BMC Geriatrics

RESEARCH Open Access

Mapping mood-improving behaviors in nursing homes: development and evaluation of the Actions to Improve Mood inventories

Inge Knippenberg[1,2*], Ruslan Leontjevas[1,2], Ine Declercq[1,2,3], Patricia De Vriendt[3,4], Anke Persoon[1], Peter Verboon[2], Jacques van Lankveld[2] and Debby Gerritsen[1]

Abstract

Background Nursing home residents and professional caregivers may engage in behaviors that, although not considered formal treatments for depression, can improve residents' mood. To support future studies aiming to explore ways to complement formal depression care in nursing homes, reliable and valid instruments for measuring these informal mood-improving behaviors are needed.

Methods This project developed and evaluated inventories to measure mood-improving behaviors in Dutch and Belgian nursing homes. Study 1 followed an iterative mixed-methods approach to develop two inventories: the Actions to Improve Mood by Residents (AIM-R) and the Actions to Improve Mood by Caregivers (AIM-C), and to assess their content validity ($N=31$ residents; $N=35$ caregivers, respectively). Study 2 evaluated test-retest agreement ($N=206$; $N=125$) and inter-rater agreement (AIM-C: $N=81$) using a test-retest design. Study 3 explored the inventories' practical application through semi-structured interviews ($N=12$; $N=6$). Data were analyzed through thematic analysis, content validity indices, and Gwet's AC2 agreement coefficients.

Results Both inventories demonstrated acceptable content validity, with moderate to very good test-retest agreement. Inter-rater agreement for most AIM-C items was classified as "fair," and appeared lower when caregivers completed the inventories for residents with moderate to severe cognitive decline compared to residents with no to mild cognitive decline. Thematic analysis suggested that using the inventories increased awareness of mood-improving behaviors and contributed to better knowledge about residents. Challenges regarding usability and interpretability were identified, along with suggestions for refinement.

Conclusions The inventories appear to adequately capture mood-improving behaviors and show consistency over time. Nevertheless, their usability and interpretability could benefit from further refinement. Pending additional research, these inventories hold promise for assessing mood-improving behaviors in nursing homes, aiding future efforts to explore new ways to enhance depression care alongside traditional treatments.

*Correspondence:
Inge Knippenberg
inge.knippenberg@radboudumc.nl

Full list of author information is available at the end of the article

Keywords AIM-R, AIM-C, Instrument development, Long-term care, Psychometric evaluation

Mapping mood-improving behaviors in nursing homes: development and evaluation of the actions to improve mood inventories

Depressive symptoms are common among nursing home residents, affecting approximately 46% of those with dementia and around 23% of those without dementia [1]. Among residents without dementia, the prevalence of major depressive disorder is estimated to be about 19% [2]. Since depression is associated with reduced quality of life [3] and increased mortality [4, 5], effective approaches for preventing and treating depression in nursing home residents are important.

Depression interventions can be broadly categorized into pharmacological and non-pharmacological treatment approaches. Despite their common use, pharmacological interventions have shown limited effects on nursing home residents [6–8]. A systematic review and meta-analysis by Nelson and Devanand [9] also indicated insufficient efficacy of antidepressants in people with dementia. Additionally, antidepressants are often associated with poor treatment adherence and negative side effects [10, 11].

In contrast, non-pharmacological treatments, including psychotherapeutic (e.g., cognitive behavioral therapy and reminiscence therapy) and psychosocial interventions (e.g., exercise programs and interventions targeting social isolation), have demonstrated effectiveness in reducing depressive symptoms, especially those that incorporate active behavioral components, with fewer negative side effects compared to pharmacological approaches [7, 8, 12–16]. However, older adults with limited cognitive capacity might encounter difficulties with psychotherapeutic interventions [17, 18], indicating a need for additional approaches.

One promising approach involves everyday behaviors by residents and caregivers that, while not considered part of formal treatment, can improve residents' mood. For example, professional caregivers using humor or residents initiating conversations have been shown to contribute to improved mood [19–21]. Because these behaviors seem relatively easy to integrate into daily routines, they warrant further exploration. Gaining insight into the use and impact of these behaviors within nursing homes is important for enhancing resident care. This requires the development of measurement instruments with acceptable psychometric properties to accurately assess these behaviors, guiding future research and aiding in the development or refinement of interventions.

These instruments need to comprehensively assess behaviors that may improve the mood of nursing home residents and are easy for residents or professional staff to engage in. This expands upon existing instruments, such as the Pleasant Events Schedule [22, 23], which focus solely on residents' engagement in pleasant activities. Moreover, instruments for mood-improving behaviors must accommodate the diversity of the resident population, considering varying degrees of physical and cognitive ability and different preferences regarding such behaviors. For example, some residents may prefer not to engage in certain behaviors themselves or may be unable to do so, but they might benefit from caregiver involvement during interactions. From a practical point of view, the instruments should also facilitate relevant suggestions for improvements in daily care. To the best of our knowledge, no such instruments currently exist.

Therefore, the aims of this project were (1) to develop inventories for mapping mood-improving behaviors in nursing home residents meeting the aforementioned requirements; (2) to examine the content validity, test-retest and inter-rater agreement of these instruments; and (3) to explore experiences with applying the inventories in daily practice in terms of potential issues with psychometrics and usability. The development of such inventories could facilitate future studies to identify specific mood-improving behaviors that are most beneficial for residents, guiding personalized and more effective interventions that can be used to improve residents' mood in addition to formal treatment approaches.

General methods

In this section, we describe the methodological aspects that apply to all three studies within this project. Further details about the methods (i.e., procedure, materials, and analysis) as well as the results and subsequent considerations are described per study.

Construct and measurement model

In this project, we define "mood-improving behaviors" as a broad spectrum of actions aimed at enhancing the mood of nursing home residents. These behaviors encompass a range of actions exhibited by both residents themselves and professional caregivers. A formative measurement model was chosen for this construct, as this type of measurement model is suitable when the construct is considered to be formed by its indicators [24]. Formative models allow for a flexible approach, as they can represent constructs that are conceptually defined as a combination of diverse, potentially unrelated indicators.

Study design and setting

We conducted three consecutive studies in nursing homes in both the Netherlands and Flanders (the

Dutch-speaking part of Belgium). Relevant findings from studies 1 and 2 were considered in the subsequent studies.

Study 1 followed an iterative, mixed-methods approach to develop two inventories, the Actions to Improve Mood by Residents (AIM-R) and the Actions to Improve Mood by Caregivers (AIM-C), for mapping mood-improving behaviors employed by residents themselves and professional nursing home caregivers, respectively. The final versions of these inventories were tested for content validity. In Study 2, the test-retest agreement and inter-rater agreement of the inventories were assessed using a repeated-measures design with two time points. Both inventories were completed twice by the same participants approximately two weeks apart (T0 and T1), which has been suggested as an appropriate time interval to assess test-retest agreement [25]. To test the inter-rater agreement of the AIM-C, two caregivers were invited to complete this inventory for the same resident. To note, as a result of using a formative measurement model, internal consistency measures (e.g., Cronbach's alpha) will not be reported. Instead, test-retest agreement and inter-rater agreement will serve as the primary indicators of reliability of the inventories. In Study 3, a qualitative approach involving semi-structured interviews with stakeholders was employed to explore their experiences in using the inventories in daily practice, specifically focusing on potential issues with psychometrics and usability.

General characteristics of participants in all studies

The research team invited nursing home residents and professional caregivers to participate in one or more studies. The inclusion criteria for residents were the ability to provide informed consent and the willingness and capacity to (a) verbally communicate their opinion about the inventories (studies 1 and 3), or (b) fill out questionnaires together with one of the researchers (study 2). Residents from short-stay units (i.e., rehabilitation care units) were excluded. We aimed to include residents with and without dementia, as well as those with and without depression, making the study inclusive of various resident profiles. The general criterion for caregivers was working primarily as employees providing direct care to residents (e.g., registered nurses and certified nurse assistants). To participate in study 2, caregivers needed to be significantly involved in the care of one or more participating residents.

In addition to residents and nurses, other professionals were involved in the studies. In study 1, the research team members, with expertise in psychology, gerontology, nursing sciences, and psychometrics, contributed to developing and refining the inventories. In study 2, nursing home staff members (e.g., psychologist, physician, and registered nurse) were consulted to provide additional information about the participating residents (e.g., level of cognitive functioning). For study 3, alongside input from residents and caregivers, we also gathered feedback from the interviewers who conducted the in-person interviews with residents in study 2, focusing on their experiences with administering the AIM-R.

