Ageing Research Reviews 81 (2022) 101742

Contents lists available at ScienceDirect

Ageing Research Reviews

journal homepage: www.elsevier.com/locate/arr

Review

Psychometric characteristics of comprehensive geriatric assessments (CGAs) for long-term care facilities and community care: A systematic review

Mauricio Molinari-Ulate [a,b,*], Aysan Mahmoudi [a,b], Manuel A. Franco-Martín [a,c], Henriëtte G. van der Roest [d]

[a] Psycho-Sciences Research Group, Institute of Biomedical Research of Salamanca, University of Salamanca, Spain
[b] Department of Research and Development, Iberian Institute of Research in Psycho-Sciences, INTRAS Foundation, Zamora, Spain
[c] Psychiatric and Mental Health Department, Zamora Healthcare Complex, Zamora, Spain
[d] Department on Aging, Netherlands Institute of Mental Health and Addiction (Trimbos Institute), Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Long-term care
Community care
Comprehensive geriatric assessment
CGA
Reliability
Validity

ABSTRACT

Background: Comprehensive Geriatric Assessments (CGAs) have been incorporated as an integrated care approach effective to face the challenges associated to uncoordinated care, risk of hospitalization, unmet needs, and care planning experienced in older adult care. As they assessed different dimensions, is important to inform about the content and psychometric properties to guide the decisions when selecting and implementing them in practice. This systematic review provides a comprehensive insight on the strengths and weaknesses of the CGAs used in long-term care settings and community care.

Methods: A systematic search was conducted in PubMed, CINAHL, and Web of Science Core Collection. Studies published up to July 13, 2021, were considered. Quality appraisal was performed for the included studies.

Results: A total of 10 different CGAs were identified from 71 studies included. Three instruments were reported for long-term care settings, and seven for community care. The content was not homogenous and differed in terms of the detail and clearness of the areas being evaluated. Evidence for good to excellent validity and reliability was reported for various instruments.

Conclusions: Setting more specific and clear domains, associated to the special needs of the care setting, could improve informed decisions at the time of selecting and implementing a CGA. Considering the amount and quality of the evidence, the instrument development trajectory, the validation in different languages, and availability in different care settings, we recommend the interRAI LTCF and interRAI HC to be used for long-term facilities and community care.

1. Background

Societies and health systems are being challenged by the current demographic transition to ageing populations, arising the need for adjustments and responses from all sectors, including public health (WHO, 2015, 2018). The complexity of older adult care associated to comorbidities, polypharmacy, multiple treatments and interventions from different health care providers, socioeconomic status, and the risk of developing functional and cognitive impairment, have implications on the quality of life and capacity for independence and autonomy of the older adult population (Bernabei et al., 2008; World Health Organization, 2015). For facing this panorama, health care professionals and public health policymakers must pursue the development of healthcare approaches that place older people's needs and preferences in the centre of service delivery (WHO, 2015).

Integrated care has been considered as an effective alternative approach to the traditional and standard service delivery, improving the quality of older adult care and positively impacting rates of institutionalization and costs (Johri et al., 2003; McDonald et al., 2013; World Health Organization, 2015). A method in which different care levels and

* Correspondence to: Ctra. de la Hiniesta 137, 49024 Zamora, Spain.
E-mail addresses: mmolinari@ides.es (M. Molinari-Ulate), amahmoudi@ides.es (A. Mahmoudi), mfranco@usal.es (M.A. Franco-Martín), hroest@trimbos.nl (H.G. van der Roest).

https://doi.org/10.1016/j.arr.2022.101742
Received 9 May 2022; Received in revised form 27 September 2022; Accepted 27 September 2022
Available online 29 September 2022

services are integrated across health care and long-term care settings covering the needs and preferences of older adults along their life course (World Health Organization, 2015). It comprises three key features, a case-management system that assess the individual's needs according to a person-centred perspective, a comprehensive care plan which aim to assist people on their treatment and care decisions, and an effective transfer of information among caregivers and settings aiming to improve coordination and integration of care (World Health Organization, 2015).

Following this approach, Comprehensive Geriatric Assessments (CGAs) became important interventions in geriatric care (Ellis et al., 2011; Pilotto et al., 2017). CGAs incorporate the main pillars of an integrated care approach and are being defined as a process of care including a coordinated multidimensional and multidisciplinary assessment, facilitating the clinical decision for the formulation of a personalized care plan to address the needs and concerns of the older person (and their family and carers) (Pilotto et al., 2017; British Geriatrics Society, 2019).

The interdisciplinary and integrated care process approach, centres its attention on the person and relatives, leading to a holistic evaluation of core domains. As this care process approach considers multiple areas of an individual, care professionals and policy makers must be aware of CGA's psychometric flaws and fortes to be able to take reliable decisions on care planning and health policy outcomes, aiming to optimize care quality. For this reason, the aim of this systematic review is to provide insight into the content and psychometric characteristics of CGAs used in long-term care settings and community care.

2. Methods

2.1. Literature search

A systematic literature search was conducted in three databases, PubMed, CINAHL and Web of Science Core Collection for studies up to July 13, 2021 (search strategy and studies selection procedure are available in Methods A.1). See Table 1 for inclusion and exclusion criteria. Reference lists of selected studies and relevant systematic reviews were scanned for potentially eligible primary studies.

2.2. Data extraction

One author (MMU) extracted the data from the final selection of papers. The following information was gathered: a) name of CGA; b) authors/year; d) description of sample; e) country; f) study setting; g) study design; h) aim of the study; i) type of validity/reliability; and j) main findings. Scale, items, indices, or domains were also extracted from the relevant studies.

Results on the domains covered by the CGAs were reported, followed by results on the evaluation of the reliability/validity of 1) complete CGAs; 2) specific domains and items; 3) scales and indexes; and 4) outcomes relevant for organization of care and clinical decision making. The domains of the CGAs were obtained from the description of the areas assessed in the papers, and from the forms or questionnaires, when available. To avoid bias on the domains' description, those areas related to demographic or administrative data (e.g., Identification, Background or Assessment information) were excluded from this analysis, as they might not be reported in the papers but included in the forms.

Table 1
Inclusion and exclusion criteria used for the study selection.

Inclusion Criteria	Exclusion Criteria
The Comprehensive Geriatric Assessment instrument must be one single test or assessment tool	An assessment that consists of a collection of single domain measures, tests or assessments, or stand-alone instruments assessing one domain (e.g, depression)
The study should report on the validation or reliability of the instrument	Studies published in languages other than English or Spanish
The instrument must target specifically people of 55 years and older	Publications such as conference abstracts, case studies, protocols, dissertations, books and systematic reviews (however, references from selected SRs were checked)
	If the entire instrument is self-report.
	Instruments developed for acute care, mental health care, palliative care, primary care or hospitalized settings. Also, those instruments that assessed transfer from or to any of the aforementioned care setting.

2.3. Risk of bias

The quality of the studies was assessed independently by two authors (MM and AM) using the "STANDARD QUALITY ASSESSMENT CRITERIA" for quantitative studies (Kmet et al., 2004) and percentage of agreement was calculated. Subsequently, the same two raters discussed the disagreements and came with a final consent agreement. Further detailed regarding the risk of bias methodology is provided in Methods A.1.

2.4. Interpretation of test scores

The validity and reliability outcomes were obtained according to the aims of the studies and the primary outcomes identified. When outcomes were unclear or multiple outcomes were reported, the researcher selected the one that best reflected the main result of the study (e.g., main scales outcomes rather than subscales). To avoid differences on the interpretation of the psychometric outcomes between the studies, the reviewers decided to use a standardized criterion based on the literature (available in Methods A.1).

3. Results

3.1. Literature search

After duplicates were removed, the titles and abstracts of 1226 records were screened, of which 115 records were analysed in detail by full text. Finally, 47 records were included in the review. After scanning the reference lists of selected studies, 24 additional studies were considered as eligible primary studies. In total, 71 papers were included in the final results (Fig. 1. PRISMA flow diagram).

