

LETTER TO THE EDITOR

doi:10.1017/S1041610223000741

Agreement between nursing home caregivers' observations of residents' depression, well-being, and quality of life

Dear Editor,

In nursing home (NH) residents, outcomes such as well-being and depression are often based on observable behaviors or signs reported by someone other than the resident. While previous studies have reported on the agreement between proxy-reported scores and self-reported scores (e.g. Leontjevas *et al.*, 2016), and between categories of proxy, such as relatives and professional caregivers (e.g. Robertson *et al.*, 2017), studies assessing agreement between professional caregivers acting as observers are scarce. Furthermore, limited attention has been paid to reporting agreement indices stratified by the level of residents' cognitive functioning. As professional caregivers commonly act as observers for resident outcomes, knowledge about the inter-rater reliability of observer-reported outcomes is important. Therefore, secondary analysis was performed on a dataset containing observer-reported outcomes in residents with and without dementia in Dutch and Flemish (Dutch speaking part of Belgium) NHs.

Eighty-one residents of 21 NHs were evaluated for depression (Nijmegen Observer-Rated Depression scale for detection of depression in nursing home residents [NORD]) (Leontjevas *et al.*, 2012), well-being (adapted version of the Social Well-being Of Nursing home residents scale [SWON-3]) (Gerritsen *et al.*, 2010), and quality of life (two subscales of the QUALIDEM, namely "social relations" and "having something to do") (Ettema *et al.*, 2007) by two professional caregivers (registered nurse or certified nurse assistant) who were involved in caring for the resident about whom the questions were answered. Most caregivers (46 out of 71) filled out the questionnaires for one resident (median, 1; range, 1–9). To assess the agreement between the pairs of caregivers, we calculated Gwet's AC1 or AC2 coefficients (Gwet, 2021) for individual items of the questionnaires using the irrCAC R package (Gwet, 2019). Individual item's agreement was calculated for more insight into whether individual items could be adjusted or deleted for improving the psychometric characteristics of an instrument. In addition, coefficients were calculated for subjective judgment regarding residents' depressive symptoms ("no," "yes, mild or

light," or "yes, severe") and for caregivers' knowledge of whether a depression diagnosis had been established ("yes," "no," or "don't know"). Intra-class correlation coefficients (ICC [1,1] and [1,2]) (Koo and Li, 2016) were calculated for the scale mean scores of the NORD (total scale), SWON-3 (three subscales and the total scale), and QUALIDEM (two subscales) using the irr R package (Gamer *et al.*, 2012).

For the total sample, Gwet's coefficients ranged from 0.29 to 0.63 for items of the NORD, from 0.32 to 0.75 for the items of the SWON-3, and from 0.45 to 0.74 for items of the QUALIDEM (see Table 1). Most items were characterized as "fair" or "moderate." Gwet's coefficients for the subjective judgment of residents' depressive symptoms and a depression diagnosis were 0.41 (fair) and 0.84 (good) respectively.

Although comparison of coefficients across different subsamples must be interpreted with caution due to relatively small sample sizes and, consequently, broad confidence intervals, the results point toward lower levels of agreement for observer-reported scores of residents with moderate to severe cognitive decline ($N = 18$, 14 of 24 analyzed items were characterized as "poor"), compared to residents with no to mild cognitive decline ($N = 58$, the most frequent item coefficients were characterized as "fair" [8 items] or "moderate" [9 items]).

Under the assumption of multiple raters, all (sub) scales showed at least moderate agreement (ICC [1,2] ≥ 0.50) for the total sample and for the subsample of residents with no to mild cognitive decline. For residents with moderate to severe cognitive decline, poor agreement (ICC [1,2] < 0.50) was found for all (sub)scales but the NORD and the subscale "social relations" of the QUALIDEM.

The limited agreement between caregivers concerning residents with moderate to severe cognitive decline underscores challenges for measurements in this population. One possible explanation is that interpretation of items or response options may be extra challenging when residents are less able to express themselves. Another explanation may be that accurate observations can be challenging if the symptoms of the outcome variables overlap with those of severe dementia (Leyhe *et al.*, 2017).