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki [26] and complied with Dutch and Belgian laws. Ethical approval for the study was obtained from the legally designated ethics committees in both countries (CMO Radboudumc, reference number: 2021–11047 in the Netherlands; CME VUB, reference numbers: EC-2021-277 and EC-2021-432 in Belgium). Prior to participation, all participants were informed about the study and provided written informed consent.

Study 1: development process and content validity

Methods

Procedure and materials

Development process The initial versions of the AIM-R and AIM-C were based on lists of informal mood-improving behaviors in nursing homes identified in two previously conducted group concept mapping studies [27]. These studies used a bottom-up method to identify, prioritize, and cluster mood-improving behaviors by nursing home residents themselves and by others important in residents' lives.

Taking into account the intended principles formulated in the introduction, members of the research team critically considered and discussed these (clusters of) behaviors, grouping them into initial themes (e.g., "exercise" and "music"). The research team also formulated initial items with response categories, which were further refined through expert consultations. Next, preliminary versions of the AIM-R and AIM-C were piloted, focusing on layout and content (i.e., relevance, applicability of the actions in daily practice, comprehensiveness, and comprehensibility [i.e., understandability of the instructions, examples, items, and response options]). For the AIM-R, two skilled interviewers (IK and ID) conducted semi-structured interviews with residents based on cognitive interviewing techniques [28]. In these interviews, parts of the AIM-R were presented to residents. The interviews were supported by a pre-defined interview guide with think-aloud and verbal probing procedures. Examples of prompt questions included "please tell me your thoughts while reading/hearing this question," and "what kinds of situations were you thinking about in coming up with this answer?" Open-ended and closed questions were included in the interviews. During the interviews, the interviewers took notes on their observations and residents' comments. Feedback on the AIM-C inventory

was solicited from caregivers via an online survey tool [29] employing questions similar to those posed to the residents.

Throughout this iterative procedure, the research team discussed the findings and made adjustments to the inventories as needed. New versions were then presented to additional participants (residents and caregivers) until no further improvements were identified, and the inventories were considered acceptable.

Content validity After the inventories were refined and the most appropriate versions were chosen, the AIM-R and AIM-C were assessed for content validity using a structured in-person interview with residents (for the AIM-R) and an online survey [29] for caregivers (for the AIM-C). To test content validity, participants were asked about the relevance ("not relevant," "somewhat relevant," "quite relevant," or "highly relevant") [30] of each theme as a whole (i.e., not the individual items within the theme in relation to residents' mood.

Analysis

To analyze the qualitative data in the development process of the AIM-R and AIM-C (i.e., responses to the open-ended questions on the survey and interviews, and additional notes from the interviews with residents), one researcher (IK) performed thematic analyses [31] in ATLAS.ti 9 [32]. A second researcher (RL) checked the initial codes attached to the text segments, after which they were discussed between the two researchers until consensus was reached.

To assess the content validity of the final versions of the inventories, we calculated Content Validity Indices (CVI) in SPSS 27 [33] for each separate theme (item-level CVI [I-CVI]) [34]. The I-CVI was computed by dividing the number of participants who rated a theme as at least "quite relevant" by the total number of participants. While an I-CVI can range between 0.00 and 1.00, an I-CVI of at least 0.78 is usually considered acceptable [35]. To reach consensus on the omission of a theme, the research team discussed the eligibility of themes that did not meet this cutoff score. The decisions were guided by the principles outlined in the introduction, while carefully considering the corresponding I-CVI level.

Results

Participants

During the iterative process, preliminary versions of the AIM-R and AIM-C were piloted with 34 residents (mean age, 81.8 years; SD, 7.4; range, 67–92) and 66 caregivers (mean age, 44.6 years; SD, 12.8; range, 21–68). Most participants were female (residents, $N = 25$ [73.5%]; caregivers, $N = 61$ [92.4%]) and lived in the Netherlands (residents, $N = 22$ [64.7%]; caregivers, $N = 59$ [89.4%]). Caregivers worked mainly as certified nurse assistants ($N = 31$ [47.0%]) or as registered nurses ($N = 19$ [43.9%]) and had a mean working experience of 14.2 years (SD, 12.9; range, 0–45). For the final versions of the inventories, 31 residents (23 of whom had already participated in the pilot with a preliminary version of the AIM-R) and 35 caregivers (22 of whom had provided feedback on a preliminary version of the AIM-C) answered questions related to the content validity of the final versions of the inventories.

Development process

In our examination of informal antidepressant behaviors, we observed a broad range of actions in which individuals engage to improve residents' mood, supporting the appropriateness of a formative measurement model [24] for the construct of "mood-improving behaviors." Regarding the scope of the inventories, the AIM-R is aimed at actions performed by residents themselves, whereas the AIM-C is directed at actions performed by caregivers. As resident care often involves multiple caregivers, we expanded the scope of the AIM-C to include actions performed by several caregivers, not solely the person completing the inventory. As a result, the AIM-C inventory inquires about actions undertaken by both the persons completing the inventory and their colleagues.

In terms of content, we deliberately chose content thematic categories and their associated behaviors for both residents and caregivers. The primary focus was on identifying themes of caregiver and resident behaviors that contribute to improved mood. This approach recognizes that the inventories can encompass elements within or extending beyond formal interventions. Consequently, we have not omitted themes that could also involve formal interventions, such as physical activity. This inclusivity is due to caregivers potentially promoting resident engagement in these activities or residents independently participating in them. By adopting a comprehensive approach, we explored a broader spectrum of behaviors that have the potential to positively influence residents' mood.

Given the aim of offering practical guidance for specific enhancements, we included identical themes in both the AIM-C and AIM-R inventories. This should enable a direct comparison between residents' and caregivers' mood-improving behaviors, which might provide better insight into the actions needed from each group. This could also facilitate an understanding of the roles and responsibilities of both residents and caregivers in enhancing mood. By identifying shared themes, it may be possible to tailor interventions more effectively and foster a collaborative approach in promoting positive mood outcomes in residents.

Final versions and content validity

The final versions of the AIM-R and AIM-C (see additional files 1 and 2) covered 18 themes, with two items per theme. Examples of themes include "physical activity," "contact with others," and "music." Each theme contains a description (e.g., AIM-R: "Activities related to music. Think of making music or listening to it attentively") and examples of possible actions (e.g., AIM-R: "Listening to the radio, singing, playing an instrument, going to a concert, or any other music-focused activity"). The description and examples are followed by two items: a question about the action's frequency (e.g., AIM-R: "concerning music-focused activities, how often do you do this in an average week?"; a 5-point scale ranging from "never" to "very often") and a question about the expected effect of the behavior on residents' mood (e.g., AIM-R: "if you do [or would do] this, does this make you feel happy?"; three response categories: "no," "yes, a bit happy," and "yes, very happy"). For the AIM-C, the descriptions, examples, and items about the action's frequency are directed at caregivers' actions, whereas the item about the expected effect is focused on residents' mood (i.e., AIM-C: "if you do [or would do] this, does this make the resident feel happy?"). Caregivers are instructed to report on actions that are performed for a specific resident in an average week.

The I-CVI for the themes was above 0.78 for all but one theme ("household activities," I-CVI = 0.73) of the AIM-R, and all but four themes of the AIM-C ("household activities," "stimulating the senses," "faith and meaning," and "doing something for someone else," I-CVI = 0.69 for the first and 0.71 for the latter three themes) (see Table 1).

