. Ming, F. Müller, C. Niederer, G. Parkes, P. Piguet, A. Repesa, C. Ritz, B. Santer, A. Schallenberg, C. Schweiger, M. Spitz, and R. Strunck. Also, many thanks to F. Murtagh for discussing the results and IPOS-Dem with us.

Author Contributions

Conceptualization: Frank Spichiger, Thomas Volken, Andrea Koppitz.

Data curation: Frank Spichiger.

Formal analysis: Frank Spichiger, André Anton Meichtry.

Funding acquisition: Andrea Koppitz.

Investigation: Frank Spichiger.

Methodology: Frank Spichiger, Thomas Volken, Philip Larkin, Andrea Koppitz.

Project administration: Andrea Koppitz.

Resources: Frank Spichiger, Andrea Koppitz.

Supervision: Thomas Volken, Philip Larkin, Andrea Koppitz.

Validation: Thomas Volken, André Anton Meichtry, Andrea Koppitz.

Visualization: Frank Spichiger.

Writing – original draft: Frank Spichiger.

Writing – review & editing: Frank Spichiger, Thomas Volken, Philip Larkin, André Anton Meichtry, Andrea Koppitz.

References

1. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. 2020; 396: 413–446. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6) PMID: 32738937
2. Sleeman KE, de Brito M, Etkind S, Nkhoma K, Guo P, Higginson IJ, et al. The escalating global burden of serious health-related suffering: Projections to 2060 by world regions, age groups, and health conditions. *Lancet Glob Health*. 2019; 7. [https://doi.org/10.1016/S2214-109X\(19\)30172-X](https://doi.org/10.1016/S2214-109X(19)30172-X) PMID: 31129125
3. Honinix E, Van den Block L, Piers R, Onwuteaka-Philipsen BD, Payne S, Szczerbińska K, et al. Large differences in the organization of palliative care in nursing homes in six European countries: Findings

