CADTH

Homelike Models in Long Term Care: A Review of Clinical Effectiveness, Cost-Effectiveness, and Guidelines

Abbreviations

ADL	Activities of daily living
ADS	Amsterdam Dementia Screening test
ANCOVA	Analysis of Covariance
BNT	Boston Naming Test
C	Comparator
CMAI	Cohen-Mansfield Agitation Inventory
CPS	cognitive performance scale
CSDD	Cornell Scale for Depression
ES	Effect size
GDS	Global deterioration scale
GDS-SF	Geriatric Depression Scale – Short Form
GH	Green house
GIP	behavioural observation scale for intramural psychogeriatry
GIT	Groningen Intelligence Test
HH	Household
IQCODE	Information Questionnaire on Cognitive Decline in the Elderly
ISE	Index of social engagement
MDS	Minimum Data Set
MEDLO	Maastricht Electronic Daily Life Observation
MMS	MDS Mood Scale
MMSE	Mini-mental state examination
MOSES	Multidimensional Observation Scale for Elderly Subjects
N/A	Not applicable
NPI-NH	Neuropsychiatric Inventory – Nursing Home Version
PHQ-9	9-item Patient health questionnaire
QoL	Quality of Life
QoL-AD	Quality of Life-Alzheimer's Disease scale
QUALIDEM	Dementia-specific quality of life
RAI	Resident Assessment Instrument
RBMT	Rivermead Behavioural Memory Test
RISE	Revised Index for Social Engagement
SCU	Special Care Unit
SD	Standard Deviation
TV	television

Context and Policy Issues

Canada's population is rapidly aging. The percentage of Canadians aged 65 years and older increased from 12.1% to 13.2% between 1996 and 2006 – by 2016 the percentage had reached 16.5%.[1] Based on these Census data,[1] it has been forecasted that 263,000 Canadians will require long-term care by 2035.[2]

Long-term care facilities offer accommodations and 24-hour care (e.g., health services, personal care, and meals) for people who are unable to live at home.[3] In Canada, long-term care is under provincial and territorial legislation, and there is wide variation in delivery and cost coverage across jurisdictions.[3]

Traditional long-term care facilities are large institutions with rigid schedules that provide little autonomy for residents, and residents in these settings have reported feeling bored, lonely, and helpless.[4]

Alternative models of long-term care have been developed to overcome the limitations of traditional models, with the aim of improving quality of life, quality of care, and satisfaction

of residents. "Homelike models of care" broadly represent one such alternative model. In terms of the physical environment, homelike care facilities are designed to feel less like medical institutions and more like homes. Although required components are not strictly defined, common elements of homelike care models include: small group living clusters; high staff-to-patient ratios; staff wearing their own clothes instead of uniforms; comfortable, homelike furnishings; and natural elements such as plants, natural sunlight, and access to the outdoors. From a patient perspective, sense of home is determined by psychological factors (i.e., feeling acknowledged, preserving one's habits and values, perceiving autonomy and control, and coping), social factors (interactions and relationships with staff, other residents, family, and friends; pets), and the built environment (i.e., private- and public space, personal belongings, technology, look and feel, outdoors spaces, and location).[5]

The objective of the report is to summarize the evidence regarding the clinical effectiveness, cost-effectiveness, and guidelines regarding homelike models of care for residents of long-term care facilities.

Research Questions

1. What is the clinical effectiveness of homelike models of care (e.g., Eden Alternative and Greenhouse Concepts of Care) for residents of long-term care facilities?

2. What is the cost-effectiveness of homelike models of care for residents of long-term care facilities?

3. What are the evidence-based guidelines for homelike models of care for residents of long-term care facilities?

Key Findings

Evidence of limited quality from nine non-randomized studies suggested there was no difference between homelike and traditional models of care with respect to depression or affective state, and findings were inconsistent for cognitive functioning, quality of life, neuropsychiatric outcomes, social engagement, and functional status. Specifically, results presented in some studies suggested no difference between homelike models of care and traditional models, whereas others showed greater effectiveness and one study showed worse effectiveness for one outcome. No evidence regarding the cost-effectiveness of homelike models of care for long-term care residents or relevant evidence-based guidelines were identified.

Methods

Literature Search Methods

A limited literature search was conducted by an information specialist on key resources including Ovid MEDLINE, the Cochrane Library, University of York Centre for Reviews and Dissemination (CRD) databases, Canadian and major international health technology agencies, as well as a focused Internet search. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were homelike models and long-term care. No filters were applied to limit the retrieval by study type. The search was also limited to English language documents published between January 1, 2009 and May 7, 2019.

Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria

Population	Residents of long-term care facilities (with any condition, e.g., dementia, mental health, general aging)
Intervention / Exposure	Homelike models of care (e.g., Eden Alternative Care, Greenhouse Concepts, Butterfly Model [Dementia Care Matters], low density of residents)
Comparator	Q1,2: Traditional models of care (e.g., nursing homes, general population care, long-term care, residential care, long-term care with high density of residents) Q3: No comparator
Outcomes	Q1: Clinical effectiveness (Physical [e.g., overall survival, mortality, activities of daily living] Psychosocial [e.g., cognitive status/functioning, mood, behavior, social activities, quality of life], nutrition and food service clinical and safety outcomes [e.g., proper nutrition, choking risk, allergies] safety [e.g., adverse events, accidental falls, pressure ulcers] Q2: Cost-effectiveness outcomes (e.g., incremental cost per quality adjusted life year or health benefit) Q3: Evidence-based guideline recommendations for homelike models of care for residents of long-term care facilities
Study Designs	Health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, non-randomized studies, economic evaluations, guidelines

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1 or they were duplicate publications. Due to the high volume of eligible studies, articles published prior to 2014 were excluded. Guidelines with unclear methodology were also excluded.

Critical Appraisal of Individual Studies

The included non-randomized studies were critically appraised using the Risk of Bias in Non-randomized Studies–of Interventions (ROBINS–I) tool.[6] Summary scores were not calculated for the included studies; rather, a review of the strengths and limitations of each included study were described narratively.

Summary of Evidence

Quantity of Research Available

A total of 372 citations were identified in the literature search. Following screening of titles and abstracts, 338 citations were excluded and 34 potentially relevant reports from the electronic search were retrieved for full-text review. No potentially relevant publications were retrieved from the grey literature search for full-text review. Of these potentially relevant articles, 25 publications were excluded for various reasons, and 9 non-randomized studies met the inclusion criteria and were included in this report. Appendix 1 presents the PRISMA[7] flowchart of the study selection.

Summary of Study Characteristics

Details regarding the characteristics of included publications are provided in Appendix 2.

Study Design

Nine non-randomized studies published between 2014 and 2018 were included in this report. These included three non-randomized controlled studies,[8-10] two matched cross-sectional studies,[11,12] two matched prospective cohort studies,[13,14] and two retrospective comparative studies.[15,16]

No eligible health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, cost-effectiveness studies, or evidence-based guidelines were identified for inclusion.

Country of Origin

The non-randomized studies were conducted in the Netherlands,[8-11,13] the US, [12,15,16] and Canada.[14]

Patient Population

Data from 811 unique participants were summarized in this report. Participants were older adults (mean ages ranged from 77.6 years [14] to 87.2 years [15,16]) living in long-term care facilities. Participants had a dementia diagnosis,[8,9,11,13,14] early or middle stage Alzheimer's disease,[14] or lived in general care units.[12]

Interventions and Comparators

All studies examined small-scale homelike long-term care as the intervention of interest. Small-scale homelike long-term care consisted of a small number (seven to 12) of residents living together in a ward[8,9,14] or stand-alone facility,[11,13] where staff were trained to focus on person-centred care[8,9] and residents were encouraged to engage in personal and household daily chores.

The exception was the study by Hermer et al., where the setting was described by authors as homelike, with private bedrooms, a central kitchen and dining area, two small central gathering areas, an inconspicuous nursing station, and access to a patio and garden. The number of residents living together in a unit was greater than the other studies at 16 residents per household.[12]

The studies conducted by de Boer et al., were conducted in designated Green Care Farms,[11,13] the study by Hermer et al., was conducted in Household Model sites,[12] and the two studies by Yoon et al., were carried out with residents living in designated Green Homes.[15,16]

The comparator in all studies consisted of traditional large-scale medicalized facilities of 20 to 30 patients living together in a unit or ward.