Table 1 Content Validity Index of the Themes of the AIM-R and AIM-C

Theme	AIM-R, *N* = 31	AIM-C, *N* = 35
1. Physical activity	0.90	1.00
2. Contact by touching	0.90	1.00
3. Taking care of the appearance	0.84	0.91
4. Relaxation	0.81	0.94
5. Healthy living	0.87	0.89
6. Household activities	0.73	0.69
7. Creating something	0.93	0.97
8. Brain stimulation	0.90	0.94
9. Music	0.93	1.00
10. Going out	0.97	0.89
11. Precious memories	0.94	0.91
12. Stimulating the senses	0.90	0.71
13. Contact with others	0.93	1.00
14. Nature	0.94	0.89
15. Faith and meaning	0.94	0.71
16. Doing something for someone else	0.90	0.71
17. Security and warmth	0.87	0.94
18. Positive attitude	0.93	0.94

AIM-R Actions to Improve Mood by Residents, *AIM-C* Actions to Improve Mood by Caregivers

Considerations for study 2

In study 1, the AIM-R and AIM-C inventories were developed and tested for their content validity. During the content validation process, we found that some themes had an I-CVI below the acceptable threshold. However, the team made a deliberate decision to retain these themes in the inventories for two reasons. First, despite certain themes falling short of the recommended standard of acceptable content validity in one inventory, they met the criterion in the other inventory. For instance, the themes "stimulating the senses," "faith and meaning," and "doing something for someone else" were deemed less relevant in the AIM-C, but demonstrated acceptable content validity in the AIM-R. This suggests that while the importance of actions within a theme may differ between residents and caregivers, the overall content validity of the theme remains intact. Second, even though some themes may have less general relevance for mood-improving behaviors, they may hold substantial importance for specific sub-populations of residents. For instance, the theme "stimulating the senses" may be particularly significant for residents with (severe) cognitive decline [36], whereas being active in "household activities" may hold promise for those with or without mild cognitive decline [37, 38]. Given our aim of developing inclusive inventories that are also suitable for residents with cognitive decline, the research team decided against excluding any themes on the basis solely of their I-CVI levels.

In conclusion, despite observing variations in I-CVI levels, all 18 themes were retained in both inventories to ensure comprehensive coverage of mood-improving behaviors and to accommodate the diverse needs of nursing home residents. Including themes relevant to specific sub-populations aligns with our goal of developing practical and inclusive tools for assessing mood-improving behaviors in residents with varying physical and cognitive abilities.

Study 2: test-retest and inter-rater agreement

Methods

Procedure and materials

In addition to demographic characteristics, residents were asked to answer the questions of the AIM-R in a structured in-person meeting with one of seven interviewers (two psychologists, one research employee, three master's students in psychology, and one master's student in gerontology; four interviewers living in the Netherlands, and the other three living in Belgium). For each participating resident, two caregivers were subsequently asked to complete the AIM-C individually with respect to actions that they or their colleagues perform for (a) specific resident(s), alongside answering questions about their demographic characteristics, either on paper or online [29]. Two weeks later, residents and caregivers

completed the AIM inventories again. Caregivers completed the inventory again for the same resident(s) as during the baseline measurement.

Additional data about the participating residents were collected by asking a therapist (in most cases, a psychologist) or registered nurse to complete an online form [29]. This form consisted of questions about the residents, including (a) type of unit ("open" or "closed" and "medical-somatic care," "psychogeriatric care," "mental-physical multimorbidity care," or "mixed"), (b) depression diagnosis ("yes," "no," or "don't know"), and (c) stage of cognitive function using the Global Deterioration Scale (GDS) [39]. All data were collected between February and December 2022.

Analysis

To assess the agreement between the two measurement points (test-retest agreement between T0 and T1) of the items of the AIM-R and AIM-C and the agreement between the two caregivers (inter-rater agreement at baseline) of the AIM-C items, Gwet's AC2 coefficients [40] were calculated using the irrCAC R package [41] in R [42]. Only responses within a time frame of 10 to 31 days between baseline and T1 were included in the assessment of test-retest agreement. Gwet's AC2 coefficient was chosen given its resilience to grey zones (i.e., areas in the data where agreement between raters is low or moderate) and variation in distributions (e.g., skewed distributions) [43–45].

Weighting schemes ranging from 0.00 (full disagreement) to 1.00 (full agreement) were specified to the distance of disagreement within measurement points (test-retest) and between caregivers (inter-rater), where a higher score indicates a higher level of agreement. A linear weighting scheme was applied to the items regarding frequency. This means that an evenly spaced weight of 0.25 was applied to each distance of disagreement (e.g., a weight of 0.75 regarding a disagreement distance of one point on the 5-point scale, and a weight of 0.50 for a disagreement distance of two points). For the items regarding the expected effect on residents' mood (a 3-point scale), adapted weights were specified for the degree of disagreement. These weights considered any disagreement with the response option "no" (weight of 0.25 for disagreement with "yes, a bit happy" or weight of 0.00 for disagreement with "yes, very happy") to be more significant than a disagreement between the response options "yes, a bit happy" and "yes, very happy" (weight of 0.75).

To characterize the magnitude of the AC2 coefficients, we used Altman's benchmarking 5-point scale ranging from "poor" to "very good" [46]. To calculate the probability that the coefficients would fall into each category, the agreement coefficient and its standard error were used. As recommended by Gwet [40], a cumulative probability above 0.95 was applied to determine the lowest expected agreement strength level.

In addition to analyses of the total sample, explorative subgroup analyses were performed. For both inventories, agreement outcomes were explored, broken down by country (the Netherlands or Belgium) and level of cognitive functioning (residents with no to mild cognitive decline [GDS score ≤ 3] and residents with moderate to severe cognitive decline [GDS score ≥ 4]).

Results

Participants

In total, 405 nursing home residents were invited to participate in study 2. At baseline, 273 residents (67.4% of the invited residents, mean age 82.4 years; SD, 9.3; median, 84.0; range, 47–102) from 30 nursing homes (16 of which are located in the Netherlands) completed the AIM-R, whereas 119 unique caregivers (mean age, 39.5 years; SD, 11.6; range, 22–65) filled out the AIM-C for at least one resident (see Tables 2 and 3). In total, the AIM-C was completed 268 times. For 187 residents (68.5%), at least one caregiver completed the AIM-C. Of those, the AIM-C was filled out by two caregivers for 81 residents (29.7% of the total participating residents). On average, a caregiver filled out the AIM-C for two residents (mode, 1; median, 1; range, 1–11). The mean time between the two administrations of the AIM-C by different caregivers was 11.0 days (SD, 12.5; mode, 1; median, 6; range, 0–65). Additional resident data about the residence unit, depression diagnosis, and stage of cognitive functioning were provided by treatment staff members for 205 residents (75.1% of the total). At the two-week measurement, 209 AIM-R and 151 AIM-C inventories were completed, of which 206 AIM-R (75.5% of completed baseline inventories) and 125 AIM-C (46.6% of completed baseline inventories) administrations fell within the acceptable time frame of 10 to 31 days between baseline and follow up.

Test-retest and inter-rater agreement

For the test-retest agreement in the total sample, the AC2 coefficients ranged from 0.55 to 0.87 for the AIM-R, and from 0.60 to 0.86 for the AIM-C inventory (see Table 4). For both inventories, the classification of the agreement strength ranged from "moderate" to "very good," where most items were characterized as "good" (23 out of 36 items for the AIM-R and 24 out of 36 items for the AIM-C). Although differences between subgroups were not statistically tested, there seemed to be slightly better test-retest agreement for the inventories administered in the Netherlands (where most items of both inventories were characterized as "good") than for those administered in Belgium (where most items were classified as "moderate" or "good" for the AIM-R and as "fair" or "moderate" for the AIM-C) (see additional file 3). In general, similar

Table 2 Characteristics of Residents for Study 2

Characteristic	T0	T1 (2 weeks)
N	273	209
Gender, female, *N*(valid %) / male, *N*	172 (63) / 101	124 (59) / 85
Age, mean (SD) [range]	82.4 (9) [47–102]	82.4 (9) [47–102]
Country, the Netherlands, *N*(%) / Belgium, *N*	140 (51) / 133	123 (59) / 86
Educational attainment		
Low, *N*(%)	135 (50)	106 (51)
Medium, *N*(%)	91 (33)	71 (34)
High, *N*(%)	47 (17)	32 (15)
Marital status		
Unmarried, *N*(%)	33 (12)	22 (11)
Married or partnered in a registered partnership, *N*(%)	55 (20)	44(21)
Widowed (after marriage or registered partnership), *N*(%)	150 (55)	115 (55)
Divorced (after marriage or registered partnership), *N*(%)	33 (12)	26 (12)
Missing, *N*(%)	2 (1)	2 (1)
Type of unit (A)		
Medical-somatic care, *N*(%)	80 (29)	75 (36)
Psychogeriatric care, *N*(%)	50 (18)	34 (16)
Mental-physical multimorbidity care, *N*(%)	25 (9)	18 (9)
Mixed, *N*(%)	118 (43)	82 (39)
Type of unit (B)		
Open, *N*(%)	187 (69)	149 (71)
Closed (code is unknown for resident), *N*(%)	59 (22)	39 (19)
Closed (code is known for resident), *N*(%)	27 (10)	21 (10)
Depression diagnosis		
Yes, *N*(%)	37 (14)	24 (12)
No, *N*(%)	160 (59)	135 (65)
Don't know, *N*(%)	8 (3)	5 (2)
Missing, *N*(%)	68 (25)	45 (22)
Global Deterioration Scale		
Stage 1: No cognitive decline, *N*(%)	42 (15)	37 (18)
Stage 2: Very mild cognitive decline, *N*(%)	69 (25)	55 (26)
Stage 3: Mild cognitive decline, *N*(%)	31 (11)	25 (12)
Stage 4: Moderate cognitive decline, *N*(%)	13 (5)	11 (5)
Stage 5: Moderately severe cognitive decline, *N*(%)	25 (9)	21 (10)
Stage 6: Severe cognitive decline, *N*(%)	25 (9)	15 (7)
Stage 7: Very severe cognitive decline, *N*(%)	0 (0)	0 (0)
Missing, *N*(%)	68 (25)	45 (22)