3.2. Risk of bias

Two raters (MM and AM) scored the quality of the 71 papers, obtaining a 78.97% of agreement. The scores ranged from 0.50 to 1 (1 maximum score), with an average score of 0.83 (Table A.1).

3.3. CGA characteristics

Ten different CGAs were identified: three focused on long-term care settings and seven focused on community care (see Table 2). Table 3 shows the number of domains assessed by each CGA and the number of studies that used criterion measures and the number of criterion measures used to validate some of their domains. The areas validated and the criterion measures used for their validation are available in Table A.2.

3.4. Psychometric characteristics of the CGAs

The following section describes the reliability and validity results. The first subsection describes the results of the long-term care facilities instruments, followed by the community care. Each subsection starts with the reliability results followed by validity findings. Results are

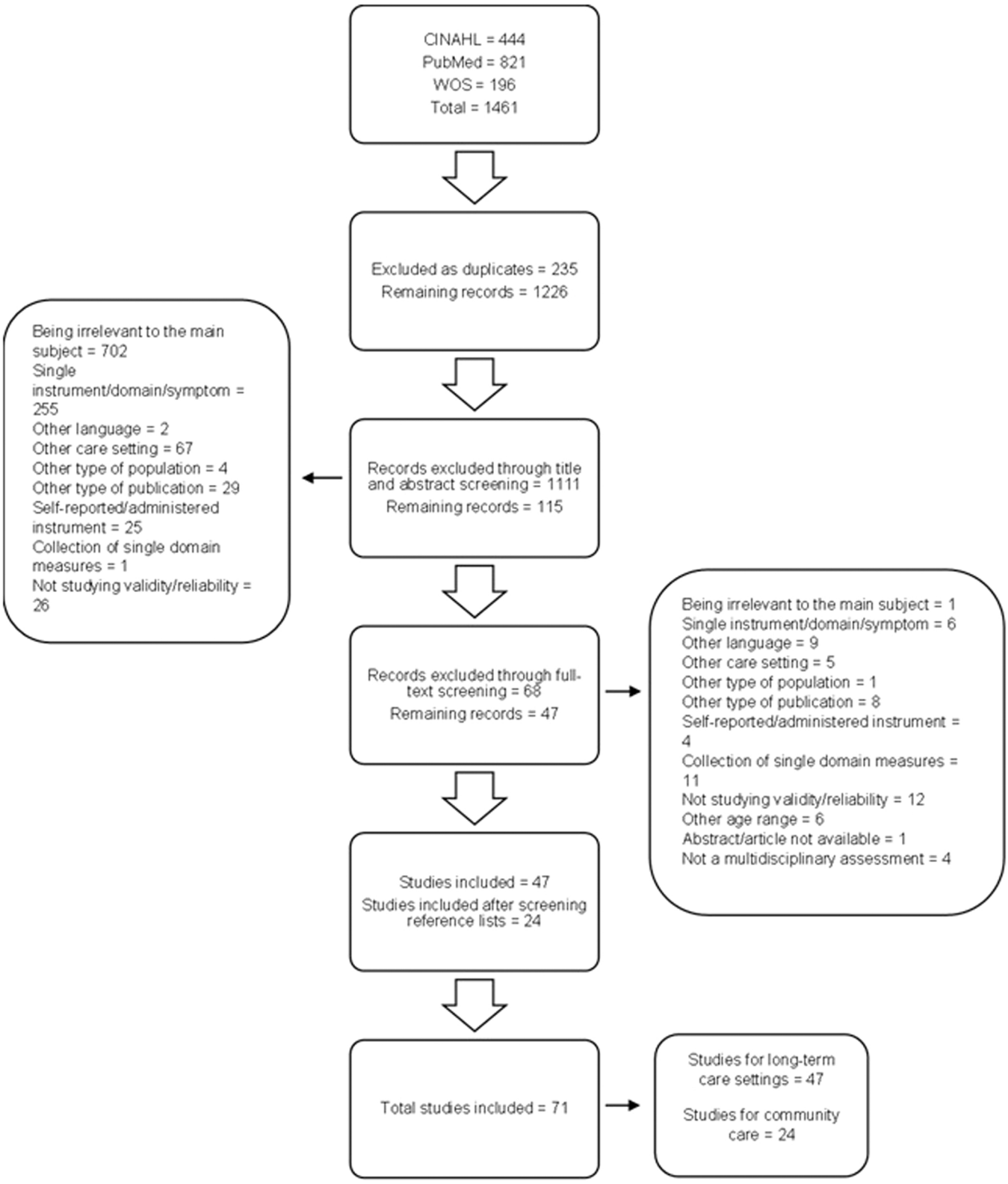

Fig. 1. PRISMA flow diagram.

described according to the four approaches identified to study the psychometric characteristics: a) complete CGAs; b) specific domains or items; c) specific scales or indices; and d) outcomes relevant for clinical decision making and organization of care. The psychometric evidence for outcomes relevant for organization of care and clinical decision making are described in a separate subsection.

3.5. CGAs for long-term care settings

3.5.1. Reliability results

The reliability of the complete CGA RAI-MDS and subsequent updated versions (n = 8), and CPAT (n = 1), was studied (Table A.3). Good to excellent inter-rater reliability was reported for more than 91.7% of the

Table 2
CGAs identified in the review, according to their care setting.

Long-term care setting	Community care
Resident Assessment Instrument-Minimum Data Set (RAI-MDS) (Morris et al., 1990) [a]RAI-MDS 2.0 (Morris et al., 1997b) [a]RAI-MDS 2.1 Chinese version (Chou et al., 2001) [ab]MDS 3.0 (Saliba and Buchanan, 2012) [a]interRAI Long-term Care Facilities (LTCF) v.9.1 (Hirdes et al., 2008) [a]interRAI LTCF Korean Version (Kim et al., 2015)	Comprehensive Assessment and Referral Evaluation (CARE) (Gurland et al., 1977)
VALutazione GRAFica (ValGraf) Residential version (Gigantesco et al., 1995)	Older American's Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire (OMFAQ) (Fillenbaum and Smyer, 1981) [a]OARS-OMFAQ Spanish version (Fibla et al., 1996)
Care Planning Assessment Tool (CPAT) (Fleming, 2008) [a]J-CPAT Japanese version (Kanegae et al., 2010)	Philadelphia Geriatric Center Multilevel Assessment Instrument (MAI) (Lawton et al., 1982)
	Popovich Scale (Grubba et al., 1990)
	Outcome and Assessment Instrument Set (OASIS) (Shaughnessy et al., 1994)
	RAI-MDS Home Care (HC) (Morris et al., 1997a) [a]RAI-MDS HC Chinese version (Kwan et al., 2000) [a]RAI-MDS HC Swiss version (Ludwig and Busnel, 2017)
	[a]interRAI HC (Hirdes et al., 2008) [a]interRAI HC Korean version (Kim et al., 2015)
	Community Assessment of Risk Instrument (CARI) (Clarnette et al., 2015)

[a] Are considered as subsequent versions of the original instrument.
[b] The MDS 3.0 is an updated version of RAI-MDS and RAI-MDS 2.0, however, it was not developed by the interRAI network.

Table 3
Number of domains included in the CGAs, number of studies using criterion measures, and number of criterion measures used for validation.

Care setting	Name of CGA	# of domains assessed by the CGA	# of studies that included criterion measures	# of criterion measures or gold standards used for validation
Long-term Care	RAI-MDS	15	14	37
	RAI-MDS 2.0	16	7	11
	MDS 3.0	16	2	2
	ValGraf	8	1	2
	CPAT	8	2	4
	interRAI LTCF	17	0	0
Community Care	CARE	3	0	0
	OARS-OMFAQ	5	1	4
	MAI	7	0	0
	Popovich Scale	4	1	2
	OASIS	4	2	5
	RAI-MDS HC	16	5	7
	interRAI HC	18	3	5
	CARI	3	0	0

Note: RAI = Resident Assessment Instrument; MDS = Minimum Data Set; ValGraf = Valutazione Grafica; CPAT = Care Planning Assessment Tool; LTCF = Long Term Care Facilities; CARE = Comprehensive Assessment and Referral Evaluation; OARS-OMFAQ = Older American's Resources and Services Multidimensional Functional Assessment Questionnaire; MAI = Philadelphia Geriatric Center Multilevel Assessment Instrument; OASIS = Outcome and Assessment Instrument Set; HC = Home Care; CARI = Community Assessment of Risk Instrument.

items for the CPAT (Fleming, 2008).

