HHS Public Access

Author manuscript

Clin Trials. Author manuscript; available in PMC 2022 February 16.

Published in final edited form as:

Clin Trials. 2022 February ; 19(1): 86–96. doi:10.1177/17407745211046672.

Methodological challenges in pragmatic trials in Alzheimer's disease and related dementias: opportunities for improvement

Monica Taljaard,
Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

Fan Li,
Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

Bo Qin,
Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

Caroline Cui,
Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

Leyi Zhang,
Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

Stuart G Nicholls,
Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

Kelly Carroll,
Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

Susan L Mitchell
Hebrew Senior Life Marcus Institute for Aging Research, Boston, Massachusetts, USA

Abstract

Background and Aims: We need more pragmatic trials of interventions to improve care and outcomes for people living with Alzheimer's disease and related dementias. However, these trials present unique methodological challenges in their design, analysis, and reporting — often, due to the presence of one or more sources of clustering. Failure to account for clustering in the design and analysis can lead to increased risks of type I and type II errors. We conducted a review to describe key methodological characteristics and obtain a "baseline assessment" of methodological quality of pragmatic trials in dementia research, with a view to developing new methods and

Corresponding author: Monica Taljaard, Clinical Epidemiology Program, Ottawa Hospital, Civic Campus, 1053 Carling Avenue, Civic Box 693, Admin Services Building, ASB 2-004, Ottawa, ON K1Y 4E9, 613-798-5555 x18618 phone, mtaljaard@ohri.ca.

Declaration of conflicting interests
None declared.

practical guidance to support investigators and methodologists conducting pragmatic trials in this field.

Methods: We used a published search filter in MEDLINE to identify trials more likely to be pragmatic and identified a subset that focused on people living with Alzheimer's disease or other dementias or included them as a defined subgroup. Pairs of reviewers extracted descriptive information and key methodological quality indicators from each trial.

Results: We identified N=62 eligible primary trial reports published across 36 different journals. There were 15 (24%) individually randomized, 38 (61%) cluster randomized, and 9 (15%) individually randomized group treatment designs; 54 (87%) trials used repeated measures on the same individual and/or cluster over time and 17 (27%) had a multivariate primary outcome (e.g., due to measuring an outcome on both the patient and their caregiver). Of the 38 cluster randomized trials, 16 (42%) did not report sample size calculations accounting for the intracluster correlation and 13 (34%) did not account for intracluster correlation in the analysis. Of the 9 individually randomized group treatment trials, 6 (67%) did not report sample size calculations accounting for intracluster correlation and 8 (89%) did not account for it in the analysis. Of the 54 trials with repeated measurements, 45 (83%) did not report sample size calculations accounting for repeated measurements and 19 (35%) did not utilize at least some of the repeated measures in the analysis. No trials accounted for the multivariate nature of their primary outcomes in sample size calculation; only one did so in the analysis.

Conclusions: There is a need and opportunity to improve the design, analysis, and reporting of pragmatic trials in dementia research. Investigators should pay attention to the potential presence of one or more sources of clustering. While methods for longitudinal and cluster randomized trials are well-developed, accessible resources and new methods for dealing with multiple sources of clustering are required. Involvement of a statistician with expertise in longitudinal and clustered designs is recommended.

Keywords

Cluster randomized trials; longitudinal data analysis; multilevel modeling; missing data; intracluster correlation; pragmatic trials; Alzheimer's disease; dementia

Introduction

There is a critical need for more pragmatic trials of interventions to improve care and clinical outcomes for people living with Alzheimer's disease or related dementias and their caregivers.[1] This vulnerable population is increasing in prevalence but is at risk of poor health outcomes and poor quality of life due to the relative absence of high quality pragmatic randomized controlled trials evaluating interventions and programs to meet their complex needs. Pragmatic trials aim to generate evidence that can directly inform decision-making by embedding the research in clinical practice and deviating as little as possible from usual care conditions.[2] The recently launched US *National Institute on Aging Imbedded Pragmatic Alzheimer's disease or Alzheimer's disease and related dementia Collaboratory* (IMPACT) aims to build capacity to conduct pragmatic trials for people living with dementia and their caregivers, and to develop statistical methodology and guidance through its Design

and Statistics Core.[3,4] As a first step, we undertook a review of the existing landscape of pragmatic trials in this field.