Outcomes

The following eight broad outcomes were assessed in the included studies: cognitive functioning (five studies),[8,9,11,14,16] depression or depressive symptoms (six studies),[8-11,14,16] quality of life (three studies),[9,11,14] neuropsychology outcomes (four studies),[9-11,14] social engagement and communication (six studies),[10-15] functional status or activities of daily living (four studies),[10,13,14,16] dementia severity (two studies),[10,13] and affective state (one

study).[12] Within neuropsychiatric outcomes, three studies reported on agitation specifically.[10,11,14]

Summary of Critical Appraisal

Additional details regarding the strengths and limitations of included publications are provided in Appendix 3.

Non-Randomized Studies

Nine non-randomized studies[8-16] were assessed using ROBINS-I[6] and several strengths and limitations were identified.

Taken together, there was a high risk of bias due to confounding in the included evidence. Five studies appeared to have adequately addressed potential confounding through design or analysis.[8-10,15,16] However, two studies reported statistically significances between groups on characteristics that could reasonably be expected to influence study outcomes and were not addressed in the analysis,[12,14] and two studies did not discuss or report potential confounders.[11,13]

Overall there was a low risk of bias due to the selection of participants into the studies.[8-16] Long-term care facilities were selected based on alignment with intervention and comparator eligibility criteria. The individuals who lived in those residences were recruited into the study if they did not meet the study exclusion criteria.

There was a low risk of bias due to the classification of interventions, with interventions and comparators clearly defined across studies.[8-16] Deviations from intended interventions were not reported.

There was a moderate risk of bias due to missing data. Four studies reported that complete data were available for all[12] or nearly all[8,11,13]. participants. However authors of one study reported that complete data were not available for 30% of intervention participants and 51% of comparator participants for unknown reasons, and only complete cases were analyzed.[9] In addition, the authors of two studies based on the same participant sample reported that there were participants with missing data or lost to follow-up which were adequately addressed via analysis, however it is not clear how much data were missing.[15,16] The topic of missing data was not discussed in the study by Lee et al.[14]

There was an unknown risk of bias in the measurement of outcomes for at least two reasons. First, outcome assessors were not blinded to intervention status in any prospective studies, however there does not appear to be a potential for the introduction of bias in most measures. Second, study authors did not consistently report the psychometric properties of the measurement tools used to assess study outcomes, however this does not necessarily mean the validity and reliability of these measures has not been established.

Finally, there was a moderate risk of bias in selection of reported results. In all included studies,[8-16] the reported effect estimates were not likely to have been selected based on results from multiple outcome measurements within outcome domains, multiple analyses, or different subgroups. However, one study assessed functional capacity at baseline, 6-, and 12-months, and yet only baseline scores were reported in the study.[10]

Summary of Findings

Appendix 4 presents a table of the main study findings and authors' conclusions.

Clinical Effectiveness of Homelike Long-Term Care

Cognitive functioning

Three of four studies that assessed cognitive functioning showed no statistically significant differences between the intervention group and the comparator.[13,14,16] The fourth study showed a mix of null findings and findings in favour of homelike long-term care.[8]

There were no statistically significant differences in cognitive functioning between residents in Green Homes and residents in traditional long-term care wards in two studies.[13,16]

For small-scale homelike facilities, findings depended on whether outcomes were assessed by patients or proxies (e.g., nursing personnel, representatives of the resident). When participants or nursing personnel provided responses, there were no significant differences in cognitive function for those in small-scale facilities versus traditional wards.[8] However, when rated by representatives of the participant, results suggested residents of small-scale facilities had significantly better cognitive functioning than those in traditional wards.[8] When taken together, a composite score of the self-reported and representative-reported questionnaires indicated better cognitive functioning for small-scale facility residents at three month follow-up but not six month follow-up as compared with traditional ward residents.[8]

In terms of specific cognitive functioning indicators, there were no significant differences between residents of small-scale versus large-scale living facilities for verbal memory, language, praxis, executive functioning, or visual perception.[8] For visual memory in the form of recognition of pictures, there was a statistically significant difference between groups, whereby small-scale facility residents had a slight improvement from baseline to three-month follow-up while residents of traditional homes had a slight decline over the same period.[8] No significant between groups differences were observed at 6-month follow-up, and no differences at either time point were apparent for visual memory in the form of recognition of faces.[8] In another study, disorientation and cognitive patterns were assessed and no significant differences between residents of small-scale homelike care and those in traditional long-term care facilities were identified.[14]

Depression

Regardless of the intervention or timeframe (baseline up to 12-month follow-up), no significant differences were observed between intervention (i.e., Green Homes, small-scale homelike care) and comparator (i.e., traditional long-term care) groups for depression or depressive symptoms in any of the six studies that reported on this outcome.[8-11,14,16]

Quality of Life

Findings were mixed for quality of life in the three studies that assessed this outcome.[9,11,14] Results from two studies indicated that residents of small-scale homelike care and traditional care facilities did not statistically differ with respect to quality of life from baseline up to 1 year follow-up.[9,14] Quality of life was assessed in multiple ways in a third study that provided mixed findings.[11] In this study, self-reported quality of life did not differ between residents in Green Care Farms and those in traditional facilities. Specific sub-domains of quality of life were also shown not to differ significantly between groups (i.e., negative affect, restless tense behaviour, or social isolation).[11] However, when assessed by proxy-report, residents of Green Care facilities had higher global quality of life scores, as well as

higher scores on the positive affect and social relations sub-domains relative to those in traditional nursing homes.[11]

Neuropsychiatric outcomes

Four studies assessed neuropsychiatric outcomes and findings were a mix of null, favourable, and unfavourable for homelike long-term care, depending on the specific sub-domain.[9-11,14]

There were no statistically significant differences in general neuropsychiatric symptoms between those in Green Homes and those in traditional nursing homes.[11]

For specific subdomains, findings varied by the subdomain. One study that compared small-scale homelike care and traditional wards showed lower frequency and severity of aberrant motor behaviour in the intervention group by 12-month follow-up, but no differences between groups for delusions, hallucinations, agitation/aggression, depression, anxiety, euphoria, apathy, disinhibition, irritability, night-time behaviour, or eating change.[10]

There was a significant group by time effect for anxious behaviour in one study, whereby the intervention group experienced a reduction in anxiety upon moving from a traditional ward to a small-scale facility, whereas anxious behavior did not change in the comparator group who remained in a traditional facility.[9]

Findings were a mix of null, favourable, and unfavourable for agitation. There were no statistically significant differences between those in Green Homes and those in traditional nursing homes for agitation.[11] In contrast, another study showed less agitation in small-scale long-term care residents versus residents of traditional long-term care facilities.[14] Another study on small-scale homelike care showed no significant difference at baseline or 6-month follow-up in physically non-aggressive behaviour, however after 12 months scores were worse for those living in small-scale homelike care facilities compared to traditional long-term care facilities.[10] The same study showed no significant differences between groups for verbally agitated behaviour or physically aggressive behaviour at either time point.[10]

There were no statistically significant differences between those who moved from large- to small-scale facilities compared with those who remained in large-scale facilities for apathy, not social behaviour, insubordinate behaviour, suspicious behaviour, or depressive behaviour.[9]

Social engagement and communication

Findings were a mix of null and favourable (i.e., favouring homelike models) for social engagement and communication outcomes in the six studies that reported these outcomes.[10-15] One study reported significantly greater social engagement at baseline and 6-month, but not 12-month follow up in small-scale homelike facilities compared with traditional facilities.[10] This is in contrast with three studies that reported no statistically significant differences between residents in Green Homes or small-scale homelike care facilities versus those in traditional nursing homes cross-sectionally or up to one year in social engagement and communication. [11,13,14] However, while engaged socially, there was a greater level of social interaction for those living in Green Homes compared with those in traditional nursing homes in the cross-sectional study.[13] Using a growth curve model one study showed that residents in Green Homes had a smaller increase in the probability of not being socially engaged over time compared with residents in traditional homes[15] and

there was no difference between the groups for the rate of increase in social engagement level.[15]

One study on small-scale homelike care showed there were no differences in social withdrawal over the course of a year when compared with traditional facility residents.[14]

When examining cognitive engagement, residents in a household model facility spent less time idle than those living in comparator facilities.[12] There were no significant differences between the groups for time displaying active engagement in activities, time engaged in socioexpressive activities, time staring blankly, or time sleeping during the day.[12]

There were no significant differences over one year in communication patterns or activity pursuits between those living in small-scale homelike care facilities versus those living in traditional long-term care.[14]

Functional Status / Activities of Daily Living

Four studies examined functional status or activities of daily living and reported a mix of null [13,14,16,10] results and those favouring homelike care.[14]

In three studies, there were no statistically significant differences in activities of daily living[10,13,16] or self-care function[14] between residents in Green Homes, Green Care Farms, or small-scale homelike facilities compared with traditional facilities at any time point.