Table 3 Characteristics of Professional Caregivers for Study 2

Characteristic	T0	T1 (2 weeks)
N	119	59
Gender, female, *N*(valid %) / male, *N*	108 (91) / 9	52 (88) / 7
Age, mean (SD) [range]	39.5 (12) [22–65]	39.2 (12) [22–65]
Country, the Netherlands, *N*(%) / Belgium, *N*	52 (44) / 67	32 (54) / 27
Educational attainment		
Low, *N*(%)	31 (26)	12 (20)
Medium, *N*(%)	64 (54)	31 (53)
High, *N*(%)	24 (20)	16 (27)
Type of health care provider		
Registered nurse, *N*(%)	33 (28)	19 (32)
Certified nurse assistant, *N*(%)	74 (62)	34 (58)
Nurse assistant / nurse aide, *N*(%)	2 (2)	1 (2)
Other, *N*(%)	7 (6)	5 (8)
Missing, *N*(%)	3 (3)	0 (0)
Years of working experience, mean (SD) [range]	10.9 (10) [0–44]	11.6 (11) [1–44]
Frequency of involvement in the care of residents with depression		
Never, *N*(%)	1 (1)	0 (0)
Occasionally, *N*(%)	28 (24)	14 (24)
Regularly, *N*(%)	57 (48)	31 (53)
Often, *N*(%)	23 (19)	8 (14)
Very often, *N*(%)	8 (7)	6 (10)
Missing, *N*(%)	2 (2)	0 (0)

The characteristics of unique caregivers are provided

levels of agreement were found when the inventories were completed for residents with no to mild cognitive decline compared with residents with moderate to severe cognitive decline (i.e., for both subgroups, the agreement levels of most items of the AIM-R were classified as "good," whereas most items of the AIM-C were categorized as "moderate" or "good").

For the inter-rater agreement of the AIM-C, the AC2 coefficients ranged from 0.21 to 0.76 for the total sample, with the lowest expected levels of agreement strength ranging from "poor" (5 items) to "good" (1 item) (see Table 5). The most frequent classification of the agreement strength was "fair" (22 out of 36 items). There seemed to be better agreement when caregivers completed the inventories for residents with no to mild cognitive decline (the most frequent item coefficients were classified as "fair" [18 items] or "moderate" [13 items]) than when inventories were completed for residents with moderate to severe cognitive decline (the most common level of agreement was "poor" [23 items]) (see additional file 4). Furthermore, comparable levels of inter-rater agreement for the AIM-C items were found among caregivers in the Netherlands and Belgium.

Table 4 Agreement Statistics for Test-Retest Agreement of the AIM-R and AIM-C

Theme	Item	AIM-R					AIM-C				
		N	% Obs	Agreement diagonal[a]	Gwet's AC2 estimate [95% CI]	Altman's benchmark scale[b]	*N*	% Obs	Agreement diagonal[a]	Gwet's AC2 estimate [95% CI]	Altman's benchmark scale[b]
1. Physical activity	Frequency	202	86.3	19 26 29 12 37	0.67 [0.60;0.74]	Good	125	84.8	14 17 18 13 4	0.64 [0.57;0.72]	Moderate
	Expected effect	199	86.9	28 47 68	0.72 [0.64;0.80]	Good	123	81.5	17 37 27	0.60 [0.48;0.73]	Moderate
2. Contact by touching	Frequency	193	85.1	35 34 21 10 11	0.65 [0.58;0.72]	Moderate	125	87.4	15 32 8 8 2	0.73 [0.67;0.79]	Good
	Expected effect	193	88.7	27 39 84	0.77 [0.69;0.85]	Good	119	84.9	12 46 27	0.69 [0.59;0.80]	Good
3. Taking care of the appearance	Frequency	199	87.8	15 55 26 8 19	0.72 [0.66;0.78]	Good	124	86.3	11 16 21 14 10	0.67 [0.59;0.75]	Good
	Expected effect	199	88.8	33 45 76	0.76 [0.68;0.84]	Good	122	87.3	11 40 41	0.74 [0.64;0.84]	Good
4. Relaxation	Frequency	189	83.5	62 12 9 5 21	0.63 [0.55;0.71]	Moderate	125	86.8	40 16 11 5 4	0.72 [0.64;0.80]	Good
	Expected effect	189	84.9	62 37 36	0.66 [0.57;0.76]	Moderate	114	83.3	30 43 9	0.67 [0.56;0.78]	Moderate
5. Healthy living	Frequency	187	85.7	15 11 11 11 61	0.67 [0.59;0.75]	Good	123	84.3	11 19 18 11 4	0.64 [0.56;0.72]	Moderate
	Expected effect	187	83.7	36 54 36	0.64 [0.55;0.74]	Moderate	120	83.3	18 53 9	0.69 [0.58;0.79]	Good
6. Household activities	Frequency	197	89.6	97 21 11 3 8	0.80 [0.74;0.86]	Good	124	88.3	52 16 5 4 2	0.77 [0.70;0.84]	Good
	Expected effect	194	86.3	69 46 27	0.70 [0.62;0.79]	Good	113	84.5	58 20 6	0.72 [0.61;0.83]	Good
7. Creating something	Frequency	198	92.3	90 30 10 4 10	0.85 [0.80;0.89]	Very good	125	87.0	27 19 12 12 1	0.70 [0.63;0.77]	Good
	Expected effect	186	87.5	56 29 59	0.73 [0.64;0.81]	Good	117	87.6	29 35 25	0.73 [0.63;0.83]	Good
8. Brain stimulation	Frequency	199	84.2	11 5 6 10 79	0.67 [0.59;0.75]	Good	125	86.0	21 16 18 10 4	0.67 [0.60;0.75]	Good
	Expected effect	199	86.6	15 39 84	0.74 [0.66;0.82]	Good	121	86.6	22 46 20	0.72 [0.63;0.82]	Good
9. Music	Frequency	198	86.0	26 24 18 15 39	0.66 [0.58;0.73]	Moderate	124	85.1	23 16 18 6 2	0.66 [0.58;0.73]	Moderate
	Expected effect	201	91.3	22 31 110	0.84 [0.78;0.90]	Good	118	84.7	20 35 28	0.67 [0.55;0.79]	Moderate
10. Going out	Frequency	194	87.8	27 55 18 4 9	0.74 [0.69;0.79]	Good	123	90.0	28 33 16 8 1	0.78 [0.70;0.85]	Good
	Expected effect	194	87.9	8 29 111	0.79 [0.72;0.87]	Good	120	91.5	14 28 54	0.83 [0.75;0.91]	Good
11. Precious memories	Frequency	198	84.3	17 34 22 7 18	0.63 [0.57;0.70]	Moderate	125	85.8	5 28 23 11 0	0.70 [0.63;0.77]	Good
	Expected effect	189	87.0	26 43 65	0.73 [0.65;0.81]	Good	123	91.7	3 50 40	0.85 [0.79;0.91]	Good
12. Stimulating the senses	Frequency	173	85.7	81 17 2 1 7	0.73 [0.66;0.81]	Good	125	90.4	65 17 3 1 1	0.84 [0.78;0.90]	Good
	Expected effect	163	84.4	61 28 28	0.66 [0.56;0.76]	Moderate	113	86.7	59 27 4	0.77 [0.67;0.86]	Good
13. Contact with others	Frequency	194	84.4	4 19 15 12 52	0.64 [0.58;0.71]	Moderate	125	84.8	6 15 22 20 2	0.64 [0.56;0.72]	Moderate
	Expected effect	193	92.1	6 36 115	0.87 [0.82;0.92]	Very good	122	85.2	8 34 42	0.70 [0.60;0.81]	Good
14. Nature	Frequency	196	83.9	16 33 17 13 25	0.61 [0.54;0.68]	Moderate	125	86.4	37 16 10 6 0	0.71 [0.64;0.78]	Good
	Expected effect	198	89.8	9 37 104	0.82 [0.76;0.88]	Good	117	84.0	31 26 26	0.64 [0.52;0.76]	Moderate
15. Faith and meaning	Frequency	189	91.1	66 27 14 3 33	0.80 [0.74;0.86]	Good	124	90.3	44 33 5 2 2	0.82 [0.76;0.88]	Good
	Expected effect	183	89.6	63 42 36	0.77 [0.70;0.84]	Good	117	82.5	52 19 11	0.66 [0.54;0.77]	Moderate
16. Doing something for someone else	Frequency	188	85.8	60 17 12 7 16	0.68 [0.60;0.75]	Good	125	89.4	52 19 2 9 2	0.79 [0.72;0.86]	Good
	Expected effect	189	86.4	33 32 68	0.71 [0.63;0.79]	Good	115	84.1	46 26 12	0.68 [0.56;0.79]	Moderate
17. Security and warmth	Frequency	173	84.8	21 28 21 10 14	0.64 [0.57;0.71]	Moderate	125	85.4	12 31 11 10 5	0.67 [0.59;0.75]	Good
	Expected effect	178	89.7	13 40 76	0.80 [0.74;0.87]	Good	121	82.9	11 38 33	0.64 [0.53;0.76]	Moderate
18. Positive attitude	Frequency	177	80.9	7 14 15 12 40	0.55 [0.47;0.64]	Moderate	125	86.2	1 6 19 31 14	0.70 [0.63;0.78]	Good
	Expected effect	181	89.2	10 38 82	0.80 [0.74;0.87]	Good	124	91.9	1 39 50	0.86 [0.81;0.91]	Very good