The RAI-MDS inter-rater reliability improved with every updated version. Originally, reliability was reported to be fair to moderate for more than half of the items (Hawes et al., 1995; Morris et al., 1990), and for the core set of items of the RAI-MDS in several countries (Sgadari et al., 1997). For the revised items and 83% of the new items in version 2.0, good to excellent inter-rater reliability was reported (Morris et al., 1997b); for version 3.0, good to excellent inter-rater reliability was found for the majority of the new and modified items (Saliba and Buchanan, 2012). The latest version, the interRAI LTCF and its Korean version (Hirdes et al., 2008; Kim et al., 2015), reported good to excellent inter-rater reliability for more than 90% of the items, and an international study reported excellent average test-retest reliability and good average inter-rater reliability (Onder et al., 2012).

The reliability of domains was reported for the RAI-MDS (n = 1), and the J-CPAT (n = 1), and *for the items* of RAI-MDS and its subsequent version 2.0 (n = 3) (Table A.4). For the 16 studied domains of the RAI-MDS, inter-rater reliability was moderate for 11, good for one, and poor for four (Hawes et al., 1995). The inter-rater and test-retest reliability for the eight J-CPAT domains were reported to be in the adequate range (Kanegae et al., 2010).

For specific RAI-MDS items related to urinary incontinence, moderate to good inter-rater reliability was reported, except for one item that was reported as poor; excellent inter-rater reliability was found for identifying different gradations of incontinence (Resnick et al., 1996). Oral/dental items reported poor or none inter-rater reliability between nursing staff and dental assessment, except for one item reporting good inter-rater reliability (Jockusch et al., 2021). Excellent test-retest reliability was found for all RAI-MDS 2.0 pain items (Fisher et al., 2002).

The reliability of scales from the RAI-MDS and subsequent versions (n = 8), and the CPAT (n = 1), *and for the indices* derived from the RAI-MDS (n = 2), was reported (Table A.5). From the eight scales of the CPAT, internal consistency was adequate for all, except for the Psychiatric Symptom Scale (Fleming, 2008).

Internal consistency was in the acceptable range for the following RAI-MDS and subsequent versions 2.0 and Korean interRAI LTCF scales: CPS, MDS-Cognition Scale (MDS-COGS), Challenging Behaviour Profile Scale (CBP) (except for one subscale), Communication, Pain and ADL (Gerritsen et al., 2008; Gruber-Baldini et al., 2000; Kim et al., 2015; Mor et al., 2011). The scales for Social Engagement, Mood, Behaviour, and Pain Scale did not reach acceptable levels for internal consistency (Mor et al., 2011). Internal consistency of the depression scales were reported as follow: insufficient for the DRS in the RAI-MDS 2.0, but acceptable in the RAI-MDS (Anderson et al., 2003; Hsiao et al., 2015), acceptable for the Depression Scale in the Korean interRAI LTCF (Kim et al., 2015), and for the composite depression measure (sum of all items in section E1, E1SUM) in version 2.0 (Koehler et al., 2005).

Inter-rater reliability was reported as excellent for all eight scales from the CPAT (Fleming, 2008). To screen dementia, the cut-off point 2 or higher of the MDS-COGS of the RAI-MDS showed high specificity, but lower sensitivity than cut-off point 1. For the cut-off point of 2, interrater and test-retest reliability for a negative screen was found to be high, but moderate for a positive screen (Zimmerman et al., 2007). For the CBP from the RAI-MDS 2.0, overall inter-rater reliability was reported as fair (Gerritsen et al., 2008). Poor to moderate test-retest reliability was shown for the DRS of RAI-MDS 2.0 (Anderson et al., 2003).

Only the reliability of indices from the RAI-MDS was studied, reporting good to excellent inter-rater reliability for seven indices (Casten et al., 1998), and an acceptable internal consistency for the Social Engagement Index (Mor et al., 1995).

3.5.2. Validity results

The validity of the complete CGA was only reported for the construct validity of the ValGraf, indicating that 52.9% of the variance was explained by 13 factors (Pascazio et al., 2009) (Table A.3).

The validity of domains was assessed for the RAI-MDS (n = 2), the CPAT (n = 1), and the ValGraf (n = 1), and *for the items* from the RAI-MDS and subsequent updated versions (n = 10) (Table A.4). Three studies of RAI-MDS and version 2.0 included samples with individuals with the characteristics of the conditions that were being validated, two of them determining the condition through previous medical diagnoses (Hendrix et al., 2003; Liang et al., 2011). Even though this was not specifically reported for the CPAT, the study sample included people from dementia and rehabilitation care settings (Kanegae et al., 2010). For the ValGraf this was not described.

The criterion validity for the RAI-MDS Behaviour and ADL domains reported medium to large effect sizes, also poor to fair agreement between RAI-MDS ADL assessments and the criterion measure (Lum et al., 2005; Snowden et al., 1999). For the CPAT Japanese version, the domains of Confusion, Self-help, Physical Problems, and Care dependency reported large effect sizes when compared against criterion scales (Kanegae et al., 2010). The concurrent validity for the Functional and Cognitive domains of the ValGraf reported large effect sizes (Pascazio et al., 2009).

Results from the validation of items from the RAI-MDS and subsequent updated versions reported difficulties on identifying their specific conditions. Risk of undernutrition and mood anxiety symptoms items from the RAI-MDS were found to underreport these conditions (Liang et al., 2011; Simmons et al., 2002). The same results have been reported for the mood indicators items from the RAI-MDS 2.0 (Hendrix et al., 2003). However, improvements in the validity of items on cognition, mood, behaviour and depression in version 3.0, and higher agreement with criterion measures as compared to the RAI-MDS 2.0 items, was reported (Saliba and Buchanan, 2012). Validity problems have been also reported for the oral/dental (Hoben et al., 2016; Jockusch et al., 2021), fall (Hill-Westmoreland and Gruber-Baldini, 2005), and pain items (Fisher et al., 2002) of the RAI-MDS 2.0.

Diagnostic accuracy for falls was in the acceptable range for only one of the two fall items (Hill-Westmoreland and Gruber-Baldini, 2005). Similar difficulties were described for the urinary tract infection (UTI) items (Stevenson et al., 2004).

Test content validity for the oral health section of the interRAI LTCF and interRAI HC was considered incomplete and items not clearly worded. Four items were considered as relevant, and two items were considered as feasible (Krausch-Hofmann et al., 2019).

The validity of scales was assessed for the CPAT (n = 1), and for the RAI-MDS and subsequent versions (n = 15), and *for the indices* from the RAI-MDS (n = 4) (Table A.5). The CPAT sample was selected from specific dementia care units, while for nine studies of the RAI-MDS and version 2.0 samples included individuals with the conditions determined before the study, four of them established by medical diagnoses (Anderson et al., 2003; Liang et al., 2011; Morris et al., 1994; Zimmerman et al., 2007).

The CPAT scales validated against criterion measures were Confusion and Self-help, which reported large effect sizes (Fleming, 2008). In the case of RAI-MDS and subsequent versions, all Cognition Scales (MDS-COGS, CPS, and Brief Interview for Mental Status), the Pain Scale, and two subscales of the CBP reported large effect sizes and good/-excellent agreement against criterion measures (Fries et al., 2001; Gerritsen et al., 2008; Gruber-Baldini et al., 2000; Hartmaier et al., 1994, 1995; Morris et al., 1994; Saliba et al., 2012). However, a medium effect size between the CPS and the Mini-Mental State Examination (MMSE) (Snowden et al., 1999), and a fair agreement between the CPS and the 7-item Global Deterioration Scale were described (Hartmaier et al., 1994). Small to medium effect sizes were found for the divergent validity from the MDS-COGS and the CPS (Gruber-Baldini et al., 2000).