We believe that practitioners and researchers should be aware of these challenges when using and interpreting observer-reported outcomes for residents with dementia. Moreover, understanding why

Table 1. Agreement statistics of nursing home caregivers' observations of residents' depression, well-being, and quality of life

Resident outcomes	Weights	Item / SCALE	Total sample					Per level of cognitive decline										
			N	% Obs	Agreement diagonal	Gwet's AC1 or AC2 ²		ICC [95% CI]		N	GDS score of 1-3		GDS score of 4-6					
						Estimate [95% CI]	Altman's benchmark scale ⁷	ICC (1,1) ⁸	ICC (1,2) ⁹		Estimate [95% CI]	Altman's benchmark scale ⁷	ICC (1,1) ⁸	ICC (1,2) ⁹				
NORD ¹	Unweighted	1. Sadness	81	64.2	22 30	0.29 [0.08;0.50]	Poor		58	0.28 [0.03;0.54]	Poor	18	0.58 [0.17;0.99]	Fair				
		2. Crying	80	76.3	9 52	0.63 [0.46;0.80]	Moderate		57	0.64 [0.44;0.85]	Moderate	18	0.44 [-0.03;0.92]	Poor				
		3. Lack of response	81	72.8	10 49	0.56 [0.37;0.75]	Fair		58	0.63 [0.43;0.83]	Moderate	18	0.40 [-0.08;0.88]	Poor				
		4. Inactivity	81	71.6	34 24	0.44 [0.24;0.64]	Fair		58	0.52 [0.29;0.75]	Fair	18	0.20 [-0.34;0.74]	Poor				
		5. Eating and sleeping problems	79	65.8	16 36	0.36 [0.14;0.58]	Fair		56	0.41 [0.15;0.66]	Poor	18	0.15 [-0.37;0.67]	Poor				
		NA	81					0.36 [0.15;0.53]	0.53 [0.27;0.70]	58			0.33 [0.08;0.54]	0.50 [0.16;0.70]	18		0.49 [0.05;0.77]	0.66 [0.10;0.87]
		TOTAL	81							58			18					
		SWON-3 ²	Modified weights: Yes, often (most of them) 1.00 Yes, sometimes (in some) 0.75 Yes, often (most of them) 0.75 Yes, sometimes (in some) 1.00 No 0.00	1. (Affection)	81	85.5	36 13 1	0.75 [0.64;0.85]	Good		58	0.78 [0.66;0.90]	Good	18	0.70 [0.52;0.89]	Moderate		
				2. (Affection)	81	77.5	9 28 8	0.55 [0.40;0.71]	Moderate		58	0.55 [0.37;0.74]	Moderate	18	0.55 [0.22;0.87]	Fair		
				3. (Affection)	81	82.1	31 12 2	0.67 [0.54;0.79]	Moderate		58	0.68 [0.54;0.82]	Moderate	18	0.63 [0.33;0.94]	Moderate		
4. (Behavioral conformation)	81			75.0	15 15 3	0.50 [0.35;0.64]	Fair		58	0.51 [0.35;0.68]	Fair	18	0.35 [-0.06;0.76]	Poor				
5. (Behavioral conformation)	81			70.1	14 17 7	0.36 [0.19;0.53]	Fair		58	0.33 [0.13;0.54]	Poor	18	0.41 [0.04;0.77]	Poor				
6. (Behavioral conformation)	79			76.3	25 13 0	0.55 [0.39;0.71]	Moderate		57	0.68 [0.54;0.82]	Moderate	17	0.08 [-0.41;0.56]	Poor				
7. (Status)	80			69.1	10 14 19	0.34 [0.17;0.51]	Poor		57	0.44 [0.24;0.65]	Fair	18	0.02 [-0.34;0.38]	Poor				
8. (Status)	80			68.8	8 17 13	0.32 [0.13;0.50]	Poor		57	0.39 [0.17;0.61]	Fair	18	0.14 [-0.26;0.54]	Poor				
9. (Status)	81			74.4	14 16 8	0.45 [0.29;0.62]	Fair		58	0.43 [0.23;0.64]	Fair	18	0.49 [0.13;0.85]	Fair				
NA	81							0.35 [0.15;0.53]	0.52 [0.26;0.69]	58			0.41 [0.18;0.60]	0.58 [0.30;0.71]	18		0.33 [-0.33;0.55]	0.23 [-0.01;0.71]
QUALIDEM ³	Linear weights	3. (Social relations)	81	83.1	14 15 24	0.67 [0.58;0.76]	Moderate		58	0.71 [0.61;0.82]	Good	18	0.59 [0.35;0.83]	Moderate				
		12. (Social relations)	79	84.0	0 10 36	0.74 [0.65;0.83]	Good		58	0.71 [0.61;0.82]	Good	16	0.83 [0.68;0.98]	Good				
		18. (Social relations)	81	79.0	8 14 9 7	0.52 [0.40;0.63]	Moderate		58	0.57 [0.44;0.71]	Moderate	18	0.36 [0.09;0.62]	Poor				
		25. (Social relations)	81	77.0	2 7 12 14	0.47 [0.35;0.60]	Fair		58	0.49 [0.34;0.65]	Fair	18	0.51 [0.29;0.73]	Fair				
		26. (Having something to do)	80	80.8	1 6 9 27	0.61 [0.49;0.73]	Moderate		57	0.67 [0.52;0.82]	Moderate	18	0.44 [0.16;0.71]	Fair				
		29. (Social relations)	81	77.4	5 12 6 13	0.47 [0.35;0.59]	Fair		58	0.53 [0.37;0.68]	Fair	18	0.32 [0.08;0.56]	Poor				
		34. (Social relations)	81	76.1	0 4 9 22	0.53 [0.41;0.66]	Moderate		58	0.58 [0.44;0.72]	Moderate	18	0.38 [0.05;0.71]	Poor				
		38. (Having something to do)	81	74.5	22 6 5 3	0.45 [0.31;0.59]	Fair		58	0.55 [0.39;0.71]	Moderate	18	0.02 [-0.21;0.26]	Poor				
		NA	81					0.60 [0.44;0.72]	0.75 [0.61;0.84]	58			0.69 [0.53;0.80]	0.82 [0.69;0.89]	18		0.35 [-0.12;0.69]	0.52 [-0.26;0.82]
		Subjective judgement of depression	Modified weights: No 1.00 Yes, mild or light 0.25 Yes, severe 0.75	Symptoms ⁴	80	68.4	22 21 0	0.41 [0.24;0.57]	Fair		58	0.52 [0.35;0.70]	Fair	17	-0.14 [-0.58;0.29]	Poor		
Diagnosis ⁵	46			89.1	6 35	0.84 [0.70;0.99]	Good		34	0.83 [0.64;1.00]	Good	9	0.85 [0.48;1.00]	Moderate				