AIM-R Actions to Improve Mood by Residents, *AIM-C* Actions to Improve Mood by Caregivers, *N* Valid number of participants, *% Obs* Percentage observed, *CI* Confidence Interval

[a] The agreement diagonal is the part of the contingency table where raters agree on the same categories

[b] Altman's benchmarking 5-point scale ranging from "poor" to "very good" was used to interpret the magnitude of the AC2 coefficients. A cumulative probability of above 0.95 was applied to determine the lowest expected agreement level

Table 5 Agreement Statistics for Inter-Rater Agreement of the AIM-C

Theme	Item	*N*	% Obs	Agreement diagonal[a]	Gwet's AC2 estimate [95% CI]	Altman's benchmark scale[b]
1. Physical activity	Frequency	81	74.4	6 9 12 3 1	0.41 [0.27;0.55]	Fair
	Expected effect	80	76.6	6 17 17	0.51 [0.35;0.67]	Fair
2. Contact by touching	Frequency	81	75.9	8 10 4 5 1	0.45 [0.33;0.57]	Fair
	Expected effect	79	75.6	5 26 10	0.51 [0.34;0.68]	Fair
3. Taking care of the appearance	Frequency	81	71.9	1 6 7 5 2	0.36 [0.23;0.48]	Fair
	Expected effect	79	80.1	3 19 16	0.61 [0.48;0.74]	Moderate
4. Relaxation	Frequency	81	77.2	18 9 4 1 1	0.49 [0.36;0.62]	Fair
	Expected effect	74	66.6	8 25 2	0.35 [0.16;0.55]	Poor
5. Healthy living	Frequency	80	68.4	0 5 4 7 0	0.27 [0.13;0.40]	Poor
	Expected effect	77	69.2	5 22 3	0.40 [0.22;0.57]	Fair
6. Household activities	Frequency	80	79.7	28 5 2 0 0	0.61 [0.49;0.73]	Moderate
	Expected effect	73	70.9	28 10 3	0.46 [0.29;0.63]	Fair
7. Creating something	Frequency	81	77.8	13 7 7 5 1	0.49 [0.37;0.62]	Fair
	Expected effect	74	77.0	15 15 13	0.49 [0.31;0.66]	Fair
8. Brain stimulation	Frequency	80	76.3	9 6 13 4 0	0.46 [0.32;0.59]	Fair
	Expected effect	77	71.4	5 21 9	0.42 [0.24;0.59]	Fair
9. Music	Frequency	81	71.3	5 5 9 4 1	0.32 [0.18;0.46]	Fair
	Expected effect	75	67.7	2 18 14	0.32 [0.11;0.53]	Poor
10. Going out	Frequency	80	75.0	8 18 3 3 1	0.43 [0.28;0.59]	Fair
	Expected effect	77	73.7	0 16 21	0.49 [0.32;0.67]	Fair
11. Precious memories	Frequency	81	77.8	0 11 12 5 1	0.52 [0.42;0.63]	Moderate
	Expected effect	79	80.7	0 22 17	0.64 [0.52;0.77]	Moderate
12. Stimulating the senses	Frequency	81	80.6	35 6 2 1 0	0.66 [0.54;0.78]	Moderate
	Expected effect	71	69.0	22 17 1	0.41 [0.23;0.59]	Fair
13. Contact with others	Frequency	81	72.8	2 7 8 5 1	0.36 [0.22;0.50]	Fair
	Expected effect	78	80.4	2 16 20	0.63 [0.50;0.76]	Moderate
14. Nature	Frequency	81	73.5	15 12 3 2 0	0.43 [0.28;0.57]	Fair
	Expected effect	73	64.4	7 11 12	0.21 [0.00;0.42]	Poor
15. Faith and meaning	Frequency	81	81.2	20 9 7 1 0	0.63 [0.52;0.73]	Moderate
	Expected effect	72	65.3	20 12 3	0.32 [0.14;0.50]	Poor
16. Doing something for someone else	Frequency	81	78.7	23 9 2 1 1	0.56 [0.43;0.68]	Moderate
	Expected effect	72	71.5	22 12 5	0.40 [0.22;0.59]	Fair
17. Security and warmth	Frequency	81	73.1	3 14 2 4 2	0.38 [0.23;0.53]	Fair
	Expected effect	76	71.7	2 17 14	0.42 [0.24;0.61]	Fair
18. Positive attitude	Frequency	81	73.8	0 1 6 8 4	0.43 [0.32;0.54]	Fair
	Expected effect	81	86.1	0 18 26	0.76 [0.68;0.84]	Good

AIM-C Actions to Improve Mood by Caregivers, *N* Valid number of participants, *% Obs* Percentage observed, *CI* Confidence Interval

[a] The agreement diagonal is the part of the contingency table where raters agree on the same categories

[b] Altman's benchmarking 5-point scale ranging from "poor" to "very good" was used to interpret the magnitude of the AC2 coefficients. A cumulative probability of above 0.95 was applied to determine the lowest expected agreement level

Considerations for study 3

Study 2 assessed the test-retest agreement and inter-rater agreement of the AIM-R and AIM-C. The items of both inventories appeared to have at least a moderate level of test-retest agreement for the total sample, which can be considered acceptable. For both inventories, the results suggested slightly worse test-retest agreement for the Belgian sub-sample than for the sub-sample from the Netherlands. Limited levels of inter-rater agreement of AIM-C items were found for the total sample as well as for all the sub-samples, especially when used for residents with moderate to severe cognitive decline. To explore these findings further, we address these topics in study 3.

Study 3: experiences with applying the inventories in daily practice

Methods

Procedure, materials, and analysis

A convenience sample of residents, caregivers, and interviewers was drawn from study 2. We used an interview guide with the following topics: (a) general experiences when administering or completing the inventory, (b)

challenging aspects (including potential issues related to limited levels of agreement), and (c) suggestions for improvements. When not mentioned spontaneously, in-depth questions about the inventories' content, questions, administration, and country-specific challenges were utilized. Thematic analysis [31] was performed via ATLAS.ti 9 [32], whereby the notes of all interviews were coded and discussed by two researchers (IK and RL).

Results

Participants

Five residents (mean age 78.3 years, 4 female) and six caregivers (mean age 40.1 years, all female) shared their opinions of positive and challenging aspects of using the AIM-R and AIM-C and advised on potential improvements. In addition, all seven interviewers from study 2 shared their opinions and suggestions for improvements regarding administering the AIM-R to residents.