The criterion validity for the DRS reported inconsistencies. One study reported a large correlation against the Geriatric Depression Scale-Short From (GDS-SF) (Hsiao et al., 2015), however, another study found some contradictory results using the same criterion measure (Liang et al., 2011). Also, the DRS did not correlate with standard measures or small to medium effect sizes as shown in two studies (Anderson et al., 2003),(Koehler et al., 2005).

Construct validity for the DRS and the CPS reported acceptable factor loadings, and the CPS also met the criteria for simplicity and face validity (Hsiao et al., 2015),(Morris et al., 1994). For the CBP, construct validity reported that it is formed by four subscales, and it was also identified as strong predictor of one year mortality (Gerritsen et al., 2008; Hsiao et al., 2015; Mor et al., 2011).

Regarding the diagnostic accuracy, AUC was excellent for CPS, MDS-COGS, and the Brief Interview for Mental Status (BIMS) (Hartmaier et al., 1994, 1995; Saliba et al., 2012), with the exception of one study that reported an insufficient AUC for the MDS-COGS (Zimmerman et al., 2007). The AUC for the DRS was also insufficient and for the Fracture Risk Scale not good (Hsiao et al., 2015; Ioannidis et al., 2017). These results are related to the Sensitivity and Specificity values reported for the different scales, which described high values for all Cognition Scales (Hartmaier et al., 1994, 1995; Morris et al., 1994; Saliba et al., 2012; Zimmerman et al., 2007), but inconsistencies and low values for the DRS that did not reach the expected minimum (Anderson et al., 2003; Hsiao et al., 2015). Also, it was identified that the DRS under detects the prevalence of depression compared to the GDS-SF (Hsiao et al., 2015).

For the RAI-MDS indices, large effect sizes were demonstrated for Cognition, ADL, Time Use and Social Engagement (Lawton et al., 1998; Mor et al., 1995), while medium and small correlations for Problem Behaviours and Depression when compared against criterion measures (Lawton et al., 1998). Discriminant validity was also demonstrated for all these indices (Lawton et al., 1998). Construct validity for Social engagement, Cognition, ADL and Time Use reported acceptable factor loadings, however, most factors failed to be replicated within a cognitive impaired sample (Casten et al., 1998; Mor et al., 1995). Lastly, Depression and Cognition were associated with the psychiatric diagnosis of depression and strongly with a dementia diagnosis, respectively (Lawton et al., 1998). Four models of mortality risk indices were validated, reporting that the Flacker models were more reliable predictors and better discriminators of mortality risk than the MDS-Mortality Risk indices (Kruse et al., 2010).

3.6. CGAs for community care

3.6.1. Reliability results

The reliability of the complete CGA was studied for the OASIS (n = 3), RAI-MDS HC and subsequent interRAI HC version (n = 5), and CARI (n = 1) (Table A.3). Two studies reported good to excellent inter-rater reliability for the OASIS (Hittle et al., 2004; Madigan and Fortinsky, 2004), nonetheless, in another study was reported poor to moderate inter-rater reliability for more than 60% of its items (Kinatukara et al., 2005).

Good to excellent inter-rater reliability has been reported for the RAI-MDS HC and interRAI HC in several countries (Hirdes et al., 2008; Kim et al., 2015; Kwan et al., 2000; Morris et al., 1997a). Data reliability reported inaccurate records in demographic and height or weight variables of the interRAI HC (Schluter et al., 2016).

Lastly, the majority of the items included in the CARI reported poor to fair inter-rater reliability (Clarnette et al., 2015).

The reliability of domains of the CARE, OASIS and CARI was targeted by three studies (Table A.4). For CARE domains, poor to excellent reliability was found, with the Psychiatric Dimension performing the best of the three domains (Gurland et al., 1977). OASIS domains internal consistency was within the acceptable range only for the functional domain (Madigan and Fortinsky, 2000), while intra-rater reliability was good to excellent for the affect and behavioural domains, and fair to good for the clinical and functional domains (Madigan and Fortinsky, 2000).

Inter-rater reliability of the CARI was poor to fair for their three domains and low agreement was reported for risk outcomes (Clarnette et al., 2015).

The reliability of scales or indices from the RAI-MDS HC and interRAI HC (n = 5), Popovich Scale (n = 1), OARS-OMFAQ (n = 1), and MAI (n = 1), were reported (Table A.5).

Internal consistency was studied for eight scales of the RAI-MDS HC Chinese version and six scales of the interRAI HC Korean version. Of the tested scales, the Pain, Communication, and Instrumental Activities of Daily Living (IADL) involvement and capacity Scales of both CGAs, and the Activities of Daily Living (ADL) and Depression Scales of the interRAI HC Korean version, reported internal consistency reliability withing acceptable ranges (Kim et al., 2015; Kwan et al., 2000; Leung et al., 2011, 2012). For all the scales of the Popovich Scale, inter-rater reliability reported large effect sizes (Grubba et al., 1990), while for the OARS-OMFAQ, it was good for the majority of the scales except for the Physical Health scale (Fillenbaum and Smyer, 1981).

Regarding the reliability of the indices, the Frailty Index of the RAI-MDS HC (Mor et al., 1995) and all indices of the MAI reported acceptable internal consistency (Lawton et al., 1982; Ludwig and Busnel, 2017)). Large effect sizes for test-retest reliability and moderate to good inter-rater reliability was also described for all scales of the MAI (Lawton et al., 1982).

3.6.2. Validity results

The validity of the complete CGA was reported only for the convergent validity of the OASIS (n = 1), indicating several inconsistencies between the OASIS and the criterion measure (Kinatukara et al., 2005) (Table A.3).

The validity of domains was studied for the OASIS (n = 2), and for the mortality data of the interRAI HC (n = 1), *and for the items* of the oral health section of the interRAI HC (n = 1) (Table A.4). When compared against gold standards, functional items of the OASIS reported medium to large effect sizes for the composite scores of ADL and IADL, large effect sizes were described for the cognitive functioning items, and moderate for the depressive symptoms items (Tullai-McGuinness et al., 2009). Regarding construct validity, only the functional domain reported adequate performance (Madigan and Fortinsky, 2000).

The criterion validity of the mortality data of the interRAI HC was reported to be consistent (Schluter et al., 2016). The test content validity of the oral health section was considered as incomplete and items not clearly worded; only four items were considered as relevant, and only two items were considered as feasible (Krausch-Hofmann et al., 2019).

The validity of scales was studied for the RAI-MDS HC and interRAI HC subsequent version (n = 7), OARS-OMFAQ (n = 2), and Popovich Scale (n = 1), and *for the indices* from the RAI-MDS HC (n = 3), and the MAI (n = 1) (Table A.5). The two studies validating the interRAI HC scales included samples with individuals diagnosed by the condition being assessed (Gee et al., 2021; Penny et al., 2016); for the other CGAs, this condition was not identified in their studies.

The criterion validity of the ADL, IADL, and Cognition Scales of the RAI-MDS HC reported large effect sizes (Carpenter et al., 2005; Landi et al., 2000). However, for the RAI-MDS HC Mood scale and the interRAI HC Depression Rating Scale (DRS), no correlation and a medium effect size were reported against the criterion measure, respectively (Carpenter et al., 2005; Penny et al., 2016). For the Economic, Mental Health, Physical Health, and Self-care capacity Scales of the OARS-OMFAQ and the Cognitive, Physical health and Social resources Subscales of the Popovich Scale, large effect sizes were reported (Fillenbaum and Smyer, 1981; Grubba et al., 1990).

Regarding content validity, RAI-MDS HC was reported to have higher data completion and better domain coverage compared to current used assessment instruments (Carpenter et al., 2005). Concerning construct validity, a good factorial structure was reported for the IADL involvement and capacity Scale and Negative Mood Scale of the RAI-MDS HC Chinese version (Leung et al., 2011, 2012). The OARS-OMFAQ Spanish version revealed a similar factor structure to the original English version, with some differences in the Self-care and Mental Health Scales (Fibla et al., 1996). For the Popovich Scale, the Economic Subscale construct validity was demonstrated (Grubba et al., 1990).