One study examined sub-types of function and showed that those in small-scale facilities had significantly better oral nutrition status after one year, which included chewing and swallowing, as compared with traditional homes.[14] The same study found no difference between groups in physical functioning and structural problems at up to one-year follow-up.[14]

Dementia severity

Two studies showed no statistically significant differences in dementia severity between residents in Green Homes or Green Care Farms versus traditional nursing homes at any time point.[10,13]

Affective state

Affective state was examined in one study, which showed no statistically significant difference between residents in the Household intervention as compared with those living in a legacy facility or in a second control facility that was unaffiliated with the household facility.[12]

Cost-Effectiveness of Homelike Long-Term Care
No relevant evidence regarding the cost-effectiveness of homelike long-term care was identified; therefore, no summary can be provided.

Guidelines
No relevant guidelines regarding homelike long-term care was identified; therefore, no summary can be provided.

Limitations

There are a number of key limitations to note with respect to the current report. First, no evidence from randomized-controlled trials was identified. As a result, conclusions are

based on evidence of limited quality from non-randomized interventions and comparative observational studies limiting certainty in the findings.

Second, eight of the nine included studies exclusively examined patients with a diagnosis of dementia or Alzheimer's disease. Only one cross-sectional study of 68 patients examined participants in general care units with other conditions, and 88% of those patients had a dementia diagnosis. Therefore, the generalizability of the current findings to populations of adults without a diagnosis of dementia or Alzheimer's disease is not known.

Third, the included interventions varied dramatically in the type of homelike setting provided (e.g., the number of residents, the setting, and the staff-to-patient ratio). As such, where findings were inconsistent, the source of the inconsistency was not always clear. Further research seeking to identify the critical elements of homelike models of care are needed to determine which are most effective.

Finally, no relevant cost-effectiveness studies or evidence-based guidelines were identified. This may be related to the lack of supportive evidence available.

Conclusions and Implications for Decision or Policy Making

This report identified evidence on the clinical effectiveness of homelike models of care for residents of long-term care facilities. No evidence-based guidelines or evidence for the cost-effectiveness of homelike models of care for residents of long-term care facilities were identified.

Evidence from nine non-randomized studies provided mixed results regarding the clinical effectiveness of homelike models of care for a variety of outcomes. There were no studies that reported statistically significant differences between homelike models of care and traditional models of care for depression or affective state. For the remaining outcomes of cognitive function, quality of life, neuropsychiatric outcomes, social engagement and communication, and functional status, findings were mixed, with most studies suggesting no difference between groups.

There is a substantial amount of uncertainty regarding the conclusions based on the inconsistency in the evidence, the lack of randomized studies, and the lack of consistency in the models of homelike care included in this report.

A previous CADTH report published in 2010 identified three studies that examined the effectiveness of one of two specific models of homelike care – the Eden Alternative and Greenhouse Concepts of Care — for residents in long-term care.[17] One study showed few statistical differences between the Greenhouse facility and other long-term care facilities for quality of life indicators. Another showed lower levels of boredom and helplessness in the Eden Alternative facility compared to a traditional long-term care facility, and a third study reported no differences between the Eden Alternative and a traditional model for functional status, infection rate, or cost of care.[17] Although a larger number of studies were identified in the current report, there remains substantial uncertainty regarding the effectiveness of homelike models of care given the limitations inherent in the included studies.

Current evidence for the clinical and cost-effectiveness of homelike models of care is limited and additional research is needed to inform clinical practice.

References

1. Statistics Canada. Canada at a glance 2017: population. 2017; https://www150.statcan.gc.ca/n1/pub/12-581-x/2017000/pop-eng.htm. Accessed 2019 Jun 4.
2. Gibbard R. Sizing up the challenge: meeting the demand for long-term care in Canada. Ottawa (ON): Conference Board of Canada; 2017: https://www.cma.ca/sites/default/files/2018-11/9228_Meeting%20the%20Demand%20for%20Long-Term%20Care%20Beds_RPT.pdf. Accessed 2019 Jun 4.
3. Health Canada. Long-term facilities-based care. 2004; https://www.canada.ca/en/health-canada/services/home-continuing-care/long-term-facilities-based-care.html. Accessed 2019 Jun 4.
4. Agency for Healthcare Research and Quality. Home-like, self-directed environment provides superior quality of life than traditional nursing homes and assisted living facilities. 2014; https://innovations.ahrq.gov/node/4392. Accessed 2019 Jun 4.
5. Rijnaard MD. The factors influencing sense of home in nursing homes: a systematic review from the perspective of residents. *J Aging Res.* 2016(2016):6143645.
6. Sterne JAC, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355(i4919).
7. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34.
8. Kok JS, van Heuvelen MJ, Berg IJ, Scherder EJ. Small scale homelike special care units and traditional special care units: effects on cognition in dementia; a longitudinal controlled intervention study. *BMC Geriatr.* 2016;16:47.
9. Kok JS, Nielen MMA, Scherder EJA. Quality of life in small-scaled homelike nursing homes: an 8-month controlled trial. *Health Qual Life Outcomes.* 2018;16(1):38.
10. Verbeek H, Zwakhalen SM, van Rossum E, Ambergen T, Kempen GI, Hamers JP. Effects of small-scale, home-like facilities in dementia care on residents' behavior, and use of physical restraints and psychotropic drugs: a quasi-experimental study. *Int Psychogeriatr.* 2014;26(4):657-668.
11. de Boer B, Hamers JPH, Zwakhalen SMG, Tan FES, Verbeek H. Quality of care and quality of life of people with dementia living at green care farms: a cross-sectional study. *BMC Geriatr.* 2017;17(1):155.
12. Hermer L, Bryant NS, Pucciarello M, Mlynarczyk C, Zhong B. Does comprehensive culture change adoption via the household model enhance nursing home residents' psychosocial well-being? *Innov Aging.* 2017;1(2):igx033.
13. de Boer B, Hamers JP, Zwakhalen SM, Tan FE, Beerens HC, Verbeek H. Green care farms as innovative nursing homes, promoting activities and social interaction for people with dementia. *J Am Med Dir Assoc.* 2017;18(1):40-46.
14. Lee SY, Chaudhury H, Hung L. Effects of physical environment on health and behaviors of residents with dementia in long-term care facilities: a longitudinal study. *Res Gerontol Nurs.* 2016;9(2):81-91.
15. Yoon JY, Brown RL, Bowers BJ, Sharkey SS, Horn SD. Longitudinal psychological outcomes of the small-scale nursing home model: a latent growth curve zero-inflated Poisson model. *Int Psychogeriatr.* 2015;27(6):1009-1016.
16. Yoon JY, Brown RL, Bowers BJ, Sharkey SS, Horn SD. The effects of the Green House nursing home model on ADL function trajectory: a retrospective longitudinal study. *Int J Nurs Stud.* 2016;53:238-247.
17. Eden alternative and green house concept of care: review of clinical effectiveness, cost-effectiveness, and guidelines. *(CADTH Rapid response: summary with critical appraisal).* Ottawa (ON): CADTH; 2010: https://www.cadth.ca/sites/default/files/pdf/L0166_Eden_Alternative_Concept_final.pdf Accessed 2019 Jun 4.