Experiences with the application of the inventories in daily practice

Although both inventories were positively evaluated as very comprehensive, challenges were mentioned in terms of their usability and interpretability. The thematic analysis suggested that the interpretation of some themes of both inventories (e.g., "healthy living," "nature," and "contact with others") may not have been worded specifically enough. Furthermore, some of the provided examples seemed to cause confusion as they could also be associated with another similar theme. For example, "listening to relaxing music" (theme "relaxation") also seemed to be interpreted as the more general "listening to music" within the theme "music." Belgian interviewers also raised concerns about language issues in the inventories. Therefore, stakeholders recommended providing more specific descriptions of the themes that were evaluated as too broad and specification or reformulation of examples within these themes.

In general, both inventories were evaluated as applicable for residents with varying degrees of physical and cognitive ability. For the AIM-R, the interviewers suggested that the interpretability of some questions may have been challenging for residents with advanced dementia, mainly because of the large amount of text in the inventory. Nevertheless, according to the interviewers, querying most themes seemed possible for a considerable number of residents with cognitive decline. Therefore, the interviewers emphasized that it is not desirable or necessary to refrain from trying to administer this inventory to residents with dementia. Taking the time to clarify the themes and questions, or administering the inventory over multiple meetings, were considered helpful strategies for obtaining answers from this resident group. Although it was considered time-consuming, administering the inventory by means of an individual structured face-to-face interview was evaluated as an adequate method for gathering relevant information from residents. The descriptions of themes and examples within each theme were also regarded as helpful for usability and understandability. To further simplify the inventories, reducing the number of examples (max. 3 common example actions per theme) and simplifying the answer options were encouraged (i.e., "never," "regularly," and "often" for the action frequency; "yes" and "no" for the expected effect on residents' mood). According to the participants, a more picture-oriented inventory could also be considered, as it may improve the usability and interpretability of the themes and questions for residents. As some themes (e.g., "household activities") were deemed applicable only for some residents, stakeholders suggested an additional answer option: "the resident is unable to perform the behavior described."

Although administering the AIM-R may have elicited negative associations (e.g., associations with unpleasant memories) in some residents, both the interviewers and the residents indicated that, in general, residents seemed to enjoy being asked about pleasant activities. It was stressed that for most residents, the inventory evoked positive associations (e.g., recent pleasant activities and fond memories). The interviewers indicated that the questions provided an opportunity for residents to discuss their experiences and preferences. In this way, it places more emphasis on residents' agency. According to both residents and interviewers, the inventory may also raise awareness of possible mood-improving behaviors and may motivate residents to adopt them. The interviewers especially noted this in residents with no to limited cognitive decline.

For the AIM-C, caregivers indicated that they were not always aware of their colleagues' activities, which complicated answering the questions. They suggested that this may have led to varying interpretations of the AIM-C items among caregivers. Caregivers also noted that some themes were less applicable for residents with advanced dementia than they were for other residents (e.g., "doing something for someone else"). They recommended the inclusion of comment fields as a means of incorporating relevant notes for practical guidance.

As a positive outcome of the AIM-C, caregivers mentioned that they had become more aware of mood-improving behaviors to employ in daily practice. They expected that the inventory may provide valuable information about caregivers' actions and their effects, and that the AIM-C may be used as a valuable guide for getting to know residents.

General discussion

In three studies, two inventories for mapping mood-improving behaviors in nursing homes were developed and evaluated: the Actions to Improve Mood by

Residents (AIM-R) and the Actions to Improve Mood by Caregivers (AIM-C). The AIM-R focuses on actions initiated by residents themselves, while the AIM-C focuses on actions initiated by caregivers. Both inventories encompass 18 similar themes, each with two items per theme (frequency and expected effect). Content validity assessment revealed that the themes reflected relevant aspects of mood-improving behaviors. Moderate to very good test-retest agreement was observed across the items of both inventories. However, only a few AIM-C items were classified as having at least moderate inter-rater agreement. Stakeholders' experiences with the inventories highlighted both positive aspects, such as increased awareness of mood-improving behaviors, as well as and challenges, such as broad formulation of themes. Their constructive feedback provided valuable insights for further enhancing the usability and interpretability of the inventories.

While the themes in both inventories were meant to capture relevant aspects of mood-improving behaviors, a few themes did not meet the cutoff score for acceptable content validity, as recommended by Lynn [35]. However, the research team deliberately decided to retain these themes, as the behaviors might hold significance for specific individuals and because our aim was to develop inclusive inventories suitable for a large proportion of nursing home residents. Nevertheless, some themes may still be considered less relevant for some residents, suggesting that some content of the inventories may not be applicable for all nursing home residents. Themes falling below the recommended content validity threshold may require further conceptual clarity and operational specification in future refinement of the inventories.

In line with this, the thematic analysis indicated that, while the inventories were evaluated as very comprehensive and, in general, applicable for residents with varying degrees of physical and cognitive ability, adding an additional answer option "not applicable" is advisable. Previous research has suggested that adding a non-applicability option could indeed improve the feasibility of inventories without affecting their content validity [47]; however, it is important to note that adding such an option may also increase nonresponse rates [48]. For practical guidance, stakeholders have also advised the addition of comment fields. Future research may investigate the value of these suggestions.

The test-retest statistics imply adequate agreement over time for items of both inventories in the total sample and for both sub-samples broken down by the level of residents' cognitive functioning. However, although not statistically tested because of the small sample sizes, the results suggest slightly lower test-retest agreement for inventory items within the Belgian sub-sample than within the Dutch sub-sample. This difference might be attributed to the limited number of Belgian participants involved in the development and pilot testing of the inventories. A related explanation may be that the level of content validity may differ between the countries. Although we were not able to compare content validity indices for both countries because of the limited sample size, feedback from stakeholders revealed that certain wording might be less applicable to Belgian participants, suggesting that even minor language differences can influence interpretability [49]. Future research may elucidate whether such differences exist, the extent to which they can explain the variations in terms of agreement over time, and how they could be reduced.

A notable challenge is the inter-rater agreement between caregivers on AIM-C items across all sub-samples, particularly for residents with moderate to severe cognitive decline. Stakeholder feedback suggested that the broad formulation and conceptual overlap of certain themes may have contributed to differing interpretations. According to caregivers, limited awareness of co-workers' actions may have further contributed to limited inter-rater agreement. In addition, caregivers were instructed to report on actions performed for the specific resident during an 'average week.' However, the interpretation of 'average week' may have differed between caregivers, and it is not clear whether they considered different shifts or weeks. This variability in how caregivers interpreted the instructions could have influenced the inter-rater agreement. This challenge, however, is not unique to the AIM-C; other inventories have also revealed limited inter-rater agreement among caregivers for resident observations, especially for residents with moderate to severe cognitive decline [50, 51]. Although the AIM-C inventory cannot be classified as an observer-reported measure, these challenges underline the importance of thoughtful utilization and interpretation of scores that go beyond self-reports. It is therefore recommended to involve at least two caregivers, allowing them to discuss and compare the scores when using the AIM-C. It is also advisable for caregivers to receive training on how to use the inventories effectively. We acknowledge, however, that due to current staffing constraints in many nursing homes [52], this approach may not always be feasible in routine practice and could be considered a best-practice approach.

The AIM-R encountered a similar challenge related to residents' cognitive functioning. Although the test-retest agreement was acceptable across different levels of cognitive functioning, stakeholders noted that some questions were difficult for residents with dementia. However, stakeholders believe that when sufficient time is taken, most themes can be effectively queried, even among residents with cognitive decline. They found the descriptions of themes and examples helpful for usability and understandability. As stakeholders proposed, reducing

the amount of text and administering the inventory over multiple meetings could further enhance usability and reliability for residents with moderate to severe dementia. Exploration of these proposals is warranted in future research. Future research may also explore the feasibility of incorporating observational instruments or proxy-reported scores, particularly for residents who experience challenges expressing themselves.

Finally, we observed measurement reactivity [53], where the use of these inventories appeared to positively influence the awareness and reflection of potential mood-improving behaviors among residents and caregivers. Thematic analysis also revealed that residents enjoyed discussing pleasant activities, whereas caregivers reported increased awareness of mood-improving behaviors that could be integrated into their daily care practices. These unintentional positive impacts highlight the potential of these inventories to contribute to residents' well-being, not only as assessment tools but also by providing a low-threshold opportunity for meaningful conversations, regardless of current psychometric limitations.