The diagnostic accuracy revealed excellent and inadequate Areas Under the Curve (AUCs) for the Cognitive Performance Scale (CPS) and the DRS of the interRAI HC, respectively (Gee et al., 2021; Penny et al., 2016). The sum between sensitivity and specificity reached the minimum expected value for the CPS but not for the DRS (Gee et al., 2021; Penny et al., 2016). In the case of the Changes in Health, End-stage disease, and Signs and Symptoms scale (CHESS), results showed it to be a strong predictor for hospitalization (Campitelli et al., 2016).

The Frailty Index of the RAI-MDS HC, was found as a strong predictor of mortality risk, admission to a long-term care facility, hospitalizations, and falls (Burn et al., 2018; Campitelli et al., 2016; Ludwig and Busnel, 2017). The studies on the seven indices from the MAI described large effect sizes for internal validity and concurrent validity, while small to large effect sizes for criterion validity for all indices with exception for the Cognition and Perceived Environment indices (Lawton et al., 1982).

3.7. Outcomes relevant for clinical decision making and organization of care

The RAI-MDS and subsequent version 2.0, interRAI-LTCF, and RAI-MDS HC and subsequent version also generate outcomes relevant for clinical decision making and organization of care. None of the other identified CGAs validated similar outcomes.

For *clinical decision making*, the Resident Assessment Protocols (RAPs) from the RAI-MDS reported reliable detection of two different urodynamic diagnosis (Resnick et al., 1996). In its Chinese version, four RAPs, Cognitive loss, ADL, Communication and Mood Symptoms, reported good concurrent validity and inter-rater and test-retest reliability (Chou et al., 2001). For the new version of the RAPs, the Clinical Assessment Protocols (CAPs) from the RAI-MDS HC, four of the 30 CAPs reported good to excellent validity, the remaining CAPs reported slight and fair agreements (Kwan et al., 2000).

RAI-MDS triggers for the detection of undernutrition were validated (Beck et al., 2001). They reported those triggers to be able to identify relevant characteristics of this condition in comparisons with those participants with no triggers (Beck et al., 2001).

For *organization of care*, the Resource Utilization Groups (RUG-III) from the RAI-MDS reported to be able to differentiate between residents according to intensity of care needs, validating RUGs with the level of Registered General Nurses care time (Carpenter et al., 2003). The Finish version of the RUG-III for version 2.0, explained 38.2% of the variance of total patient cost per diem; also, they found an ambiguity in terms of the inter-rater reliability as it varied from fair to good according to the type of professionals, agreement was slightly better when assessors were personal nurses, as compared to personal nurse and outsider evaluator (Björkgren et al., 1999). For the interRAI HC version, good convergent validity with the Resource Utilization in Dementia instrument (RUD-Lite) (Wimo et al., 1998) for the estimations of societal costs of care (including healthcare and nursing, welfare and informal care) derived from interRAI-HC data was reported (van Lier et al., 2016).

Applicable for quality improvement and benchmarking purposes, the 22 Quality Indicators (QIs) from the RAI-MDS 2.0 reported fair to excellent inter-rater reliability, except for two (Mor et al., 2003). Except for three of the 100 single items used for measuring the quality indicators, good to excellent inter-rater reliability was reported for all (Mor et al., 2003). In addition, the content validity of the QIs reported pressure ulcers as the most practice sensitive QI; variations were found between the QIs considered as most sensitive for physician and nursing care, while none of them were considered to be most sensitive for policy/decision makers (Estabrooks et al., 2013).

Lastly, the accuracy of the RAI-MDS in identifying hospitalization

events and payment sources was evaluated, reporting that the RAI-MDS is not ideal for the identification of these elements without the support of supplemental information from claims data (Cai et al., 2011).

4. Discussion

This systematic review aimed to provide insight into the content and psychometric characteristics of CGAs to enable potential users to make an informed decision when a CGA is considered to be implemented.

We found that some of the CGAs included broad domains which makes it difficult to identify what they are specifically assessing. On the contrary, other assessments were more specific or clearer by referring and evaluating the domains in more detail. We consider it necessary to be more specific on the domains as these are screening tools that could identify risk factors for potential deterioration and take into account the complexity associated with older adult care, which is associated with comorbidity, polypharmacy, multiple treatments, etc., and could guide clinical decision making and interventions (Bernabei et al., 2008; World Health Organization, 2015; Scanlan, 2005), so if the information is clear, and the dimensions and warnings are well defined, the relevant clinical information will be gathered by the clinicians. This approach also allows to compare how comprehensive a CGA is against proposed lists of domains that are considered relevant to be evaluated by a CGA (Pilotto et al., 2017; British Geriatrics Society, 2019) and could facilitate the clinicians and policy makers decision on which instruments are more suitable for their specific situation and aim.

By acknowledging the comprehensiveness of the CGAs, a big amount of data could be gathered supporting improvements and developments of algorithms that could be used to incorporate automatic learning to extract and identify useful information through large databases (Dipnall et al., 2016). Consequently, this could guide the development of decision models for medical and care procedures, such as prognosis, diagnosis, and treatment planning, which should be embedded into CGAs as systematic support components (Góngora et al., 2018), optimizing personalized treatments and improving evidence-based decisions making among clinicians and scientists, identifying the causes of unmet care of older adults and more effective treatment approaches (Góngora et al., 2018).

The most studied instruments for long term care were the RAI-MDS and its subsequent versions, and for community care were the RAI-MDS HC and subsequent version, and OASIS. The RAI-MDS and RAI-MDS HC and their subsequent versions have been studied, validated, and adapted globally. Only the CPAT and OARS-OMFAQ were adapted for other non-English speaking countries, but not reaching the global covering of the interRAI network instruments. Another distinguishing feature was that interRAI instruments used specific population samples clinical characteristics (e.g., depression, cognition, dementia) necessarily to validate the outcomes of interest.

Although the number of studies and psychometric results positively support the RAI-MDS and subsequent versions for its used in long-term care facilities as compared to the other identified CGAs, some considerations must be made. Its reliability improved along the evolution to the next versions, reporting good to excellent results in the later stages. Overall, the main strengths rely on the psychometric characteristics of the Cognition, ADL, Time Use, and CBP items, and the outcomes relevant for clinical decision making and organization of care (e.g., RAPs and RUG III); however, flaws remain on items such as oral/dental problems, risk of undernutrition, urinary tract infection, and depression and mood.

For community care, the difference on the number of studies validating the CGAs was not as vast as for long-term care settings, and the psychometric characteristics appear to achieve high standards for all, except for the CARI and the CARE. However, some considerations must be contemplated: a) The CARE, the CARI, the MAI and the OARS-OMFAQ were not reported to be validated against gold standards or criterion measures; b) the CARE, the CARI, the Popovich Scale, the OASIS, and the OARS-OMFAQ were studied with suboptimal study designs, using small samples; c) items of the scales of the Popovich Scale and OASIS were validated against OARS-OMFAQ scales as the criterion measure, for which strong psychometric evidence is not available; d) inconsistency was revealed on the validation and reliability of the OASIS, some studies reported high values but this was not confirmed by other studies; and e) the RAI-MDS HC and subsequent version reported similar flaws to the aforementioned for the RAI-MDS and subsequent versions on areas such as mood and depression and the oral health section.

In this review mainly studies performed in high income countries have been included. Studies on CGAs performed in low to middle income countries, such as African, Southeast Asian, and Latin-American countries have not been found, despite Spanish language was one of the inclusion criteria. This might imply that in these regions, where health systems are not optimally developed, an integrated care approach towards ageing has not been implemented yet, despite that these regions are also confronted with an ageing population (Prina et al., 2019).

4.1. Strengths and limitations

The strength of this review relies on reporting on a comprehensive panorama of the CGAs available for long-term and community care, with insights on the psychometric properties, the content, and the variety of instruments. The results presented here support reliable decision-making on care planning and health policy outcomes, impacting the vulnerable older adult population that live in these care settings.