Pragmatic trials involving people living with dementia and their caregivers raise unique statistical complications for several reasons. *First*, the nature of interventions and the setting (e.g., health system interventions implemented in nursing homes) may necessitate or encourage the use of cluster randomization; alternatively, individual randomization may be feasible but the intervention may need to be delivered by a common provider or therapist within a group setting.[5] Cluster randomized trials and individually randomized group treatment trials have special requirements for design and analysis due to the presence of intracluster correlation, i.e., the nonindependence of observations from multiple individuals belonging to the same cluster.[6] *Second*, repeated outcome assessments on the same individual may be necessary to examine the effect of an intervention over time.[7] In cluster randomized trials, the need for greater statistical efficiency may encourage adoption of stepped wedge or other multiple period designs that measure outcomes repeatedly in the same cluster (and/or the same individuals) over time.[8] To reap the benefits of repeated measures, statistical methods that incorporate all available measurements and account for correlations over time are required. *Third*, the intervention may specifically target the patient-caregiver dyad with outcomes assessed on each patient and their caregiver.[9] For example, depressive symptoms or quality of life may be assessed on both the patient and their caregiving spouse to allow for mutual influences of patients and caregivers on the response to an intervention. Ideally, such trials should be designed and analyzed using multivariate approaches that take the correlation between the outcomes into account.[10] In the presence of missing data — a particular challenge in aging research — multivariate approaches can also help mitigate loss of information due to missing values for one member of the dyad.[11] *Fourth*, the intervention may be tested for its effect on a range of outcomes, e.g., multiple domains within a questionnaire-based scale.[12] An estimate of the intervention effect can be obtained with greater precision by taking the correlation between the outcomes into account.[13,14] Furthermore, co-primary endpoints may be necessary,[15] for example, to demonstrate an intervention's effect on both cognitive and functional endpoints; analyzing such outcomes using a multivariate approach has several advantages. In summary, pragmatic trials in dementia research may be characterized by one or more sources of clustering due to the choice of cluster randomization, delivery of treatment in groups, use of repeated measures, interventions targeting patient-caregiver dyads, and multivariate primary trial outcomes. Failing to account for these sources of clustering can lead to increased risks of either type I or type II errors. Table 1 summarizes these sources of clustering, and their implications for sample size calculation and analysis.

This manuscript describes the results from a systematic literature survey to assess the methodological conduct and reporting of pragmatic trials in people living with dementia published over the past five years with a special emphasis on how sources of clustering were handled and reported. CONSORT reporting guidelines for cluster randomized and stepped wedge designs[16,17] require clear descriptions of how clustering was accounted for in the sample size calculation and analysis. They also require estimates for the intracluster correlation coefficient and correlations in repeated measures over time to be reported as these correlation coefficients can usefully inform sample size calculations for future trials.

Our main objectives were to a) describe characteristics and design features of pragmatic trials in this field; b) assess the extent to which clustering was accommodated in sample size calculations; c) assess the extent to which clustering was accommodated in the analysis; and d) describe the prevalence of reporting estimates for correlation coefficients. We also examined the extent to which covariates used in the randomization were adjusted for in the analysis,[18] and whether methods to mitigate effects of missing data were used.

Methods

Database of pragmatic trials

This study was embedded within the scope of a larger project about ethical and methodological issues in pragmatic trials.[19] As part of that project, we derived and validated a search filter to locate pragmatic trials in MEDLINE,[20] implemented the filter on 3 April 2019 to cover the period 1 January 2014 to that date, and conducted a landscape analysis of the resulting set of 4337 eligible trials.[21] We chose 2014 as the start date as this was the year that the National Library of Medicine began indexing articles as “pragmatic clinical trial as publication type” as well as topic. The methods for identifying eligible trials and flow diagram for trials included in the larger landscape analysis have been described in detail elsewhere.[21] In brief, we included primary reports of pragmatic randomized controlled trials with at least 100 individuals (excluding pilot or feasibility studies, protocols, and secondary analyses).

Identification of pragmatic trials in Alzheimer’s disease and related dementias

For the present manuscript, we aimed to identify —within the larger pragmatic trials database — trials that a) specifically focused on people living with dementia, or b) focused on a broader cohort of older adults which included a subgroup with dementia and conducted a stratified or subgroup analysis on that cohort. To locate trials of type a), we applied a search filter from the Cochrane dementia and cognitive improvement group.[22] To efficiently identify trials of type b), we used MeSH terms to identify trials exclusively in the elderly (aged 65 and over). Appendix 1 in the supplemental material provides full details of our search. Five reviewers (BQ, CC, LZ, MT and FL) then jointly screened 15 potentially eligible trials as part of a training process. After discussing discrepancies, the remaining trials were distributed amongst the first three reviewers who independently screened them with one reviewer per trial. Each trial screened in was classified as an individually randomized, cluster randomized, or individually randomized group treatment trial.