Appendix 1: Selection of Included Studies

```
┌─────────────────────────────────┐
│ 372 citations identified from   │
│ electronic literature search    │
│ and screened                    │
└─────────────────────────────────┘
           │
           │────────────────►  ┌─────────────────────────────┐
           │                   │ 338 citations excluded      │
           │                   └─────────────────────────────┘
           ▼
┌─────────────────────────────────┐
│ 34 potentially relevant         │
│ articles retrieved for scrutiny │
│ (full text, if available)       │
└─────────────────────────────────┘

┌──────────────────────┐           │
│ 0 potentially        │           │
│ relevant reports     │──────────►│
│ retrieved from       │           │
│ other sources (grey  │           │
│ literature, hand     │           │
│ search)              │           │
└──────────────────────┘           ▼
                   ┌─────────────────────────────────┐
                   │ 34 potentially relevant reports │
                   └─────────────────────────────────┘
                              │
                              │       ┌──────────────────────────────────────────┐
                              │       │ 25 reports excluded:                     │
                              │──────►│ -irrelevant population (1)               │
                              │       │ -irrelevant outcomes (2)                 │
                              │       │ -other (design, review articles,         │
                              │       │  editorials) (7)                         │
                              │       │ -published before 2014 (15)              │
                              ▼       └──────────────────────────────────────────┘
                   ┌─────────────────────────────────┐
                   │ 9 reports included in review    │
                   └─────────────────────────────────┘
```

Appendix 2: Characteristics of Included Publications

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
Kok, 2018[9] The Netherlands	Non-randomized controlled trial Period of data collection not reported	N = 84 with complete data Inclusion criteria: Residents with dementia living in one of two special care units of one mental health care institute Participant characteristics: Residents did not differ significantly on age, gender, education, global cognitive function, or mood. Intervention vs. comparator: Location: rural vs. small city Mean age: 83.4 vs. 82.8 Gender: 69% vs. 72% female; 31% vs. 28% male All patients had moderate to severe dementia; Types of Dementia: NOS, 23% vs. 38%; Alzheimers, 31% vs. 19%; Mixed dementia, 8% vs. 16%; Vascular, 7% vs. 12%; Lewy body, 1% vs. 2% Frontotemporal, 0% vs. 6%; Other: 5% vs. 2%	Intervention: One traditional SCU of a nursing home; 20 to 30 residents / ward The group moved after 2 months to a small-scale homelike SCU; 7 to 8 residents per ward in a large building; staff were trained to focus on person-centred care Note: the move to small-scale homelike care was previously planned by the organization and the researchers capitalized on the opportunity to study the potential effects of the move. This was not an investigator-initiated intervention. No change in resident-to-staff ratio Comparator: One traditional SCU of a nursing home within the same the same mental health care institute; 20 to 30 residents / ward	Outcomes: **Cognitive functioning** Standardized MMSE; Collected from residents by trained research assistants; Higher scores reflect better cognitive functioning; Maximum score = 30 **Mood (Depression)** GDS-SF (15-items); Self-report; Dichotomously scored (yes/no) Scores ≥5 indicate presence of depression **QoL** QUALIDEM (40 items); Assessed by trained primary nurses; Study authors reported sufficient reliability and validity (scores not reported) Scores rated on a 4-point scale; Higher scores indicate higher QoL **Neuropsychiatry** 6 subscales of the Behavioral Observation Scale for Intramural Geriatric Psychiatry: Not social behaviour (maximum score 24) Apathy (maximum score 18) Insubordinate behaviour (maximum score 15) Suspicious behaviour (maximum score 21) Depressive behaviour (maximum score 18) Anxious behaviour (maximum score 18) Scores rated from 1 to 4 Higher scores = worse outcomes Follow-up: Assessed at baseline, 3 months post-move and 6 months post-move

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
de Boer, 2017[13] The Netherlands	Matched 2-arm prospective cohort study Period of data collection not reported	N = 56 Inclusion criteria: Dementia diagnosis Living in a non-profit, collectively funded nursing home Participant characteristics: Baseline mean age = 82 years vs. 83 years; 68% women, 32% men vs. 62% women, 38% men	Intervention *Green Care Farms* Combined agriculture and care activities. Stand-alone, small-scale facilities; groups of 8 residents per home on the farm; residents encouraged to help with agricultural tasks and daily living tasks; patient-centred care; autonomy *Regular small-scale living facility* (no relevant comparative data reported) Comparator *Traditional nursing home wards* Groups of ≥20 residents / ward; differentiated tasks for caregivers; daily life mainly determined by routines and rules of the organization.	Outcomes: **Social engagement** *Engaging in social/communication related activity* Assessed by ecological momentary assessments using MEDLO-tool; observed over 2 weeks; percentage of "yes" (i.e., activity performed) responses were calculated; Reported to be valid and reliable by study authors *Social interaction during the activity being observed* Rated by observers as 0 = no social interaction or attempted interaction without response; 1 = yes, social interaction with one or more persons; Assessed with MEDLO-tool **Cognitive functioning** Assessed with Standardized MMSE; scores ranged from 0 to 30; higher scores indicate better cognition; Psychometric properties not reported by study authors **ADL** Assessed with Barthel index Total scores range from 0 to 20, higher scores indicate less dependence; Psychometric properties not reported by study authors **Dementia severity** Assessed with GDS; scores ranged from 1 (normal) to 7 (highly severe); Psychometric properties not reported by study authors Follow-up: Assessed at baseline and 6 months
de Boer, 2017[11] The Netherlands	Matched 2-arm cross-sectional study Data were	N = 115 (includes participants in de Boer 2017 cohort study plus additional regular small-scale facility comparator)	Intervention *Green Care Farms* Combined agriculture and care activities. Stand-alone, small-scale	Outcomes: **QoL** Assessed using the 13-item QoL-AD; 4-point Likert scale by proxy- and self-report, scores range from

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
	collected between April and October 2014	Inclusion criteria: Residents in long-term care with a formal dementia diagnosis Participant characteristics: Mean age = 83.8 years (range 59-97 years) Gender = 75% female, 25% male; Barthel index = 9.7 Standardized-MMSE = 8.4	facilities; groups of 8 residents per home on the farm; residents encouraged to help with agricultural tasks and daily living tasks; patient centred care; autonomy *Regular small-scale living facility* (no relevant data reported) Comparator *Traditional nursing home wards* Groups of ≥20 residents / ward; differentiated tasks for caregivers; daily life mainly determined by routines and rules of the organization.	13 to 52, higher scores indicate better QoL; differences of ≥3 points considered meaningful; Authors reported acceptable psychometric properties (data not reported) Assessed by 7-day recall, 37-item QUALIDEM; 4-point Likert scale Authors reported acceptable psychometric properties (data not reported) **Social Engagement** Assessed using the 6-item RISE 7-day recall; scores range from 0 to 6 (minimum to maximum social engagement) Authors reported good reliability. Validity not reported. **Neuropsychiatric symptoms** Assessed by caregivers using 1-month recall NPI-NH questionnaire. Assesses 12 symptoms Presence of symptoms yes/no; Frequency of symptoms rated from 1 to 4 (rarely to very often); Severity of symptoms rated from 1 to 3 (mild to severe) **Agitation** Assessed using the 29-item, 2-week recall CMAI using a 7-point Likert scale; rated from 1 to 7 (never to several times per hour); scores range from 29 to 203; higher scores indicate more agitated behaviour Authors reported acceptable psychometric properties (values not reported)

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
				Depressive Symptoms Assessed using the 19-item, 7-day recall CSDD; rated from 0 to 2 (absent to severe); Summed scores range from 0 to 38; higher scores indicate more depressive symptoms Authors reported the scale was valid and reliable (values not reported) Follow-up: N/A
Hermer, 2017[12] US	Matched cross-sectional study Data collected summer 2016	N = 68 Inclusion criteria: Residents on general care units Exclusion criteria: Advanced dementia; Severe aphasia or otherwise unable to communicate; on hospice; unable to be matched with another resident due to combination of primary diagnoses, depressive symptoms, and cognitive status Participant characteristics: Significant difference between groups for: *Race*: household model, 100% Caucasian; institutional legacy, 92.3% Caucasian; second institution, 70.5% Caucasian; *PHQ-9 depression score*: household model, 1; institutional legacy, 2.6 No difference between	Intervention: *HH* 2 x 16-resident households separated by hallways; each site had a central kitchen and dining area; private bedrooms + 1 double suite, 2 small central gathering areas, inconspicuous nursing station, locked patio with vegetable and flower garden; most paid privately for rooms and care Comparator: *C1: Legacy control* Older nursing home owned by same organization; certified Eden facility, 90% of staff received person-centred care training; 3 neighbourhoods with 20-21 residents each; semi-private rooms; conspicuous central nursing station; shared large common area for both neighbourhoods with natural light, a piano, and TV, a patio and garden; most paid privately for	Outcomes: **Affective State** Assessed by direct observation using the Observed Emotion Rating Scale; Authors reported the scale as validated with elderly populations and used extensively for older adults with dementia **Cognitive Engagement** Assessed by direct observation using the Menorah Park Engagement Scale; Authors reported the scale as validated Follow-up: N/A