Note: % Obs = percentage observed; CI = confidence interval; ICC = intraclass correlation coefficient; GDS = global deterioration scale; N = valid number of caregiver pairs.

1. NORD: Nijmegen Observer-Rated Depression scale for detection of depression in nursing home residents. Response options: "Yes," "No."
2. SWON-3: Social Well-being Of Nursing home residents scale.
3. QUALIDEM: Response options: "Never," "Rarely," "Sometimes," "Frequently."
4. Symptoms: 1-item ("Do you think this resident has depressive symptoms?").
5. Diagnosis: 1-item ("Has a depression diagnosis been established?"). Response options: "Yes," "No," "Don't know" (the option "don't know" was treated as missing).
6. Gwet's AC1 was used for calculating unweighted coefficients, and Gwet's AC2 was used for calculating weighted coefficients.
7. Altman's benchmarking 5-point scale ranging from "poor" to "very good" was used to interpret the magnitude of the AC1 and AC2 coefficients. A cumulative probability of above 0.95 was applied to determine the lowest expected agreement level.
8. ICC (1,1): Intraclass correlation coefficient for absolute agreement, 1-way random effects model, single rater. This is informative for planning measurements from a single rater.
9. ICC (1,2): Intraclass correlation coefficient for absolute agreement, 1-way random effects model, two raters. This is informative for the use of a mean value of two raters as the basis of the actual measurement (Koo and Li, 2016).