Data extraction

The data extraction form was pilot tested on samples of three individually randomized trials, three cluster randomized trials, and three individually randomized group treatment trials as part of a training and calibration exercise. All five reviewers, (BQ, CC, LZ, MT and FL) completed these extractions and then met to review discrepancies and refine the extraction form. Thereafter, the remaining studies were distributed amongst three reviewers (BQ, CC, LZ) with two reviewers per trial independently extracting information from each trial. The

reviewers met to discuss discrepancies; MT and FL were consulted and reached agreement when discrepancies could not be resolved.

The data extraction form is attached as Appendix 2. It included sections on general study characteristics, study design, sample size calculation, and analysis. General study characteristics included country of study conduct, setting, study and control interventions, data collection procedures, number of individuals (or dyads) randomized, and journal impact factor. The study design section included items on the unit of randomization, method of allocation, and type of primary outcome. The primary outcome was classified as a univariate outcome measured on either the patient or caregiver, bivariate outcome measured on both the patient and caregiver, or multivariate outcome measured on the same participant. If a primary outcome was not identified, we used the outcome in the sample size calculation, and if a sample size calculation was not provided, the first outcome listed under the objectives. Cluster randomized trials with repeat measurements were classified as closed cohort (i.e., the same individuals measured over time), open cohort (i.e., a mixture of the same and different individuals), or cross-sectional (i.e., different individuals measured each time).

We extracted whether a sample size or power calculation was presented. If a cluster randomized or individually randomized group treatment trial, we classified whether clustering was accounted for in the calculation. Clustering was considered accounted for if the calculation was done at the individual-level and clearly stated that an intracluster correlation coefficient or design effect was used to adjust for clustering, or if the calculation was done at the cluster-level. If the trial had repeated measures, we extracted whether the sample size calculation accounted for them (e.g., by assuming a correlation with baseline, or by using design effects for longitudinal designs such as cluster cross-over or stepped wedge designs). If the trial had a dyadic or multivariate primary outcome, we extracted whether the sample size calculation was based on the dyadic or multivariate outcome accounting for correlation in the outcomes.

For cluster randomized trials and individually randomized group treatment trials, we extracted whether the analysis accounted for clustering. Clustering was considered accounted for if analysis was at individual-level and clearly accounted for intracluster correlation or was conducted at the cluster-level. For repeated measures designs, we extracted whether the analysis was based on repeated measures (e.g., using analysis of covariance or longitudinal regression accounting for correlation in repeated measures). If the trial had a dyadic or multivariate primary outcome, we extracted whether the analysis used a multivariate approach accounting for correlation in outcomes. We also extracted whether any correlation coefficients (within clusters, over time, or between multivariate outcomes) were reported. Adjustment for any covariates used in the randomization, presence of missing data, and use of any missing data method in both the primary analysis of the primary outcome, and secondary or sensitivity analyses for the primary outcome were recorded.

Analysis

We described categorical variables using frequencies and percentages, and continuous variables using range, mean and standard deviation, and median and interquartile range

(Q1-Q3). Average cluster size was calculated as the number of randomized individuals or dyads divided by the number of clusters.

Results

Identification of eligible trials

From our database of 4337 pragmatic trials, 488 trials were identified as potentially relevant to Alzheimer's disease and related dementias using a composite of either the Cochrane filter (273 trials) and the MeSH aged filter (281 trials). From these trials, reviewers screened in N=62 as eligible: 60 focused solely on people living with dementia and 2 included such populations as a defined subgroup. These trials were published across 36 different journals (see Appendix 3) with the highest frequency in the Journal of the American Geriatrics Society (7 trials).

Study characteristics

Table 2 presents a summary of the study characteristics. The majority were conducted in the European Union, United Kingdom or United States, and took place in a nursing home setting. The types of experimental interventions were diverse: most commonly patient non-pharmacological interventions (47%), followed by educational interventions targeted at health professionals (31%), interventions targeted at the health care organization (27%), interventions targeting the caregiver (19%) and the patient-caregiver dyad (13%), and patient pharmacological interventions (5%). The most commonly used data collection procedures were patient-focused questionnaires (73%); more than half used mental or physical examinations not required for normal patient care (53%); 36% used caregiver-focused questionnaires; 34% used routinely collected data or reviews of patient medical records. The majority were multicenter trials and enrolled a median of 267 individuals or dyads.