Several limitations regarding this systematic review must be considered. First, our definition of a CGA encompassed single multidisciplinary assessment instruments targeting different domains, not a conglomerate of single domain instruments assessing each area individually or those developed for being used for one single discipline. Therefore, a number of CGAs used in practice might not be included in our results. This definition of a CGA might be closer to the integrated care approach proposed at the introduction (Johri et al., 2003; McDonald et al., 2013; World Health Organization, 2015), which focusses on the facilitation of data sharing across care settings, offering a core set of items avoiding different score systems, and thereby facilitating an easy transfer between care settings. Second, several statistical methods and strategies were used to assess the psychometric characteristics, however we proposed a single criterion to report, compare and analyse these results. Also, we only used specificity, sensitivity, and AUC regarding diagnostic accuracy. Third, CGAs in other languages than English and Spanish were not considered. Finally, some of the study designs and types of validity and reliability presented at the tables are according to the researchers' judgements as they were underreported by some of the studies.

5. Conclusion

The timely detection of clinical problems, side effects or comorbidity is strategic for a good quality care, so, it is highly relevant to considered reliable tools with clear and specific domains to support clinical decisions.

Due to the study characteristics such as the sample size, number of studies, instrument development trajectory, validation in several countries, and availability in different care settings, we recommend the interRAI LTCF and interRAI HC to be used for long-term care facilities and community care, respectively. Also, it has been demonstrated their potential for predicting mortality, hospitalizations, admission, urinary infections, and detecting cognitive problems, falls, and nutritional risk factors. Nonetheless, health care professionals must be aware of the flaws reported for mood, depression, oral health, risk of undernutrition, and urinary tract infection.

Funding

This work was supported by the Marie Skłodowska Curie Actions Innovative Training Network H2020-MSCA-ITN [grant agreement number 813196].

CRediT authorship contribution statement

MMU, HR, and MF conceptualized and designed the review. MMU and AM screened the titles and abstracts of the records resulted from the search. MMU and AM evaluated the records for risk of bias. MMU extracted the data from the included studies, and it was verified by HR. MMU, HR, and MF drafted and critically revised the manuscript. HR and MF provided overall study supervision. All authors authorized the final version of the manuscript that was submitted. All authors read and approved the final manuscript.

Conflict of interest

MF and HR are fellows of interRAI Network (www.interRAI.org). HR is a co-author in one of the studies included in this review, however, she was not involved in the risk of bias assessment, which was performed independently by MMU and AM.

Data availability

No data was used for the research described in the article.

Acknowledgement

Not applicable.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.arr.2022.101742.

References

Anderson, R.L., Buckwalter, K.C., Buchanan, R.J., Maas, M.L., Imhof, S.L., 2003. Validity and reliability of the minimum data set depression rating scale (MDSDRS) for older adults in nursing homes. Age Ageing 32, 435–438.
Beck, A.M., Ovesen, L., Schroll, M., 2001. Validation of the Resident Assessment Instrument triggers in the detection of under-nutrition. Age Ageing 30, 161–165.
Bernabei, R., Landi, F., Onder, G., Liperoti, R., Gambassi, G., 2008. Second and third generation assessment instruments: the birth of standardization in geriatric care. J. Gerontol. 63A, 308–313.
Björkgren, M.A., Häkkinen, U., Finne-Soveri, U.H., Fries, B.E., 1999. Validity and reliability of Resource Utilization Groups (RUG-III) in Finnish long-term care facilities. Scan. J. Public Health 27, 228–234.
British Geriatrics Society, 2019. Comprehensive Geriatric Assessment Toolkit for Primary Care Practitioners.
Burn, R., Hubbard, R.E., Scrase, R.J., Abey-Nesbit, R.K., Peel, N.M., Schluter, P.J., Jamieson, H.A., 2018. A frailty index derived from a standardized comprehensive geriatric assessment predicts mortality and aged residential care admission. BMC Geriatr. 18, 319.
Cai, S., Mukamel, D.B., Veazie, P., Temkin-Greener, H., 2011. Validation of the minimum data set in identifying hospitalization events and payment source. J. Am. Med. Dir. Assoc. 12, 38–43.
Campitelli, M.A., Bronskill, S.E., Hogan, D.B., Diong, C., Amuah, J.E., Gill, S., Seitz, D., Thavorn, K., Wodchis, W.P., Maxwell, C.J., 2016. The prevalence and health consequences of frailty in a population-based older home care cohort: a comparison of different measures. BMC Geriatr. 16, 133.
Carpenter, I., Challis, D.J., Swift, C., 2005. "Single" assessment for older people: comparison of the MDS-HC with current auditable methods in the home care setting. J. Integr. Care 13, 35–41.
Carpenter, I., Perry, M., Challis, D., Hope, K., 2003. Identification of registered nursing care of residents in English nursing homes using the Minimum Data Set Resident Assessment Instrument (MDS/RAI) and Resource Utilisation Groups version III (RUG-III). Age Ageing 32, 279–285.
Casten, R., Lawton, M.P., Parmelee, P.A., Kleban, M.H., 1998. Psychometric characteristics of the minimum data set i: confirmatory factor analysis. J. Am. Geriatr. Soc. 46, 726–735.
Chou, K.-L., Chi, I., Leung, A.C.-t, Wu, Y.M., Liu, C.-P., 2001. Validation of minimum data set for nursing home in Hong Kong Chinese elders. Clin. Gerontol. 23, 43–54.
Clarnette, R.M., Ryan, J.P., O'Herlihy, E., Svendrovski, A., Cornally, N., O'Caoimh, R., Leahy-Warren, P., Paul, C., Molloy, D.W., 2015. The community assessment of risk instrument: investigation of inter-rater reliability of an instrument measuring risk of adverse outcomes. J. frailty Aging 4, 80–89.
Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N., Meyer, D., 2016. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. PLOS ONE 11, e0148195.
Ellis, G., Whitehead, M.A., Robinson, D., O'Neill, D., Langhorne, P., 2011. Comprehensive geriatric assessment for older adults admitted to hospital: meta-analysis of randomised controlled trials. BMJ 343, d6553.
Estabrooks, C.A., Knopp-Sihota, J.A., Norton, P.G., 2013. Practice sensitive quality indicators in RAI-MDS 2.0 nursing home data. BMC Res. Notes 6.
Fibla, G.G., Patiño, P.E., Domínguez, A.C., 1996. Versión española del OARS multidimensional functional assessment questionnaire: adaptación transcultural y medida de la validez. Aten. Prima 17.
Fillenbaum, G.G., Smyer, M.A., 1981. The development, validity, and reliability of the OARS multidimensional functional assessment qustionnaire. J. Gerontol. 36, 428–434.
Fisher, S.E., Burgio, L.D., Thorn, B.E., Allen-Burge, R., Gerstle, J., Roth, D.L., Allen, S.J., 2002. Pain assessment and management in cognitively impaired nursing home residents: association of certified nursing assistant pain report, Minimum Data Set pain report, and analgesic medication use. J. Am. Geriatr. Soc. 50, 152–156.
Fleming, R., 2008. The reliability and validity of the care planning assessment tool. Australas. J. Ageing 27, 209–211.
Fries, B.E., Simon, S.E., Morris, J.N., Flodstrom, C., Bookstein, F.L., 2001. Pain in U.S. nursing homes: validating a pain scale for the minimum data set. Gerontologist 41, 173–179.
Gee, S., Croucher, M., Cheung, G., 2021. Performance of the cognitive performance scale of the resident assessment instrument (interRAI) for detecting dementia amongst older adults in the community. Int. J. Environ. Res. Public Health 18.
Gerritsen, D.L., Achterberg, W.P., Steverink, N., Pot, A.M., Frijters, D.H., Ribbe, M.W., 2008. The MDS challenging behavior profile for long-term care. Aging Ment. Health 12, 116–123.
Gigantesco, A., Morosini, P., Alunno Pergentini, S., Minardi, A., Cavagnaro, P., 1995. Validazione di un semplice strumento per la valutazione funzionale dell'anziano: Il ValGraf. . Giorn. Gerontol. 43, 379–385.
Góngora, S., de la Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D.C., Nozaleda, L.M., Franco, M., 2018. Data mining algorithms and techniques in mental health: a systematic review. J. Med. Syst. 42, 161.
Grubba, C.J., Popovich, B., Jirovec, M.M., 1990. Reliability and validity of the popovich scale in home health care assessments. Appl. Nurs. Res. 3, 161–165.
Gruber-Baldini, A.L., Zimmerman, S.I., Mortimore, E., Magaziner, J., 2000. The validity of the minimum data set in measuring the cognitive impairment of persons admitted to nursing homes. J. Am. Geriatr. Soc. 48, 1601–1606.
Gurland, B., Kuriansky, J., Sharpe, L., Simon, R., Stiller, P., Birkett, P., 1977. The comprehensive assessment and referral evaluation (CARE)–rationale, development and reliability. Int J. Aging Hum. Dev. 8, 9–42.
Hartmaier, S.L., Sloane, P.D., Guess, H.A., Koch, G.G., 1994. The MDS cognition scale: a valid instrument for identifying and staging nursing home residents with dementia using the minimum data set. J. Am. Geriatr. Soc. 42, 1173–1179.
Hartmaier, S.L., Sloane, P.D., Guess, H.A., Koch, G.G., Mitchell, M., Phillips, C.D., 1995. Validation of the minimum data set cognitive performance scale: agreement with the mini-mental state examination. J. Gerontol.: Med. Sci. 50A, M128–M133.
Hawes, C., Morris, J.N., Phillips, C.D., Mor, V., Fries, B.E., Nonemaker, S., 1995. Reliability estimates for the minimum data set for nursing home resident assessment and care screening (MDS). Gerontologist 35, 172–178.
Hendrix, C.C., Sakauye, K.M., Karabatsos, G., Daigle, D., 2003. The use of the minimum data set to identify depression in the elderly. J. Am. Med. Dir. Assoc. 4, 308–312.
Hill-Westmoreland, E.E., Gruber-Baldini, A.L., 2005. Falls documentation in nursing homes: agreement between the minimum data set and chart abstractions of medical and nursing documentation. J. Am. Geriatr. Soc. 53, 268–273.
Hirdes, J.P., Ljunggren, G., Morris, J.N., Frijters, D.H., Finne Soveri, H., Gray, L., Bjorkgren, M., Gilgen, R., 2008. Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. BMC Health Serv. Res 8, 277.
Hittle, D.F., Shaughnessy, P.W., Crisler, K.S., Powell, M.C., Richard, A.A., Conway, K.S., Stearns, P.M., Engle, K., 2004. A study of reliability and burden of home health assessment using OASIS. Home Health Care Serv. Q 22, 43–63.
Hoben, M., Poss, J.W., Norton, P.G., Estabrooks, C.A., 2016. Oral/dental items in the resident assessment instrument - minimum Data Set 2.0 lack validity: results of a retrospective, longitudinal validation study. Popul Health Metr. 14, 36.
Hsiao, C.Y., Lan, C.F., Chang, P.L., Li, I.C., 2015. Development of the psychometric property of a Minimum Data-Set-Based Depression Rating Scale for use in long-term care facilities in Taiwan. Aging Ment. Health 19, 129–135.
Ioannidis, G., Jantzi, M., Bucek, J., Adachi, J.D., Giangregorio, L., Hirdes, J., Pickard, L., Papaioannou, A., 2017. Development and validation of the Fracture Risk Scale (FRS) that predicts fracture over a 1-year time period in institutionalised frail older people living in Canada: an electronic record-linked longitudinal cohort study. BMJ Open 7, e016477.
Jockusch, J., Hopfenmüller, W., Sobotta, B.A.J., Nitschke, I., 2021. Interrater reliability and concurrent validity of oral/dental items in the resident assessment instrument minimum data set 2.0. Gerodontology 38, 66–81.
Johri, M., Beland, F., Bergman, H., 2003. International experiments in integrated care for the elderly: a synthesis of the evidence. Int J. Geriatr. Psychiatry 18, 222–235.