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
		groups for: Mean age (ranged from 85.8 years to 86.9 years); % female / male gender (ranged from 58.8%/41.2% to 84%/16%); Dementia diagnosis (ranged from 88% to 88.4%); Cognitive functioning score (ranged from 6.8 to 6.9); Depression diagnosis or symptoms (ranged from 60% to 68%)	rooms and care *C2: Traditional control* Institutional layout 2 neighbourhoods with 30 residents each; semi-private rooms located along a long hallway; dining room also served as common area, TV area, activities area; no access to the outdoor space; rooms paid via a government program for those with low income	
Kok, 2016[8] The Netherlands [same participants as Kok, 2018]	Non-randomized controlled trial Time frame for recruitment and data collection not reported Intervention participants were recruited and assessed 1 year prior to comparator	N = 115 Inclusion criteria: Residents with dementia living in one of two special care units of one mental health care institute Participant characteristics: No difference between groups for demographics, mood, or cognition (values not reported) All patients had moderate to severe dementia; Types of Dementia: NOS, 23% vs. 38%; Alzheimer's, 31% vs. 19%; Mixed dementia, 8% vs. 16%; Vascular, 7% vs. 12%; Lewy body, 1% vs. 2% Frontotemporal, 0% vs. 6%; Other: 5% vs. 2%	Intervention: Small-scale homelike SCU Patients in a traditional SCU of a nursing home with 20 to 30 residents / ward moved after 2 months to a small-scale homelike SCU; 7 to 8 residents per ward in a large building; staff were trained to focus on person-centred care; patients were more engaged in daily chores, encouraged to do their own cooking and washing No change in resident-to-staff ratio Comparator: One traditional SCU of a nursing home within the same the same mental health care institute; 20 to 30 residents / ward; meals from institution kitchen, no participation in household activities	Outcomes: **Mood (Depression)** GDS-SF (15-items); Self-report; Dichotomously scored (yes/no) Scores ≥5 indicate presence of depression **Cognitive functioning** Standardized MMSE (19 items); Collected from residents by trained research assistants; Maximum score = 30 Higher scores reflect more correct answers and better cognitive functioning; <10 - severe dementia 10 to 19 – moderate dementia, 20 to 26 – mild dementia >26 – normal cognitive function **Proxy reported cognitive function** *Proxy: nursing personnel* GIP (12 items); Authors reported valid for patients in nursing homes *Proxy: representatives* IQCODE Scored from –3 to 3 (much worse to much better) *Verbal memory*

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
				ADS Test - Eight Word Verbal Memory Test Assesses immediate recall (maximum score 40), delayed recall (maximum score 8) and recognition (maximum score 8); Higher scores indicate more correct responses Authors indicated test is validated for older people with dementia *Visual memory* RBMT subtests Assess visual memory and recognition of pictures (maximum score 10) and recognition of faces (maximum score 5); higher scores indicate more correct responses *Language* BNT-short (29-items); Maximum score 29 Higher scores indicate more correct responses *Praxis* van Heugten diagnostic test for apraxia (maximum score 90) Higher scores are more favourable *Executive functioning* Trail Making Test A and B (maximum score 25 for each of A and B), The Category Fluency Task from the GIT (maximum possible score not reported) Clock Drawing test (maximum score 15) *Visual perception* GIT subtask – Incomplete Drawings (maximum score 20) Higher scores indicate more correct responses

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
				Follow-up: Baseline (before relocation), 3-, and 6-months after relocation
Lee, 2016[14] Canada	Matched prospective cohort study Data were collected between March 2012 and April 2013	N = 12 Inclusion criteria: Aged ≥60 years; Early or middle stage of Alzheimer's or a related dementia; Able to walk without assistive device Exclusion: Bed-bound or staying in private rooms during daytime Participant characteristics: Mean ages = 82.9 years (small-scale) and 77.6 years (large-scale); Length of stay = 28.9 months (small-scale) vs. 29 months (large scale)	Intervention: A small-scale homelike long-term care facility for people with dementia; 12 residents per unit; single bedrooms Comparator: A traditional large-scale long-term care facility; 30 residents per unit; mix of single/semi-private bedrooms	Outcomes: **Functioning, Cognitive Status and Psychosocial Behaviours** assessed with MOSES proxy report questionnaire (completed by care aids familiar with resident); Good interrater reliability; validity not reported by authors; items rated on a scale from 1 (negative to great degree) to 4- or 5 (positive to great degree). **Residents' functions, needs, and latent risks** Assessed with MDS 2.0 proxy report questionnaire by RN at each facility 12 domains that inform individualized care planning and monitoring: cognitive patterns, communication patterns, mood/behaviour patterns, psychosocial wellbeing, physical functioning and structural problems, continence, disease diagnoses, health conditions, oral status, skin condition, activity pursuit patterns, medications Authors did not report on psychometric properties Follow-up: Assessed at 4-, 8-, and 12-month follow up
Yoon, 2016[16] US	Retrospective comparative study Secondary analysis of data from the Study of Changes in ADL Assistant Levels in	N = 242 Inclusion: Resided in a selected home for ≥6 months Participant characteristics: No significant difference between participants for	Intervention: GH Multi-faceted intervention Homelike environment, Highly trained and empowered staff, Individualized care that respects residents' choices, encourages self-care and independence,	Outcomes: **ADL function** Assessed with the 7-item ADL long-form scale. Items rated from 0 to 4. Total score ranges from 0 to 28 (complete independence to total dependence) **Depressive mood**

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
	Traditional Nursing Homes and The Green House Project sites Data originally collected between June 2014 and September 2009	age (87.2 vs. 85.8 years), gender (both 73% women), dementia (55.9% vs. 50.0%), or proportions of Medicaid payers (value not reported)	Private rooms and bathrooms, Communal meals in a dining area, Involvement of residents in unit routines (e.g., laundry) Small number of beds compared with traditional homes (not specified) Comparator: Traditional nursing home No differences between homes in ownership, organization, location	Assessed with MMS; Scores range from 0 to 8 (higher scores = more depressed mood) **Cognitive function** Assessed with CPS; Scores range from 0 to 6 (intact vs. very severe impairment) Psychometrics not reported by authors Follow-up: Assessed at admission, 3, 6, 9, 12, and 18 months
Yoon, 2015[15] US	Retrospective comparative study Secondary analysis of data from the Study of Changes in ADL Assistant Levels in Traditional Nursing Homes and The Green House Project sites Data originally collected between June 2014 and September 2009	N = 242 Inclusion: Resided in a selected home for ≥6 months Exclusion: residents were admitted for short-term rehab or hospice upon moving into home Participant characteristics: No significant difference between participants for age (87.2 vs. 85.8 years), gender (both 73% women), dementia (55.9% vs. 50.0%), ADLs, cognitive function, or proportions of Medicaid payers (value not reported)	Intervention: GH – a small-scale homelike long-term care facility for people with dementia; 12 residents per unit; single bedrooms Multi-faceted intervention Homelike environment, Highly trained and empowered staff, Individualized care that respects residents' choices, encourages self-care and independence, Private rooms and bathrooms, Communal meals in a dining area, Involvement of residents in unit routines (e.g., laundry) Small number of beds compared with traditional homes (not specified) Comparator: A traditional large-scale long-term care facility; 30 residents per unit; mix of single/semi-private bedrooms No differences between	Outcomes: **Social engagement** Assessed using the 6-item ISE; total scores range from 0 to 6, higher scores indicate greater social engagement. Authors reported good internal consistency and inter-rater reliability (Other outcomes reported by Yoon 2016 not extracted here) Follow-up: 6-, 12-, and 18-month assessments were analyzed