Strengths, constraints, and generality

A key strength of this research is its comprehensive, multi-phase, iterative, and mixed-methods approach. The combination of both quantitative data and profound qualitative insights enriched our understanding of selecting response options and allowed evaluation of reliability (in terms of test-retest and inter-rater agreement) and assessment of the inventories' usability in practical settings. Another notable strength of this study is the active involvement of nursing home residents and caregivers in the process of co-creation throughout the inventories' development, testing, and evaluation phases. Their valuable perspectives and experiences have not only contributed to the current refinement of the inventories, but also hold promise for future enhancements. Finally, the development of two inventories – each reflecting a different perspective (i.e., resident and caregiver behaviors) – considers the diverse population of nursing home residents, with varying degrees of physical and cognitive ability, and different preferences regarding mood-improving behaviors. The combination of these two perspectives may strengthen insight into additional and personalized strategies to improve the mood of residents.

However, it is also crucial to recognize certain limitations. First, the utilization of a formative measurement model for the inventories, while suitable for capturing the broad essence of the construct and providing practical guidance, introduces constraints. While formative measurement models are adept at evaluating specific indicators (i.e., themes within the inventories) by examining their distinct components, unlike reflective models, they do not facilitate the computation of a unified composite score for the overarching construct [24]. Consequently, instead of deriving a total score for mood-improving behaviors, each item must be evaluated independently. This approach introduces challenges in establishing the construct and predictive validity of the overall construct, introduces sensitivity to measurement errors, and limits the ability to effectively summarize and compare data. In addition, the internal consistency of the inventories could not be assessed as the items are potentially unrelated.

Furthermore, the inventories encompass numerous specific activities within different themes, thus exploring a wide range of behaviors while limiting the number of items. The development of alternative instruments is necessary for assessing specific behaviors (e.g., watering the plants).

Methodological limitations include the variation in data collection methods during study 1, whereby nursing home residents were interviewed in person while caregivers completed an online survey. Although similar questions were employed, this variation, driven by practical considerations, may have affected the outcomes. Future research could benefit from conducting in-person cognitive interviews with caregivers or comprehensive online response process evaluations [54]. Additionally, exploring inter-rater agreement for the AIM-C through techniques such as cognitive interviewing can provide valuable insights into potential disparities between caregivers' responses.

Moreover, in this study, the AIM-R was completed during a structured face-to-face meeting with one of the interviewers. Although the interviewers were trained to administer the inventory, and each participating nursing home resident was asked the same questions in the same order, the existence of 'interviewer error' (biases or inaccuracies introduced by interviewers during the administration process) [55] cannot be completely ruled out. Notably, the AIM-R has thus not been evaluated for use by residents alone or together with family members or professional caregivers. Future research may explore alternatives, such as residents completing the AIM-R themselves as a self-report measure, or with family members or professional caregivers helping them complete it.

With respect to psychometric properties, it is essential to note that concurrent validity, which assesses the correctness of answers, was not established in this study. While the test-retest agreement was deemed acceptable, the study does not ascertain whether reported behaviors were indeed performed and experienced as pleasant by residents. Ensuring response reliability is a fundamental prerequisite for instrument validity [56], necessitating future validity assessments alongside reliability evaluations.

Finally, the application of these inventories to residents with severe cognitive decline presents challenges. Consequently, our findings may not be generalizable to all nursing home residents. Future measurement strategies such as shorter interviews should take these challenges into account. For the AIM-R, it may be beneficial to test proxy-reported inventories alongside resident self-reports and to consider incorporating visual aids to simplify the administration process.

After refining the inventories, we recommend comparing scores between the AIM-R and AIM-C to explore differences and similarities in actions undertaken by caregivers and nursing home residents. This comparative analysis may yield valuable insights into essential themes for nursing home residents, as observed from different perspectives, and offer practical guidance for enhancing the mood of individual residents.

Conclusions

Overall, the results suggest that the inventories adequately capture mood-improving behaviors among both nursing home residents and caregivers, demonstrate consistent assessments over time, and can potentially increase awareness of potential mood-improving practices in nursing homes. Recommendations are provided to enhance their usability and interpretability for future research and practical application. Given the low levels of inter-rater reliability, we advise caution when using the AIM-C in its current form and recommend involving at least two caregivers to discuss and compare scores. While further research is pending, these inventories show promise for measuring mood-improving behaviors, aiding future studies in uncovering new insights into improving depression care in nursing homes alongside formal treatment options.

Abbreviations

AIM-C	Actions to Improve Mood by Caregivers
AIM-R	Actions to Improve Mood by Residents
CVI	Content Validity Index
GDS	Global Deterioration Scale

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-026-07472-0.

Additional file 1: Actions to Improve Mood by Residents inventory (AIM-R).

Additional file 2: Actions to Improve Mood by Caregivers inventory (AIM-C).

Additional file 3: Agreement statistics for the test-retest agreement of the AIM-R and AIM-C by country and level of cognitive functioning.

Additional file 4: Agreement statistics for the inter-rater agreement of the AIM-C by country and level of cognitive functioning.

Acknowledgements

The authors would like to thank all the contacts at the participating nursing home locations and all participants (residents, caregivers, and other professionals). We also extend our gratitude to the interviewers who contributed to this research but are not co-authors of this paper: Marijke Akkerman, Marissa Bauwens, Merle Borremans – de Bruin, Karin Fleuren, and Ellis Vermeule.

Authors' contributions

IK: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Writing-original draft. RL: Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Supervision; Validation; Writing-review and editing. ID: Conceptualization; Investigation; Methodology; Writing-review and editing. PdV: Conceptualization; Methodology; Writing-review and editing. AP: Conceptualization; Methodology; Writing-review and editing. PV: Conceptualization; Methodology; Writing-review and editing. JvL: Conceptualization; Funding acquisition; Methodology; Supervision; Writing-review and editing. DG: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing-review and editing.

Funding

This study was funded by the Junior Researcher project 2019 – 159 from the Radboud University Medical Center, Radboud Institute for Health Sciences.

Data availability

Research materials, anonymized data, and analysis code are available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki [26] and complied with Dutch and Belgian laws. Ethical approval for the study was obtained from the legally designated ethics committees in both countries (CMO Radboudumc, reference number: 2021–11047 in the Netherlands; CME VUB, reference numbers: EC-2021-277 and EC-2021-432 in Belgium). Prior to participation, all participants were informed about the study and provided written informed consent.

Consent to publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Primary and Community Care, Radboud university medical center, Research Institute for Medical Innovation, Radboudumc Alzheimer Center, Nijmegen, Netherlands
[2]Faculty of Psychology, Open University of the Netherlands, Heerlen, Netherlands
[3]Department of Gerontology and Frailty in Ageing (FRIA) Research Group, Mental Health and Wellbeing (MENT) Research Group, Vrije Universiteit Brussel, Brussels, Belgium
[4]Research Group Health and Care, Artevelde University of Applied Sciences, Ghent, Belgium

Received: 27 September 2024 / Accepted: 8 April 2026
Published online: 16 April 2026

References

1. Van Asch IFM, Nuyen J, Veerbeek MA, Frijters DHM, Achterberg WP, Pot AM. The diagnosis of depression and use of antidepressants in nursing home residents with and without dementia. Int J Geriatr Psychiatry. 2013;28(3):312–8.