Kanegae, S., Koizumi, S., Fleming, R., Ichimaru, N., Nagashio, T., 2010. Development of a Japanese version of the care planning assessment tool. Australas. J. Ageing 29, 27–32.
Kim, H., Jung, Y.-I., Sung, M., Lee, J.-Y., Yoon, J.-Y., Yoon, J.-L., 2015. Reliability of the interRAI Long Term Care Facilities (LTCF) and interRAI Home Care (HC). Geriatr. Gerontol. Int. 15, 220–228.
Kinatukara, S., Rosati, R.J., Huang, L., 2005. Assessment of OASIS reliability and validity using several methodological approaches. Home Health Care Serv. Q 24, 23–38.
Kmet, L.M., Lee, R.C., Cook, L.S., 2004. STANDARD QUALITY ASSESSMENT CRITERIA for Evaluating Primary Research Papers from a Variety of Fields. Alberta Heritage Foundation for Medical Research, Alberta, Canada.
Koehler, M., Rabinowitz, T., Hirdes, J., Stones, M., Carpenter, G.I., Fries, B.E., Morris, J. N., Jones, R.N., 2005. Measuring depression in nursing home residents with the MDS and GDS: an observational psychometric study. BMC Geriatr. 5, 1.
Krausch-Hofmann, S., De Almeida Mello, J., Declerck, D., Declercq, A., De Lepeleire, J., Tran, T.D., Lesaffre, E., Duyck, J., 2019. The oral health-related section of the interRAI: Evaluation of test content validity by expert rating and assessment of potential reasons for inaccurate assessments based on focus group discussions with caregivers. Gerodontology 36, 382–394.
Kruse, R.L., Oliver, Parker, Mehr, D., Petroski, D.R., Swenson, G.F., D.L., Zweig, S.C., 2010. Using mortality risk scores for long-term prognosis of nursing home residents: caution is recommended. J. Gerontol. Ser. A, Biol. Sci. Med. Sci. 65, 1235–1241.
Kwan, C.-W., Chi, I., Lam, T.-P., Lam, K.-F., Chou, K.-L., 2000. Validation of minimum data set for home care assessment instrument (MDS-HC) for Hong Kong Chinese elders. Clin. Gerontol. 21, 35–48.
Landi, F., Tua, E., Onder, G., Carrara, B., Sgadari, A., Rinaldi, C., Gambassi, G., Lattanzio, F., Bernabei, R., 2000. Minimum data set for home care a valid instrument to assess frail older people living in the community. Med. Care 38, 1184–1190.
Lawton, M.P., Moss, M., Fulcomer, M., Kleban, M.H., 1982. A research and service oriented multilevel assessment instrument. J. Gerontol. 37, 91–99.
Lawton, M.P., Casten, R., Parmelee, P.A., Van Haitsma, K., Corn, J., Kleban, M.H., 1998. Psychometric characteristics of the minimum data set 11: validity. J. Am. Geriatr. Soc. 46, 736–744.
Leung, D.Y., Leung, A.Y., Chi, I., 2011. An evaluation of the factor structure of the instrumental activities of daily living involvement and capacity scales of the minimum data set for home care for elderly Chinese community dwellers in Hong Kong. Home Health Care Serv. Q. 30, 147–159.
Leung, D.Y., Leung, A.Y., Chi, I., 2012. A psychometric evaluation of a negative mood scale in the MDS-HC using a large sample of community-dwelling Hong Kong Chinese older adults. Age Ageing 41, 317–322.
Liang, C.K., Chen, L.K., Tsai, C.F., Su, T.P., Lo, Y.K., Lan, C.F., Hwang, S.J., 2011. Screening depression among institutionalized older Chinese men by minimum data set: we need a new instrument. J. Geriatr. Psychiatry Neurol. 24, 179–183.
van Lier, L.I., van der Roest, H.G., van Hout, H.P., van Eenoo, L., Declercq, A., Garms-Homolova, V., Onder, G., Finne-Soveri, H., Jonsson, P.V., Hertogh, C.M., Bosmans, J. E., 2016. Convergent validity of the interRAI-HC for societal costs estimates in comparison with the RUD Lite instrument in community dwelling older adults. BMC Health Serv. Res 16, 440.
Ludwig, C., Busnel, C., 2017. Derivation of a frailty index from the resident assessment instrument - home care adapted for Switzerland: a study based on retrospective data analysis. BMC Geriatr. 17, 1–10.
Lum, T.Y., Lin, W., Kane, R.L., 2005. Use of proxy respondents and accuracy of minimum data set assessments of activities of daily living. J. Gerontol.: Med. Sci. 60A, 654–659.
Madigan, E.A., Fortinsky, R.H., 2000. Additional psychometric evaluation of the Outcomes and Assessment Information Set (OASIS). Home Health Care Serv. Q 18, 49–62.
Madigan, E.A., Fortinsky, R.H., 2004. Interrater reliability of the outcomes and assessment information set: results from the field. Gerontologist 44, 689–692.
McDonald, K.M., Schultz, E.M., Chang, C., 2013. Evaluating the state of quality-improvement science through evidence synthesis: insights from the closing the quality gap series. Perm. J. 17, 52–61.
Mor, V., Intrator, O., Unruh, M.A., Cai, S., 2011. Temporal and Geographic variation in the validity and internal consistency of the Nursing Home Resident Assessment Minimum Data Set 2.0. BMC Health Serv. Res 11, 78.
Mor, V., Angelelli, J., Jones, R., Roy, J., Moore, T., Morris, J., 2003. Inter-rater reliability of nursing home quality indicators in the U.S. BMC Health Serv. Res. 3, 20.
Mor, V., Branco, K., Fleishman, J., Hawes, C., Phillips, C., Morris, J., Fries, B., 1995. The structure of social engagement among nursing home residents. J. Gerontol. B Psychol. Sci. Soc. Sci. 50, P1–P8.
Morris, J.N., Fries, B.E., Mehr, D.R., Hawes, C., Phillips, C.D., Mor, V., Lipsitz, L.A., 1994. MDS cognitive performance scale. J. Gerontol.: Med. Sci. 49, M174–M182.