different raters reach different conclusions regarding the same resident is important for interpreting observer-reported outcomes. We therefore argue that future research should explore the reasons why observer-reported scores may differ between caregivers, especially concerning residents with moderate or severe dementia. Both cognitive interviewing and other forms of in-depth interviews with caregivers are recommended to better understand their interpretation of items and to discover the actual reasons for differences between caregivers' scores.

Conflict of interest







None.

Source of funding

None.

References

- Ettema, T. P., Dröes, R. M., De Lange, J., Mellenbergh, G. J. and Ribbe, M. W.** (2007). QUALIDEM: development and evaluation of a dementia specific quality of life instrument-validation. *International Journal of Geriatric Psychiatry*, 22, 424–430.
- Gamer, M., Lemon, J., Fellows, I. and Singh, P.** (2012). *irr: Various Coefficients of Interrater Reliability and Agreement* (version 0.84.1) [software, R package]. Available from <https://cran.r-project.org/web/packages/irr/irr.pdf>.
- Gerritsen, D. L., Steverink, N., Frijters, D. H., Ooms, M. E. and Ribbe, M. W.** (2010). Social well-being and its measurement in the nursing home, the SWON-scale. *Journal of Clinical Nursing*, 19, 1243–1251.
- Gwet, K.** (2021). *Handbook of Inter-Rater Reliability: the Definitive Guide to Measuring the Extent of Agreement among Raters: Vol 1*. Gaithersburg, USA: Analysis of Categorical Ratings, AgreeStat Analytics.
- Gwet, K. L.** (2019). *irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC)* (version 1.0) [software, R package]. Available from <https://cran.r-project.org/web/packages/irrCAC/irrCAC.pdf>.
- Koo, T. K. and Li, M. Y.** (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163.
- Leontjevas, R., Gerritsen, D. L., Vernooij-Dassen, M. J., Teerenstra, S., Smalbrugge, M. and Koopmans, R. T.** (2012). Nijmegen observer-rated depression scale for detection of depression in nursing home residents. *International Journal of Geriatric Psychiatry*, 27, 1036–1044.
- Leontjevas, R., Teerenstra, S., Smalbrugge, M., Koopmans, R. T. and Gerritsen, D. L.** (2016). Quality of life assessments in nursing homes revealed a tendency of proxies to moderate patients' self-reports. *Journal of Clinical Epidemiology*, 80, 123–133.
- Leyhe, T. et al.** (2017). A common challenge in older adults: classification, overlap, and therapy of depression and dementia. *Alzheimers Dementia*, 13, 59–71.
- Robertson, S., Cooper, C., Hoe, J., Hamilton, O., Stringer, A. and Livingston, G.** (2017). Proxy rated quality of life of care home residents with dementia: a systematic review. *International Psychogeriatrics*, 29, 569–581.

INGE KNIPPENBERG,^{1,2}  RUSLAN LEONTJEVAS,^{1,2} 
 INE DECLERCQ,^{1,2,3}  JACQUES VAN LANKVELD,² 
 PATRICIA DE VRIENDT,^{3,4}  AND DEBBY GERRITSEN¹ 

¹Department of Primary and Community Care, Radboud University Medical Center, Radboud Institute for Medical Innovation, Radboudumc Alzheimer Center, Nijmegen, Netherlands

²Faculty of Psychology, Open University of the Netherlands, Heerlen, Netherlands

³Department of Gerontology and Frailty in Ageing (FRIA) Research Group, Mental Health and Wellbeing (MENT) Research Group, Vrije Universiteit Brussel, Brussels, Belgium

⁴Research Group Health and Care, Artevelde University of Applied Sciences, Ghent, Belgium

Email: inge.knippenberg@radboudumc.nl