- from the PACE cross-sectional study. *BMC Palliat Care*. 2021; 20: 131. <https://doi.org/10.1186/s12904-021-00827-x> PMID: 34433457
4. Knaut FM, Farmer PE, Krakauer EL, De Lima L, Bhadelia A, Jiang Kwete X, et al. Alleviating the access abyss in palliative care and pain relief—An imperative of universal health coverage: the Lancet Commission report. *Lancet*. 2018; 391: 1391–1454. [https://doi.org/10.1016/S0140-6736\(17\)32513-8](https://doi.org/10.1016/S0140-6736(17)32513-8) PMID: 29032993
 5. van der Steen JT, Radbruch L, Hertogh CM, de Boer ME, Hughes JC, Larkin P, et al. White paper defining optimal palliative care in older people with dementia: A Delphi study and recommendations from the European Association for Palliative Care. *Palliat Med*. 2014; 28: 197–209. <https://doi.org/10.1177/0269216313493685> PMID: 23828874
 6. Eisenmann Y, Golla H, Schmidt H, Voltz R, Perrar KM. Palliative care in advanced dementia. *Front Psychiatry*. 2020; 11. <https://doi.org/10.3389/fpsy.2020.00699> PMID: 32792997
 7. Grünig A. Indikationskriterien für spezialisierte Palliative Care [Quality indicators for specialised palliative care]. Bundesamt für Gesundheit, Schweizerische Konferenz der kantonalen Gesundheitsdirektorinnen und -direktoren, editors. BBL, Vertrieb Bundespublikationen; 2014. <https://www.bag.admin.ch/dam/bag/de/dokumente/nat-gesundheitsstrategien/strategie-palliative-care/grundlagen/spezialisierte-indikationskriterien.pdf.download.pdf/indik-spez-pc.pdf>
 8. Schmidt T. Palliative Aspekte bei Demenz [Palliative Care in Dementia]. *Z Für Prakt Psychiatr Neurol*. 2022; 25: 26–31. <https://doi.org/10.1007/s00739-021-00773-6>
 9. Deuschl G, Maier W. S3-Leitlinie Demenzen [S3 Dementia Guideline]. Leitlinien Für Diagnostik und Therapie in der Neurologie, Deutsche Gesellschaft für Neurologie, editors. 2016. <https://www.dgn.org/leitlinien>
 10. Husebø BS, Ballard C, Sandvik R, Nilsen OB, Aarsland D. Efficacy of treating pain to reduce behavioural disturbances in residents of nursing homes with dementia: Cluster randomised clinical trial. *The BMJ*. 2011; 343: d4065. <https://doi.org/10.1136/bmj.d4065> PMID: 21765198
 11. Shim SH, Kang HS, Kim JH, Kim DK. Factors associated with caregiver burden in dementia: 1-year follow-up study. *Psychiatry Investig*. 2016; 13: 43–9. <https://doi.org/10.4306/pi.2016.13.1.43> PMID: 26766945
 12. Australian Government Department of Health. PACOP for clinicians—University of Wollongong—UOW. In: Palliative aged care outcomes programme [Internet]. 2022 [cited 22 Apr 2022]. <https://www.uow.edu.au/ahsri/pacop/pacop-for-clinicians/>
 13. Backhaus R, Hoek LJM, de Vries E, van Haastregt JCM, Hamers JPH, Verbeek H. Interventions to foster family inclusion in nursing homes for people with dementia: A systematic review. *BMC Geriatr*. 2020; 20: 434. <https://doi.org/10.1186/s12877-020-01836-w> PMID: 33126855
 14. Ellis-Smith C, Evans CJ, Bone AE, Henson LA, Dzingina M, Kane PM, et al. Measures to assess commonly experienced symptoms for people with dementia in long-term care settings: a systematic review. *BMC Med*. 2016; 14: 38. <https://doi.org/10.1186/s12916-016-0582-x> PMID: 26920369
 15. Ellis-Smith C, Higginson IJ, Daveson BA, Henson LA, Evans CJ. How can a measure improve assessment and management of symptoms and concerns for people with dementia in care homes? A mixed-methods feasibility and process evaluation of IPOS-Dem. *PLoS One*. 2018; 13: e0200240. <https://doi.org/10.1371/journal.pone.0200240> PMID: 29995932
 16. Ecoplan. Grundlagen für eine Nationale Demenzstrategie [Swiss national dementia strategy fundamentals]; Demenz in der Schweiz: Ausgangslage. Bern: Bundesamt für Gesundheit (BAG) / Schweizerische Konferenz der kantonalen Gesundheitsdirektorinnen und -direktoren (GDK); 2013. <https://www.bag.admin.ch/dam/bag/de/dokumente/nat-gesundheitsstrategien/nationale-demenzstrategie/grundlagen-nds.pdf.download.pdf/03-d-grundlagen-nds.pdf>
 17. Vettori A, von Stokar T, Petry C, Britt D. Mindestanforderungen für Pflegebedarfserfassungssysteme [Minimum requirements for care tariff systems]. 2017. https://www.infras.ch/media/filer_public/32/8c/328cd5ec-af19-4b41-ab8c-31119b51a440/mindestanforderungen_fur_pflegebedarfserfassungssysteme-1.pdf
 18. Vellani S, Zuniga F, Spilsbury K, Backman A, Kusmaul N, Scales K, u. a. Who's in the House? Staffing in Long-Term Care Homes Before and During COVID-19 Pandemic. *Gerontology and Geriatric Medicine*. 1. April 2022; 8:23337214221090804.
 19. Zúñiga F, Favez L, Baumann S. SHURP 2018 –Schlussbericht. Personal und Pflegequalität in Pflegeinstitutionen in der Deutschschweiz und Romandie [Swiss nursing homes resources project—final report. Human resources and quality of care in german- and french-speaking Switzerland]. Universität Basel; 2021. <https://shurp.unibas.ch/shurp-2018-publikationen/>
 20. McCance TV. Caring in nursing practice: The development of a conceptual framework. *Res Theory Nurs Pract*. 2003; 17: 101–116. <https://doi.org/10.1891/rtnp.17.2.101.53174> PMID: 12880216
 21. Bausewein C, Schildmann E, Rosenbruch J, Haberland B, Tänzler S, Ramsenthaler C. Starting from scratch: Implementing outcome measurement in clinical practice. *Ann Palliat Med*. 2018; 7: S253–S261. <https://doi.org/10.21037/apm.2018.06.08> PMID: 30180734