Morris, J.N., Nonemaker, S., Murphy, K., Hawes, C., Fries, B.E., Mor, V., Phillips, C.D., 1997b. A commitment to change: revision of HCFA's RAI. J. Am. Geriatr. Soc. 45, 1011–1016.
Morris, J.N., Hawes, C., Fries, B.E., Phillips, C.D., Mor, V., Katz, S., Murphy, K., Drugovich, M.L., Friedlob, A.S., 1990. Designing the national resident assessment instrument for nursing homes. Gerontol. Soc. Am. 30, 293–307.
Morris, J.N., Fries, B.E., Steel, K., Ikegami, N., Bernabei, R., Carpenter, I., Gilgen, R., Hirdes, J.P., Topinková, E., 1997a. Comprehensive clinical assessment in community setting: applicability of the MDS-HC. J. Am. Geriatr. Soc. 45, 1017–1024.
Onder, G., Carpenter, I., Finne-Soveri, U.H., Gindin, J., Frijters, D., Henrard, J.C., Nikolaus, T., Topinková, E., Tosato, M., Liperoti, R., Landi, F., Bernabei, R., 2012. Assessment of nursing home residents in Europe: the Services and Health for Elderly in Long TERm care (SHELTER) study. BMC Health Serv. Res. 12.
Pascazio, L., Morosini, P., Bembich, S., Nardone, I., Clarici, A., Barbina, L., Zuttion, R., Gigantesco, A., 2009. Description and validation of a geriatric multidimensional graphical instrument for promoting longitudinal evaluation. Arch. Gerontol. Geriatr. 48, 317–324.
Penny, K., Barron, A., Higgins, A.M., Gee, S., Croucher, M., Cheung, G., 2016. Convergent validity, concurrent validity, and diagnostic accuracy of the interRAI depression rating scale. J. Geriatr. Psychiatry Neurol. 29, 361–368.
Pilotto, A., Cella, A., Pilotto, A., Daragjati, J., Veronese, N., Musacchio, C., Mello, A.M., Logroscino, G., Padovani, A., Prete, C., Panza, F., 2017. Three decades of comprehensive geriatric assessment: evidence coming from different healthcare settings and specific clinical conditions. J. Am. Med Dir. Assoc. 18 (192), e191–e192 e111.
Prina, A.M., Mayston, R., Wu, Y.T., Prince, M., 2019. A review of the 10/66 dementia research group. Soc. Psychiatry Psychiatr. Epidemiol. 54, 1–10.
Resnick, N.M., Brandeis, G.H., Baumann, M.M., Morris, J.N., 1996. Evaluating a national assessment strategy for urinary incontinence in nursing home residents: reliability of the minimum data set and validity of the resident assessment protocol. Neurourol. Urodyn. 15, 583–598.
Saliba, D., Buchanan, J., 2012. Making the investment count: revision of the minimum data set for nursing homes, MDS 3.0. J. Am. Med. Dir. Assoc. 13, 602–610.
Saliba, D., Buchanan, J., Edelen, M.O., Streim, J., Ouslander, J., Berlowitz, D., Chodosh, J., 2012. MDS 3.0: brief interview for mental status. J. Am. Med. Dir. Assoc. 13, 611–617.
Scanlan, B.C., 2005. The value of comprehensive geriatric assessment. Care Manag. J. 6, 2–8.
Schluter, P.J., Ahuriri-Driscoll, A., Anderson, T.J., Beere, P., Brown, J., Dalrymple-Alford, J., David, T., Davidson, A., Gillon, D.A., Hirdes, J., Keeling, S., Kingham, S., Lacey, C., Menclova, A.K., Millar, N., Mor, V., Jamieson, H.A., 2016. Comprehensive clinical assessment of home-based older persons within New Zealand: an epidemiological profile of a national cross-section. Aust. N. Z. J. Public Health 40, 349–355.
Sgadari, A., Morris, J.N., Fries, B.E., Ljunggren, G., Jónsson, P.V., DuPaquier, J., Schroll, M., 1997. Efforts to establish the reliability of the resident assessment instrument. Age Ageing 26, 27–30.
Shaughnessy, P.W., Crisler, K.S., Schlenker, R.E., Arnold, A.G., Kramer, A.M., Powel, M. C., Hittle, D.F., 1994. Measuring and assuring the quality of home health care. Health Care Financ. Rev. 16, 35–67.
Simmons, S.F., Lim, B., Schnelle, J.F., 2002. Accuracy of minimum data set in identifying residents at risk for undernutrition: oral intake and food complaints. J. Am. Med. Dir. Assoc. 3, 140–145.
Snowden, M., McCormick, W., Russo, J., Srebnik, D., Comtois, K., Bowen, J., Teri, L., Larson, E.B., 1999. Validity and responsiveness of the Minimum Data Set. J. Am. Geriatr. Soc. 47, 1000–1004.
Stevenson, K.B., Moore, J.W., Sleeper, B., 2004. Validity of the minimum data set in identifying urinary tract infections in residents of long-term care facilities. J. Am. Geriatr. Soc. 52, 707–711.
Tullai-McGuinness, S., Madigan, E.A., Fortinsky, R.H., 2009. Validity testing the Outcomes and Assessment Information Set (OASIS). Home Health Care Serv. Q 28, 45–57.
Wimo, A., Wetterholm, A.L., Mastery, V., Winblad, E., 1998. Evaluation of the resource utilization and caregiver time in Anti-dementia drug trials – a quantitative battery, in: Wimo, A., Karlsson, G., Jönsson, B., Winblad, B. (Eds.), The Health Economics of Dementia. Wiley's, London, UK.
World Health Organization, 2015. World Report on Ageing and Health. World Health Organization.
World Health Organization, 2018. Integrated care for older people: realigning primary health care to respond to population ageing. World Health Organization.
Zimmerman, S., Sloane, P.D., Williams, C.S., Dobbs, D., Ellajosyula, R., Braaten, A., Rupnow, M.F., Kaufer, D.I., 2007. Residential care/assisted living staff may detect undiagnosed dementia using the minimum data set cognition scale. J Am Geriatr Soc 55, 1349–1355.