Types of trial designs

Table 3 presents the types of trial designs, methods of random allocation and use of repeated measures. There were 15 (24%) individually randomized, 38 (61%) cluster randomized, and 9 (15%) individually randomized group treatment designs. The majority were parallel arm designs; two used a stepped wedge design. Over a third (39%) did not use stratification or some other type of balancing constraint. Twelve trials (19%) did not clearly identify a primary trial outcome. The nature of the primary outcome was most often continuous (81%), and univariate assessed on the patient (63%) or caregiver (10%); 5 trials (8%) had a dyadic primary outcome assessed on both the patient and their caregiver, and 12 (19%) a multivariate primary outcome measured on the same participant. Only 8 trials (13%) did not have repeated measures on the primary outcome. Among the 34 cluster randomized trials with repeated measures, almost all (94%) were closed cohort designs. The median number of clusters randomized was 22 ranging from 2 to 168. The median cluster size was 14 (3 to 708).

Sample size calculation methods

Table 4 presents details about reporting and conduct of sample size calculation procedures. Fifty trials (81%) reported a sample size or power calculation. Among the 38 cluster randomized trials, 22 (58%) accounted for the intracluster correlation while 6 (16%) did not; the remaining 10 (26%) did not present sample size calculations. Among the 9 individually randomized group treatment trials, all presented sample size calculations but 6 (67%) did not account for the intracluster correlation. Among 54 trials with repeated measurements, 9 (17%) accounted for repeated measures in the sample size calculation, while 33 (61%) did not; the remaining 12 (22%) did not present sample size calculations. Among 17 trials with dyadic or multivariate outcomes, none presented sample size calculations that accounted for the dyadic or multivariate nature of their primary outcomes.

Methods of analysis

Table 5 presents details about the methods of analysis for the primary outcome and reporting of correlation coefficients. Among the 38 cluster randomized trials, 13 (34%) did not clearly account for the intracluster correlation in the analysis; among the 9 individually randomized group treatment trials, 8 (89%) did not account for intracluster correlation in the analysis. Among the 54 trials with repeated measures, 19 (35%) did not utilize at least some of the repeated measures in the analysis. Among the 17 trials with dyadic or multivariate outcomes, 16 (94%) did not account for the dyadic or multivariate nature of the outcome in the analysis. Reporting of measures of correlation was poor: 17 (36%) of studies using cluster randomization and group treatment reported an estimate of the intracluster correlation; no studies reported measures of correlation over time and no studies reported measures of correlation between patient and caregiver or multivariate outcomes. The most frequently used method of analysis was mixed effects regression.

Covariate adjustment and missing data

Table 6 presents details about adjustment for covariates used in the randomization and missing data. Among the 38 trials using restricted randomization methods to balance allocations between the arms, 14 (37%) adjusted for all balancing factors as covariates in the analysis. Nearly all trials (95%) had missing data on the primary outcome; 34 of these (58%) did not use any method to account for missing data in the primary analysis of the primary outcome. When considering both primary and sensitivity analyses, 32 trials (54%) did not use any method to account for missing data. Nearly half reported a statistically significant result for the primary outcome.

Handling of clustering by number and sources of clustering

We conducted a post-hoc analysis tabulating the adequacy of sample size and analysis methods according to the number and source of clustering (see Appendix 4). Three trials (5%) had no sources of clustering; 12 (19%) had one, 38 (61%) two, and 9 (24%) had three sources of clustering. Among trials with one, two, and three sources of clustering, the percentages accounting for all sources of clustering in their sample size calculations were 33%, 11% and 0% respectively, while the percentages accounting for all sources of clustering in the analysis were 75%, 24% and 0% respectively.

Discussion

Summary of principal findings

We reviewed recently conducted pragmatic trials in Alzheimer's disease and related dementias focusing on their key statistical and reporting requirements. Trials commonly had multiple sources of clustering. Despite several decades of availability of methods to account for intracluster correlation, many cluster randomized trials and individually randomized group treatment trials did not account for intracluster correlation in their sample size calculations and analysis. Trials with repeated measures were common but this design aspect was often ignored during sample size calculation and analysis. Few trials with dyadic or multiple primary outcome domains used methods for multivariate outcomes. Sample size and analysis methods were particularly poor when there were multiple sources of clustering. Estimates for relevant correlation coefficients were seldom reported. Nearly all trials were subject to missing data but methods to account for missing data were infrequently used.