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
			homes in ownership, organization, location	
Verbeek, 2014[10] the Netherlands	Non-randomized controlled study Matched on cognitive and functional status	N = 259 Inclusion criteria: Residents with dementia diagnosis living in long-term institutional nursing care for ≥1 month Participant characteristics: Residents differed on (1) Living arrangement prior to admission. I.e., intervention group were more likely living in another institution/ unknown (70%) vs. living at home (30%) and comparator group were more likely living at home (59%) vs. in another institution/ unknown (41%), $P < 0.001$, and (2) length of stay. I.e., was shorter in the intervention group (mean = 15.7 months) vs. the comparator group (mean = 24.4 months), $P = 0.047$; Groups were similar for other characteristics (i.e., age, sex, dementia type, cognition, function, and comorbidities)	Intervention: Small-scale long-term care Selection criteria: ≤8 residents / house; Daily household duties centred on ADLs; Staff performed medical and personal care, household chores, and organized activities; Small consistent team of staff; Daily life largely determined by residents, family caregivers, staff; Physical environment resembled a house Comparator: Traditional psychogeriatric wards in nursing homes Selection criteria: ≥20 residents / ward; Staff had specialized tasks with medical and personal care focus; Daily life dictated by the nursing home; Located near the small-scale facilities	Outcomes: **Neuropsychiatric symptoms** Assessed by caregivers using NPI-NH questionnaire. Assesses delusions, hallucinations, aggression/agitation, depression, anxiety, euphoria, apathy, disinhibition, irritability, aberrant motor behavior, sleep, and eating disturbances Presence of symptoms yes/no; Frequency of symptoms rated from 1 to 4 (rarely to very often); Severity of symptoms rated from 1 to 3 (mild to severe) Total score = Frequency x severity; ranges from 0 to 12 for each symptom; higher scores indicate more frequent and severe symptoms *Agitation* assessed using CMAI. Frequency of 29 agitated behaviours rated from 1 to 7 (never to several times per hour) during previous 2 weeks. Types of agitation assessed: physically non-aggressive (total score ranges from 7 to 49), physically aggressive (total score ranges from 8 to 56), and verbally agitated (total score ranges from 8 to 56) **Depressive symptoms** Assessed by 19-item CSDD. Symptom severity ranges from 0 to 2 (absent to severe) Total score ranges from 0 to 38; higher scores indicate more depressive symptoms;

Table 2: Characteristics of Included Non-Randomized Primary Clinical Studies

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
				Total score of ≥10 indicates major depressive symptoms **Social engagement** Assessed with the 6-item ISE subscale from RAI-MDS. Scored as 0 (absent) or 1 (present). Types of social engagement assessed: social involvement, autonomy Total scores ranged from 0 to 6, with higher scores indicating greater social engagement **Dementia severity** Assessed with GDS, scores range from 1 to 7. Higher scores indicate more severe dementia **Functional status / ADL-capacity** Assessed with ADL-hierarchy subscale from RAI-MDS -Data not presented Follow-up: 6- and 12-months post-baseline assessments

ADL = activities of daily living; ADS = Amsterdam Dementia Screening test; BNT = Boston Naming Test; C = comparator; CMAI = Cohen-Mansfield Agitation Inventory; CPS = cognitive performance scale; CSDD = Cornell Scale for Depression; GDS = global deterioration scale; GDS-SF = Geriatric Depression Scale – Short Form; GIP = behavioural observation scale for intramural psychogeriatry; GH = Green Home; GIT = Groningen Intelligence Test; HH = household; ISE = index of social engagement; IQCODE = The Information Questionnaire on Cognitive Decline in the Elderly; MDS = Minimum Data Set; MEDLO = Maastricht Electronic Daily Life Observation; MMSE = mini-mental state examination; MMS = MDS mood scale; MOSES = Multidimensional Observation Scale for Elderly Subjects; N/A = not applicable; NPI-NH = Neuropsychiatric Inventory – Nursing Home Version; PHQ-9 = 9-item Patient Health Questionnaire; QoL-AD = quality of life – Alzheimer's Disease scale; QoL = quality of life; QUALIDEM = Dementia-specific quality of life; RAI = Resident Assessment Instrument; RBMT = Rivermead Behavioural Memory Test; RISE = revised Index for Social Engagement; SCU = special care unit; TV = television.

Appendix 3: Critical Appraisal of Included Publications

Table 3: Strengths and Limitations of Non-Randomized Studies using ROBINS-I[6]

Strengths	Limitations
Kok, 2018[9]	
The potential for confounding of the effect of the intervention was adequately controlled. Selection of intervention participants into the study was not based on participant characteristics observed after the start of intervention. There was a low risk of bias due to classification of interventions. Intervention groups were clearly defined a priori; classification of intervention status could not have been affected by knowledge of the outcomes. No deviations from interventions were reported. Quality of life was assessed by trained nurses using the QUALIDEM. Authors reported sufficient reliability and validity (values not reported); methods of outcome assessment were the same across groups. There was low risk of bias in selection of the reported result. It is unlikely multiple measurements were made within the outcome domain or multiple analyses of the intervention-outcome relationship. Subgroups were not calculated.	Complete data were unavailable for 30% of intervention participants and 51% of comparator participants; the proportion missing was imbalanced and the reasons missing were largely unknown (i.e., no data available); only complete cases (i.e. data at baseline and 2 follow-up points) were analyzed. Outcome assessors were not blinded to intervention status and there was a possibility of bias in the measurement method.
de Boer[13]	
The intervention group did not differ significantly from comparator groups on potential confounders examined at baseline and subsequent analyses controlled for potential confounders that were assessed. The selection of intervention participants was based on residence in the selected home and not based on participant characteristics observed after the start of intervention. Comparator group participants were matched to intervention group participants. The groups were clearly defined, information used to define groups was recorded at the start of the study, and classification could not have been affected by knowledge of the outcomes. Complete data were available for nearly all (87%) participants. Outcome assessments were the same across intervention groups. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within outcome domains, multiple analyses, or different subgroups.	There was no evidence that the outcome assessors were blinded to intervention received by study participants.

Table 3: Strengths and Limitations of Non-Randomized Studies using ROBINS-I[6]

Strengths	Limitations
de Boer[11]	
The intervention group did not differ significantly from comparator groups on potential confounders examined at baseline and subsequent analyses controlled for potential confounders that were assessed. The selection of intervention participants was based on residence in the selected home and not based on participant characteristics observed after the start of intervention. Comparator group participants were matched to intervention group participants. The groups were clearly defined, information used to define groups was recorded at the start of the study, and classification could not have been affected by knowledge of the outcomes. Complete data were available for all participants. Authors reported good psychometric properties for outcome measures. Methods of measurement were comparable across groups. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within outcome domains, multiple analyses, or different subgroups.	It is possible all potential confounders were not measured. E.g., it is not known if socio-economic status of residents differed between the groups. Outcome assessors were probably not blinded to intervention received by study participants.
Hermer, 2017[12]	
The selection of intervention participants was based on residence in the selected home and not based on participant characteristics observed after the start of intervention. Comparator group participants were matched to intervention group participants. The groups were clearly defined, information used to define groups was recorded at the start of the study, and classification could not have been affected by knowledge of the outcomes. There were no missing outcome data. Outcome assessors were blinded to the study purpose. Methods of outcome assessment were the same for all groups. There were no apparent systematic errors in outcome assessment related to intervention received. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within the outcome domain, multiple analyses of the intervention-outcome relationship, or different subgroups.	Participants in each of the two control conditions were matched to intervention group participants on presence or absence of a dementia diagnosis, cognitive status, presence or absence of depressive symptoms and diagnosis, and to the extent possible: primary diagnosis, age within 5 years, gender, and race. Significant differences between groups were observed for race and depression that were not accounted for in the analysis. Additionally, most residents of the household model and legacy comparator paid for their care and accommodation privately, while most residents in the traditional facility comparator paid for their care by government support offered to people with limited resources.