22. Murtagh FE, Ramsenthaler C, Firth A, Groeneveld EI, Lovell N, Simon ST, et al. A brief, patient- and proxy-reported outcome measure in advanced illness: Validity, reliability and responsiveness of the Integrated Palliative care Outcome Scale (IPOS). *Palliat Med.* 2019; 33: 1045–1057. <https://doi.org/10.1177/0269216319854264> PMID: 31185804
23. Ellis-Smith C, Evans CJ, Murtagh FE, Henson LA, Firth AM, Higginson IJ, et al. Development of a caregiver-reported measure to support systematic assessment of people with dementia in long-term care: The Integrated Palliative care Outcome Scale for Dementia. *Palliat Med.* 2017; 31: 651–660. <https://doi.org/10.1177/0269216316675096> PMID: 28618899
24. Hodiament F, Hock H, Ellis-Smith C, Evans C, de Wolf-Linder S, Jünger S, et al. Culture in the spotlight—Cultural adaptation and content validity of the integrated palliative care outcome scale for dementia: A cognitive interview study. *Palliat Med.* 2021; 35: 962–971. <https://doi.org/10.1177/02692163211004403> PMID: 33863246
25. Wicki MT, Riese F. Prevalence of dementia and organization of dementia care in Swiss disability care homes. *Disabil Health J.* 2016; 9: 719–723. <https://doi.org/10.1016/j.dhjo.2016.05.008> PMID: 27431767
26. Spichiger F, Keller Senn A, Volken T, Larkin P, Koppitz A. Integrated Palliative Outcome Scale for People with Dementia: Easy language adaption and translation. *J Patient-Rep Outcomes.* 2022; 6: 14. <https://doi.org/10.1186/s41687-022-00420-7> PMID: 35169943
27. Spichiger F, Koppitz AL, Wolf-Linder SD, Murtagh FEM, Volken T, Larkin P. Improving caring quality for people with dementia in nursing homes using IPOS-Dem: A stepped-wedge cluster randomized controlled trial protocol. *J Adv Nurs.* 2021 [cited 16 Jul 2021]. <https://doi.org/10.1111/jan.14953> PMID: 34235765
28. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud.* 2011; 48: 661–671. <https://doi.org/10.1016/j.ijnurstu.2011.01.016> PMID: 21514934
29. WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects. 2013. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
30. Gattinger H, Ott S, Saxer S. Interrater-Reliabilität und Übereinstimmung der Schweizer RAI-MDS Version 2.0. *Pflege.* 2014; 27: 19–29. <https://doi.org/10.1024/1012-5302/a000336>
31. Gattinger H, Ott S, Saxer S. Comparison of BESA and RAÍ: evaluating the outcomes of two assessment instruments for long-term residential care needs. *Pflege.* 2014; 27: 31–40.
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009; 42: 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010> PMID: 18929686
33. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: A conceptual framework for understanding multiple proxy perspectives. *Med Care.* 2005; 43: 493–499. <https://doi.org/10.1097/01.mlr.0000160419.27642.a8> PMID: 15838415
34. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019; 4: 1686. <https://doi.org/10.21105/joss.01686>
35. R core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33: 159–174. <https://doi.org/10.2307/2529310> PMID: 843571
37. Brennan RL. Generalizability theory. New York, NY: Springer New York; 2001. <https://doi.org/10.1007/978-1-4757-3456-0>
38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv; 2014.
39. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 1995; 4: 293–307. <https://doi.org/10.1007/BF01593882> PMID: 7550178
40. Fleiss JL, Levin B, Myunghee CP. Chapter 18: The measurement of interrater agreement. 3rd ed. *Statistical Methods for Rates and Proportions.* 3rd ed. John Wiley & Sons, Ltd; 2003. pp. 598–626.
41. Nunnally JC. Chapter 7: The assessment of reliability. In *Psychometric theory.* 2nd ed. New York: McGraw-Hill; 1978. pp. 264–266.
42. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: A practical guide.* Cambridge: Cambridge University Press; 2011.
43. Giraudeau B, Mary JY. Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med.* 2001; 20: 3205–3214. <https://doi.org/10.1002/sim.935> PMID: 11746313

44. Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: Reliability, validity, and ordinality. *Int Psychogeriatr*. 1992; 4: 55–69. <https://doi.org/10.1017/s1041610292001157> PMID: 1504288
45. Ettema TP, Dröes R-M, de Lange J, Mellenbergh GJ, Ribbe MW. QUALIDEM: Development and evaluation of a dementia specific quality of life instrument. Scalability, reliability and internal structure. *Int J Geriatr Psychiatry*. 2007; 22: 549–556. <https://doi.org/10.1002/gps.1713> PMID: 17152121
46. Dichter MN, Dortmann O, Halek M, Meyer G, Holle D, Nordheim J, et al. Scalability and internal consistency of the German version of the dementia-specific quality of life instrument QUALIDEM in nursing homes—A secondary data analysis. *Health Qual Life Outcomes*. 2013; 11: 91. <https://doi.org/10.1186/1477-7525-11-91> PMID: 23738658
47. Dichter MN. QUALIDEM: Userguide. Witten, Germany: German Center for Neurodegenerative Diseases (DZNE), Witten, Germany; 2016.
48. Dichter MN, Schwab CGG, Meyer G, Bartholomeyczik S, Halek M. Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements—A systematic review. *J Clin Epidemiol*. 2016; 70: 233–245. <https://doi.org/10.1016/j.jclinepi.2015.08.002> PMID: 26319270
49. Bieri G, Silva-Lima S, Widmer B. Begleitung, Betreuung, Pflege und Behandlung von Personen mit Demenz [Care, caring and therapy for people with dementia]. Bern: Bundesamt für Gesundheit (BAG) / Schweizerische Konferenz der kantonalen Gesundheitsdirektorinnen und -direktoren (GDK); 2020 p. 25. https://www.bag.admin.ch/dam/bag/de/dokumente/nat-gesundheitsstrategien/nationale-demenzstrategie/hf-angebote/3_5_langzeitpflege/empfehlungen-langzeitpflege.pdf.download.pdf/Brosch%C3%BCre_Demenz_Empfehlung_Langzeitpflege_DE.pdf
50. Pivodic L, Smets T, Van den Noortgate N, Onwuteaka-Philipsen BD, Engels Y, Szczerbińska K, et al. Quality of dying and quality of end-of-life care of nursing home residents in six countries: An epidemiological study. *Palliat Med*. 2018; 32: 1584–1595. <https://doi.org/10.1177/0269216318800610> PMID: 30273519
51. van Soest-Poortvliet MC, van der Steen JT, Zimmerman S, Cohen LW, Klapwijk Maartje S, Bezemer M, et al. Psychometric properties of instruments to measure the quality of end-of-life care and dying for long-term care residents with dementia. *Qual Life Res*. 2012; 21: 671–684. <https://doi.org/10.1007/s11136-011-9978-4> PMID: 21814875
52. Zimmerman S, Cohen L, van der Steen JT, Reed D, van Soest-Poortvliet MC, Hanson LC, et al. Measuring end-of-life care and outcomes in residential care/assisted living and nursing homes. *J Pain Symptom Manage*. 2015; 49: 666–679. <https://doi.org/10.1016/j.jpainsymman.2014.08.009> PMID: 25205231
53. Kupeli N, Candy B, Tamura-Rose G, Schofield G, Webber N, Hicks SE, et al. Tools Measuring quality of death, dying, and care, completed after death: Systematic review of psychometric properties. *Patient—Patient-Centered Outcomes Res*. 2019; 12: 183–197. <https://doi.org/10.1007/s40271-018-0328-2> PMID: 30141020
54. van Soest-Poortvliet MC, van der Steen JT, Zimmerman S, Cohen LW, Munn J, Achterberg WP, et al. Measuring the quality of dying and quality of care when dying in long-term care settings: A qualitative content analysis of available instruments. *J Pain Symptom Manage*. 2011; 42: 852–863. <https://doi.org/10.1016/j.jpainsymman.2011.02.018> PMID: 21620642
55. Rababa M. The role of nurses' uncertainty in decision-making process of pain management in people with dementia. *Pain Res Treat*. 2018; 2018: 1–7. <https://doi.org/10.1155/2018/7281657> PMID: 30155298
56. Gräske J, Meyer S, Wolf-Ostermann K. Quality of life ratings in dementia care—A cross-sectional study to identify factors associated with proxy-ratings. *Health Qual Life Outcomes*. 2014; 12: 177. <https://doi.org/10.1186/s12955-014-0177-1> PMID: 25495548
57. Arons AMM, Wetzels RB, Zwijsen S, Verbeek H, van de Ven G, Ettema TP, et al. Structural validity and internal consistency of the Qualidem in people with severe dementia. *Int Psychogeriatr*. 2017; 1–11. <https://doi.org/10.1017/S1041610217001405> PMID: 28866990
58. Dichter MN, Schwab CGG, Meyer G, Bartholomeyczik S, Halek M. Item distribution, internal consistency and inter-rater reliability of the German version of the QUALIDEM for people with mild to severe and very severe dementia. *BMC Geriatr*. 2016; 16: 126. <https://doi.org/10.1186/s12877-016-0296-0> PMID: 27317476
59. Russell S, Dawkins M, de Wolf S, Bunnin A, Reid R, Murtagh F. Evaluation of the outcome assessment and complexity collaborative (oacc) train the trainers workshops. *BMJ Support Amp Palliat Care*. 2016; 6: A29–A30. <https://doi.org/10.1136/bmjspcare-2016-001245.81>