2. Fornaro M, Solmi M, Stubbs B, Veronese N, Monaco F, Novello S, et al. Prevalence and correlates of major depressive disorder, bipolar disorder and schizophrenia among nursing home residents without dementia: systematic review and meta-analysis. Br J psychiatry: J mental Sci. 2020;216(1):6–15.
3. Sivertsen H, Bjorklof GH, Engedal K, Selbaek G, Helvik AS. Depression and Quality of Life in Older Persons: A Review. Dement Geriatr Cogn Disord. 2015;40(5–6):311–39.
4. Gilman SE, Sucha E, Kingsbury M, Horton NJ, Murphy JM, Colman I. Depression and mortality in a longitudinal study: 1952–2011. CMAJ. 2017;189(42):E1304–10.
5. Kane KD, Yochim BP, Lichtenberg PA. Depressive symptoms and cognitive impairment predict all-cause mortality in long-term care residents. Psychol Aging. 2010;25(2):446–52.
6. Boyce RD, Hanlon JT, Karp JF, Kloke J, Saleh A, Handler SM. A review of the effectiveness of antidepressant medications for depressed nursing home residents. J Am Med Dir Assoc. 2012;13(4):326–31.
7. Declercq IJN, Leontjevas R, Verboon P, De Vriendt P, Gerritsen DL, van Hooren S. A bayesian network meta-analysis to explore modifying factors in randomized controlled trials: what works for whom to reduce depression in nursing home residents? BMC Geriatr. 2024;24(1):518.
8. Simning A, Simons KV. Treatment of depression in nursing home residents without significant cognitive impairment: a systematic review. Int Psychogeriatr. 2017;29(2):209–26.
9. Nelson JC, Devanand DP. A systematic review and meta-analysis of placebo-controlled antidepressant studies in people with depression and dementia. J Am Geriatr Soc. 2011;59(4):577–85.
10. Brown MN, Lapane KL, Luisi AF. The management of depression in older nursing home residents. J Am Geriatr Soc. 2002.
11. Hartikainen S, Lönnroos E, Louhivuori K. Medication as a risk factor for falls: critical systematic review. journals Gerontol Ser Biol Sci Med Sci. 2007;62(10):1172–81.
12. Burley CV, Burns K, Lam BCP, Brodaty H. Nonpharmacological approaches reduce symptoms of depression in dementia: A systematic review and meta-analysis. Ageing Res Rev. 2022;79:101669.
13. Cuijpers P, Ciharova M, Miguel C, Harrer M, Ebert DD, Brakemeier EL, et al. Psychological treatment of depression in institutional settings: A meta-analytic review. J Affect Disord. 2021;286:340–50.
14. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: A meta-analysis. Clin Psychol Rev. 2007;27(3):318–26.
15. Gramaglia C, Gattoni E, Marangon D, Concina D, Grossini E, Rinaldi C, et al. Non-pharmacological Approaches to Depressed Elderly With No or Mild Cognitive Impairment in Long-Term Care Facilities. A Systematic Review of the Literature. Front Public Health. 2021;9:685860.
16. Noone D, Stott J, Aguirre E, Llanfear K, Spector A. Meta-analysis of psychosocial interventions for people with dementia and anxiety or depression. Aging Ment Health. 2019;23(10):1282–91.
17. Rostamzadeh A, Kahlert A, Kalthegener F, Jessen F. Psychotherapeutic interventions in individuals at risk for Alzheimer's dementia: a systematic review. Alzheimers Res Ther. 2022;14(1):18.
18. Tonga JB, Karlsoeen BB, Arnevik EA, Werheid K, Korsnes MS, Ulstein ID. Challenges With Manual-Based Multimodal Psychotherapy for People With Alzheimer's Disease:A Case Study. Am J Alzheimer's Disease Other Dementias®. 2016;31(4):311–7.
19. Haugan G, Innstrand ST, Moksnes UK. The effect of nurse-patient interaction on anxiety and depression in cognitively intact nursing home patients. J Clin Nurs. 2013;22(15–16):2192–205.
20. Knippenberg IAH, Reijnders JSAM, Gerritsen DL, Leontjevas R. The association between specific activity components and depression in nursing home residents: the importance of the social component. Aging Ment Health. 2021;25(1):118–25.
21. Meeks S, Looney SW. Depressed nursing home residents' activity participation and affect as a function of staff engagement. Behav Ther. 2011;42(1):22–9.
22. Logsdon RG, Teri L. The Pleasant Events Schedule-AD: psychometric properties and relationship to depression and cognition in Alzheimer's disease patients. Gerontologist. 1997;37(1):40–5.
23. Meeks S, Shah SN, Ramsey SK. The Pleasant Events Schedule – Nursing Home Version: A useful tool for behavioral interventions in long-term care. Aging Ment Health. 2009;13(3):445–55.
24. Fried EI. What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. Health Psychol Rev. 2017;11(2):130–4.
25. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
26. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–4.
27. Knippenberg IAH, Leontjevas R, Stoyanov S, Persoon A, Verboon P, Vermeulen H et al. Informal antidepressant strategies for nursing home residents: two group concept mapping studies. Aging Ment Health. 2022:1–12.
28. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. Pub Opin Q. 2007;71(2):287–311.
29. LimeSurvey Project Team and Schmitz C. LimeSurvey: An Open Source survey tool. 2015.
30. Davis LL. Instrument review: Getting the most from a panel of experts. Appl Nurs Res. 1992;5(4):194–7.
31. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol. 2006;3:77–101.
32. ATLAS.ti Scientific Software Development GmbH. ATLAS.ti (version 24) [Qualitative data analysis software]. 2023.
33. IBM Corporation. IBM SPSS Statistics for Windows, Version 22.0 Armonk. NY: IBM Corp; 2013. https://www.ibm.com/spss.
34. Yusoff MSB. ABC of Content Validation and Content Validity Index Calculation. Education in Medicine Journal. 2019.
35. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–5.
36. van Weert JC, van Dulmen AM, Spreeuwenberg PM, Ribbe MW, Bensing JM. Behavioral and mood effects of snoezelen integrated into 24-hour dementia care. J Am Geriatr Soc. 2005;53(1):24–33.
37. Liu KP, Chan CC, Chu MM, Ng TY, Chu LW, Hui FS, et al. Activities of daily living performance in dementia. Acta Neurol Scand. 2007;116(2):91–5.
38. den Ouden M, Bleijlevens MHC, Meijers JMM, Zwakhalen SMG, Braun SM, Tan FES, et al. Daily (In)Activities of Nursing Home Residents in Their Wards: An Observation Study. J Am Med Dir Assoc. 2015;16(11):963–8.
39. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry. 1982;139(9):1136–9.
40. Gwet K. Handbook of Inter-Rater Reliability: the Definitive Guide to Measuring the Extent of Agreement among Raters: Vol 1. Analysis of Categorical Ratings. 5th ed. Gaithersburg, USA: AgreeStat Analytics; 2021.
41. Gwet KL, irrCAC. Computing Chance-Corrected Agreement Coefficients (CAC) (version 1.0) [software, R package]. 2019.
42. R Core team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2022.
43. Quarfoot D, Levine RA. How Robust Are Multirater Interrater Reliability Indices to Changes in Frequency Distribution? Am Stat. 2016;70(4):373–84.
44. Tran D, Dolgun A, Demirhan H. Weighted inter-rater agreement measures for ordinal outcomes. Commun Stat - Simul Comput. 2020;49(4):989–1003.
45. Tran QD, Dolgun A, Demirhan H. The impact of grey zones on the accuracy of agreement measures for ordinal tables. BMC Med Res Methodol. 2021;21(1):70.
46. Altman D. Practical statistics for medical research. London: Chapman and Hall; 1991.
47. Wessels RD, De Witte LP. Reliability and validity of the Dutch version of QUEST 2.0 with users of various types of assistive devices. Disabil Rehabil. 2003;25(6):267–72.
48. Kmetty Z, Stefkovics Á. Assessing the effect of questionnaire design on unit and item-nonresponse: evidence from an online experiment. Int J Soc Res Methodol. 2022;25(5):659–72.
49. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract. 2011;17(2):268–74.
50. Knippenberg I, Leontjevas R, Declercq I, van Lankveld J, De Vriendt P, Gerritsen D. Agreement between nursing home caregivers' observations of residents' depression, well-being, and quality of life. Int Psychogeriatr. 2023:1–3.

51. Dichter MN, Schwab CGG, Meyer G, Bartholomeyczik S, Dortmann O, Halek M. Measuring the quality of life in mild to very severe dementia: Testing the inter-rater and intra-rater reliability of the German version of the QUALIDEM. Int Psychogeriatr. 2014;26(5):825–36.
52. Miller M, Almomani Y, Hopwood P, Haghighi P, Davis A, Littler E, et al. The impact of staffing structures in long-term care homes on the quality of work-life and work outcomes of care-workers: A narrative scoping review. Int J Nurs Stud. 2026;174:105304.
53. French DP, Sutton S. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? Br J Health Psychol. 2010;15(Pt 3):453–68.
54. Wolf MG, Ihm E, Maul A, Taves A. Survey Item Validation. The Routledge Handbook of Research Methods in the Study of Religion. 2019.
55. Davis RE, Couper MP, Janz NK, Caldwell CH, Resnicow K. Interviewer effects in public health surveys. Health Educ Res. 2009;25(1):14–26.
56. Hand DJ. Measurement Theory and Practice: The World Through Quantification. Wiley; 2010.

Publisher's note