Table 3: Strengths and Limitations of Non-Randomized Studies using ROBINS-I[6]

Strengths	Limitations
Kok, 2016[8]	
The potential for confounding of the effect of the intervention was adequately controlled Selection of intervention participants into the study was not based on their residence in one of the selected facilities, and not on participant characteristics observed after the start of intervention. Intervention groups were clearly defined a priori; classification of intervention status could not have been affected by knowledge of the outcome. No deviations from interventions were reported. Outcome data were available for 87.01% and 70.59% of intervention and comparator participants, respectively. Reasons for missing were similar across groups. There were no differences between those analyzed and those who were lost to follow-up on study outcomes. Objective cognitive tests were unlikely to have been influenced by knowledge of the intervention received; methods of outcome assessment were comparable across groups. Reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within the outcome domain, multiple analyses of the intervention outcome relationship, or different subgroups.	Participants lost to follow-up were not included in analyses. Proportions of lost-to follow-up were not similar across groups. More participants were lost to follow-up in the comparator group. There was a large amount of missing data for individual outcomes at follow-up time points. It is possible certain outcomes (e.g., the proxy report outcomes) could have been influenced by knowledge of the intervention received; outcome assessors were aware of the intervention received by study participants; psychometric properties were not reported for most outcomes.
Lee, 2016[14]	
Intervention groups were clearly defined. Classification of intervention status could not have been affected by knowledge of the outcome or risk of the outcome. There is no evidence for any deviations from the intended interventions. No co-interventions were reported. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within the outcome domain, multiple analyses, or different subgroups.	Potential confounders were not adequately described or considered. One homelike care residence and one traditional residence were selected purposively. Any number of characteristics could have influenced study findings. Participants were selected into the study based on their residence in one of the purposively selected long-term care facilities. It is possible that any number of participant characteristics contributed to their eventual residence in a given facility and that these characteristics could have differed between the intervention and comparator facility. Participant characteristics and distributions between groups were not reported. It is unclear if outcome data were available for all participants. Missing data were not discussed. Other: It is not likely that those asked to participate were representative of the entire population from which they were drawn. Ten nursing homes were randomly selected within a regional

Table 3: Strengths and Limitations of Non-Randomized Studies using ROBINS-I[6]

Strengths	Limitations
	sampling frame. From those, the home that most resembled homelike care and the home that most resembled traditional institutional care were selected purposively. Authors did not report the psychometric properties of outcome measures.
Yoon, 2016[16]	
The potential for confounding of the effect of the intervention was adequately controlled. Selection of intervention participants into the study was not based on their residence in one of the selected facilities, and not on participant characteristics observed after the start of intervention. Intervention groups were clearly defined. Classification of intervention status could not have been affected by knowledge of the outcome or risk of the outcome. There do not appear to have been any deviations from the intended interventions. No co-interventions were reported. The impact of missing data due to drop-out or other missing outcome values was considered in the analysis. Given the retrospective nature of this study, it was not possible for outcome measures to have been influenced by knowledge of the intervention received; methods of outcome assessment were comparable across groups. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within the outcome domain, multiple analyses, or different subgroups.	There were patients lost to follow-up and those with missing outcome data, however values and distributions (e.g., n, percentage) were not reported. Psychometric properties of outcome measures were not reported by study authors. Therefore, it is unclear if they were accurate and reliable.
Yoon, 2015[15]	
The potential for confounding of the effect of the intervention was adequately controlled. Selection of intervention participants into the study was not based on their residence in one of the selected facilities, and not on participant characteristics observed after the start of intervention. Intervention groups were clearly defined. Classification of intervention status could not have been affected by knowledge of the outcome or risk of the outcome. There do not appear to have been any deviations from the	There were patients lost to follow-up and those with missing outcome data, however values and distributions (e.g., n, percentage) were not reported at all time points. Drop-out was 62% at 18-month follow-up. Characteristics of patients lost to follow-up were not described. The amount of missing data was not reported.

Table 3: Strengths and Limitations of Non-Randomized Studies using ROBINS-I[6]

Strengths	Limitations
intended interventions. No co-interventions were reported. The impact of missing data due to drop-out or other missing outcome values was considered in the analysis. Psychometric properties of outcome measures were not reported by study authors. Therefore, it is unclear if they were accurate and reliable. Given the retrospective nature of this study, it was not possible for outcome measures to have been influenced by knowledge of the intervention received; methods of outcome assessment were comparable across groups. The reported effect estimates were not likely to have been selected on the basis of results from multiple outcome measurements within the outcome domain, multiple analyses, or different subgroups.	
Verbeek, 2014[10]	
The potential for confounding of the effect of the intervention was adequately controlled. Selection of intervention participants into the study was not based on their residence in one of the selected facilities, and not on participant characteristics observed after the start of intervention. It is possible participants were selected into the different facilities on the basis of certain characteristics. However, the few differences were corrected for in the analysis. Thus, it is unlikely bias was introduced in the selection of participants into the study. Interventions were clearly defined, information used to define intervention groups was defined a priori, and intervention group status could not have been affected by knowledge of the outcomes. All participants who provided baseline data were included in the analyses. Missing data at each time point appeared balanced across groups. Outcome measures were consistent across study groups. The reported effect estimates were not likely to have been selected on the basis of the results from multiple outcome measurements within the outcome domain, multiple analyses of the intervention-outcome relationship, or different subgroups.	There was a moderate risk of bias due to potential deviations from intended interventions. Authors used a questionnaire to assess adherence to principles of small-scale homelike care (scores ranged from 18 to 90). The median score for small-scale homelike facilities was 66 and for traditional facilities it was 40. The range of scores was not reported, and therefore it is unclear how well the care received by the participants in each group adhered to the respective eligibility criteria. There was a moderate risk of bias due the measurement of outcomes. Most outcomes were assessed by staff in the respective institutions. Staff satisfaction was not assessed; however, small-scale homelike care facilities are intended to improve staff satisfaction. If accomplished, this may have influenced staff ratings of residents' social engagement and neuropsychiatric symptoms; authors did not report psychometric properties for outcome measures. There was a high risk of reporting bias. Participants were selected into the comparator group based on matching criteria that included baseline scores on a measure of functional capacity. Functional capacity was also assessed at 6- and 12-months and these data were not reported as outcomes.

QUALIDEM = Dementia-specific quality of life.

Appendix 4: Main Study Findings and Authors' Conclusions

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
Kok, 2018[9]	
Small-scale homelike care vs. traditional ward **QoL** QUALIDEM No significant differences between groups for any QoL outcome: i.e., care relationship, positive affect, negative affect, restless tense behaviour, positive self-image, social relations, social isolation, feeling at home, having something to do **Mood (Depression)** GDS-SF No significant difference between groups for mood **Neuropsychiatry** Behavioral Observation Scale for Intramural Geriatric Psychiatry *Anxious behaviour* Time x group effect: Eta squared = 0.086; P = 0.008 (Significant difference between groups over time) No significant differences between groups over time for apathy, not social behaviour, insubordinate behaviour, suspicious behaviour, or depressive behaviour	"This study demonstrates that moving to a small-scale care facility is associated with a reduction of anxiety in residents with dementia. These findings add to growing evidence supporting the benefits of small, homelike care facilities on well-being of these residents. The experience of decreased anxiety in residents in small-scaled homelike facilities was clinically relevant" (p. 6).
De Boer, 2017[13]	
Green homes vs. traditional nursing homes **Engaging in social/communication-related activity** No significant differences between groups after controlling for age, gender, cognition, and ADL independence (data not extracted) **Level of social interaction** **Beta = −11.8 (SE = 3.5), 95% CI = −19.4 (−4.1 to 124.6); P = 0.006, ES = 1.1** Significantly greater social interaction in Green homes vs. traditional nursing homes after controlling for age, gender, cognition, and ADL independence Green homes (n = 30) vs. traditional nursing homes (n = 26)	"In conclusion, green care farms have demonstrated that they are a valuable alternative to traditional nursing homes, as they provide residents with engagement in activities, social interaction, physical activity, and increased opportunities to go outside" (p.45).

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
Cognitive functioning No statistically significant differences between groups **ADL** No statistically significant differences between groups **Dementia severity** No statistically significant differences between groups	
De Boer, 2017[11]	
Green Care Farms vs. traditional facility **QoL** QoL-AD - Self-report No statistically significant difference between groups QoL-AD - Proxy-report Green care farm vs. traditional nursing home; M (SD) 32.9 (4.5) vs. 29.1 (4.9), $P < 0.05$; ES = 0.8 QUALIDEM *Positive affect* 15.8 (3.6) vs. 12.9 (3.5), $P < 0.05$, ES > 0.7 *Social relations* 13 (3.5) vs. 10.4 (3.8), $P < 0.05$, ES > 0.7 No statistically significant differences between groups for negative affect, restless tense behaviour, or social isolation **Social Engagement** RISE No statistically significant differences between groups **Neuropsychiatric symptoms** NPI-NH No statistically significant differences between groups **Agitation** CMAI No statistically significant differences between groups **Depressive Symptoms** CSDD	"Green care farms seem to be a valuable alternative to existing nursing homes. This is important as people with dementia are a heterogeneous group with varying needs. In order to provide tailored care there also is a need for a variety of living environments" (p.1).

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
No statistically significant differences between groups	
Hermer, 2017[12]	
HH vs. Legacy Control facility vs. Second Control Facility **Affective State** Observed Emotion Rating Scale No significant difference in time displaying positive affect between HH and comparator groups. **Cognitive Engagement** Menorah Park Engagement Scale No significant differences between HH and comparator groups for time displaying active engagement in activities, time engaged in socioexpressive activities, time staring blankly, or time sleeping during the day. HH spent less *time idle* than comparators: Percentage of time (SD) HH, 10.1% (6.6%) vs. C1, 20.3% (14.4%) vs. C2, 33.1% (20.2%) $F_{(2,38)} = 20.14$, $P \leq 0.00005$	"It revealed that, compared to residents at two control facilities with a traditional environment and an average degree of culture change adoption, HH-model residents experienced a distinct set of enhancements to their daily life" (p. 10).
Kok, 2016[8]	
Small-scale homelike care vs. traditional ward **Mood (Depression)** GDS-SF No significant difference between groups at 3- or 6-month follow up **Cognitive functioning** Standardized MMSE No statistically significant difference between groups at 3- or 6-month follow up *Proxy: nursing personnel* GIP No statistically significant difference between groups at 3- or 6-month follow up *Proxy: representatives* IQCODE Scored from −3 to 3 (much worse to much better) 5.5 (17.3) vs. −5.4 (26.0); $P = 0.03$, partial eta square = 0.07 (i.e., Significantly better cognitive functioning in the small-scale homelike care group vs. the traditional ward)	"The findings of the present study suggest that there is no difference between two types of care facilities for demented residents concerning (a decline in) global and specific cognitive functions over a certain period of time" (p.7).

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
Cluster - Global Cognitive Functioning Combined Standardized MMSE and IQCODE Global cognitive functioning was significantly greater in small-scale homelike SCU vs. traditional SCU at 3-month follow up. 12.8 (17.2) vs. −1.4 (29.5); P = 0.03, partial eta square = 0.07 There was no significant difference between groups at 6-month follow-up **Verbal memory** ADS Test - Eight Word Verbal Memory Test No significant difference between groups for *immediate recall, delayed recall,* or *recognition* **Visual memory** RBMT *Recognition of pictures* (maximum score 5); Mean (SD) 13.2 (3.7) vs.11.3 (4.1); P = 0.003, partial eta square = 0.19 ANCOVA (covariate = baseline scores) Small-scale homelike SCU had a slight improvement from baseline to 3-month follow-up while traditional homes had a slight decline over the same period. No significant between groups differences were observed at 6-month follow-up. *Recognition of faces* No significant difference between groups at 3- or 6-month follow up **Cluster – Episodic Memory** Combined RBMT pictures, RBMT faces, Eight word verbal memory test No statistically significant difference between groups **Language** BNT-short No significant difference between groups **Praxis** van Heugten diagnostic test for apraxia No statistically significant difference between groups **Executive functioning** Trail Making Test A and B No statistically significant difference between groups The Category Fluency Task from the GIT No statistically significant difference between groups Clock Drawing test	

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
No statistically significant difference between groups **Cluster – Executive Functions** Combined the Category Fluency Test, Clock Drawing Tests, and the Trail Making Tests No statistically significant difference between groups **Visual perception** GIT subtask – Incomplete Drawings No statistically significant difference between groups	
Lee, 2016[14]	
Small-scale homelike care vs. traditional long-term care **Self-care function** MOSES No significant differences between groups **Cognitive status** MOSES *Disorientation*: No significant differences between groups **Irritability**: MOSES $F_{2,10}$ = 4.76; $P < 0.05$ between unit comparisons favours intervention **Depressed/ anxious mood**: MOSES No significant differences between groups **social withdrawal** MOSES No significant differences between groups **Residents; functions, needs, and latent risks** MDS 2.0 *Oral/nutrition status (includes chewing/swallowing)*: $F_{2,10}$ = 11.75; $P < 0.01$ No significant differences between groups for *cognitive patterns, communication patterns, mood/behaviour patterns, psychosocial well-being, physical functioning and structural problems, or activity pursuit patterns*	"Findings of the current study indicate that small-scale facilities have positive effects on health and behavior of residents in long-term care facilities" (p.90).
Yoon, 2016[16]	
GH vs. traditional home; mean (SD)	"Although GH nursing homes are an innovative model to make the nursing home environment more person-centered

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
No significant differences between groups for *ADL function*, *cognitive function*, or *depressive symptoms* at any time point.	(Zimmerman & Cohen, 2010), this study did not demonstrate significant differences in ADL function for the GH home model residents compared to traditional nursing home residents" (p.11).
Yoon, 2015[15]	
GH vs. traditional home Method of analysis: Zero Inflated Poisson growth curve model **Social engagement** Standardized beta = -0.274, *P* = 0.010 GH had less increase in the probability of "not being socially engaged" over time compared with traditional homes Standardized beta = -0.010, *P* = 0.913 There was no difference between groups for rate of increase in social engagement level	"Major findings were that GH residents had a higher rate of increase of depressive symptoms, and a lower rate of increase in the probability of not being socially engaged over time relative to those in traditional nursing homes" (p.7).
Verbeek, 2014[10]	
Small-scale homelike care vs. traditional ward **Neuropsychiatric symptoms** NPI-NH *Aberrant motor behaviour* Group effect: MD (adjusted) = 0.19; 95% CI = 0.02 to 0.35; *P* = 0.02 No significant difference between groups for other sub-domains: *delusions, hallucinations, agitation/aggression, depression, anxiety, euphoria, apathy, disinhibition, irritability, night-time behaviour, eating change* ***Agitation*** CMAI *Physically non-aggressive behaviour* (total scores; range 4 to 28) Baseline: No significant difference between groups (data not reported) 6-months: No significant difference between groups (data not reported) 12-months: 6.82 (4.93) vs. 5.82 (3.43) MD (adjusted) = 2.58; CI = 1.00 to 4.17; *P* = 0.001 Small-scale residents had more physically non-aggressive behaviour	"Effects on behavior were mixed and suggest more social engagement for residents in small-scale living facilities, although this effect disappeared over time in the total group of residents. Furthermore, physically non-aggressive behavior (such as aimless wandering) was more present after 12 months, and more aberrant motor behavior (e.g. repetitive behaviors such as picking, handling buttons, and wrapping strings) was found for all measurements in small-scale living facilities compared with traditional wards. No effects were found for the other neuropsychiatric or depressive symptoms" (p.665).

Table 4: Summary of Findings of Included Non-Randomized Studies

Main Study Findings	Authors' Conclusion
No significant differences between groups for *physically aggressive behaviour or verbally agitated behaviour* **Depressive symptoms** CSDD No significant difference between groups for depressive symptoms **Social engagement** ISE subscale from RAI-MDS Baseline: 3.87 (1.67) vs. 2.93 (1.81) MD (adjusted) = 0.93; 95% CI = 0.48 to 1.38; *P* < 0.001 Intervention group significantly more engaged at baseline 6-months: 3.44 (1.83) vs. 3.06 (1.64) MD (adjusted) = 0.50; 95% CI = 0.03 to 0.98; *P* = 0.038 Intervention group significantly more engaged at 6 months 12-months: No significant difference between groups **Dementia severity** GDS No significant difference between groups **Functional status / ADL-capacity** ADL-hierarchy subscale from RAI-MDS No significant difference between groups	

ADL = activities of daily living; ADS = Amsterdam Dementia Screening test; ANCOVA = analysis of covariance; BNT = Boston Naming Test; C = comparator; CMAI = Cohen-Mansfield Agitation Inventory; CSDD = Cornell Scale for Depression; GDS = Global deterioration scale; GDS-SF = Geriatric Depression Scale – Short Form; GH = Green House; GIP = behavioural observation scale for intramural psychogeriatry; GIT = Groningen Intelligence Test; HH = household; ISE = Index of Social Engagement; IQCODE = Information Questionnaire on Cognitive Decline in the Elderly; MD = mean difference; MDS = minimum data set; MMSE = mini-mental state examination; MOSES = Multidimensional Observation Scale for Elderly Subjects; NPI-NH = Neuropsychiatric Inventory – Nursing Home Version; QoL = quality of life; QoL-AD = Quality of Life-Alzheimer's Disease scale; QUALIDEM = Dementia-specific quality of life; RAI = Resident Assessment Instrument; RISE = Revised index for social engagement; RBMT = Rivermead Behavioural Assessment Instrument; SD = standard deviation; SCU = special care units.

Note: If there were no statistically significant differences between groups, data were not extracted.