Comparison with other studies

We are unaware of any other published methodological reviews of pragmatic trials in this field. Previous reviews of cluster randomized trials, focused on single sources of clustering, have consistently found that quality tends to be poor. Diaz-Ordaz and colleagues[23] reviewed 73 cluster randomized trials in nursing homes until the end of 2010, of which 27% reported accounting for clustering in sample size calculations and 74% in the analyses (compared to our review of 62 trials published 2014–2019 in which 58% reported accounting for clustering in sample size calculation and 66% in the analyses); they found that only 11% of trials reported intracluster correlation coefficients (compared to 36% in our review). Kahan and colleagues[18] found that 26% of 258 individually randomized trials published in four major medical journals in 2010 adjusted for all balancing factors in their primary analyses (compared to 37% in our review). We are unaware of any published reviews of clustered longitudinal trials or trials with multiple sources of clustering.

Limitations

Locating pragmatic trial reports in the literature is challenging. We used our published search filter to locate trials that are more likely to be pragmatic.[20] The search filter was designed to capture not only trials that self-declare their pragmatic intention in the title or abstract, but also trials that only have design features suggesting a pragmatic intention (without explicit identification as pragmatic in the title or abstract). We did not score each trial (e.g., using the PRECIS-2 tool)[24] to confirm that the trial was mostly pragmatic. Even if retrospective scoring of a large database of trials were feasible, there is no objective threshold for determining when a trial can be considered sufficiently "pragmatic".[25] Furthermore, no reporting guidelines require authors to label their trials as "pragmatic" in the title or abstract;[26] thus, our search filter may not have captured all trials with pragmatic intentions in this field.

We extracted data mainly from the primary trial report and did not access protocols. It is possible that appropriate methods were used in the design; however, CONSORT guidelines are clear in their requirements to provide explicit details about methods in the trial report.

Implications of our findings

There is a recognized need for more pragmatic trials in Alzheimer's disease and related dementias and there is now an opportunity to build capacity through the IMPACT Collaboratory.[27] Our review of the existing landscape of pragmatic trials in this field reveals a critical need for improvement. Failure to account for clustering in sample size calculations may mean that many trials are under-powered, while failure to account for repeated measures may imply over-recruitment in a vulnerable population and a waste of resources. Failure to account for intracluster correlation in analysis may mean that some trials have overstated the statistical significance of their findings, while failure to adopt methods that account for repeated measures imply a missed opportunity for statistical efficiency, and potentially useful treatments being declared ineffective. While our focus has been on pragmatic trials in dementia research, we expect similar considerations to apply in pragmatic trials in other clinical areas involving vulnerable populations and their caregivers for example, in hemodialysis where attempts are being made to advance the conduct of more pragmatic trials.[28]

Recommendations

We make several recommendations based on these results:

1. Trialists conducting clustered designs should collaborate with statisticians experienced in methods for these designs and should ensure that all sources of clustering are accounted for in the sample size calculation and analysis.[29] When cluster randomization is necessary, trialists should consider the use of more efficient multiple period designs which can substantially reduce the required number of clusters, although this should be balanced against the increase in the number of measurements.

2. When a single primary outcome is not adequate to inform a decision about effectiveness of an intervention, a limited number of co-primary outcomes may be considered. However, multiple primary outcomes raise complex challenges for type I and type II error control during the design and analysis. An important consideration is whether the intervention will be declared a success if at least one primary outcome is significant or whether joint significance on all primary outcomes is required. When joint significance is required, power decreases as the number of outcomes being evaluated increases. It is useful to consider the correlation between the outcomes during sample size calculation as it may lessen the increase in the required sample size. When the outcomes have very different effect sizes and the correlation is not very strong, there may not be much practical benefit to incorporating correlations between the outcomes in the sample size calculation as the smallest effect size will primarily determine the sample size.[30,31,32] Nevertheless, adopting multivariate methods which account for correlation between the outcomes has additional advantages, for example, when there are missing values for one outcome but not others.

3. Restricted randomization techniques such as covariate-constrained randomization should be considered to improve balance in cluster randomized

trials with few clusters[33] and variables balanced by design should be adjusted in the primary analysis to obtain correct p-values and improve power and efficiency.[34]

4. Methods to account for missing data should be used in the analysis and where relevant, should account for sources of clustering.[35,36,37]

5. There is a need for new methods for trials with multiple sources of clustering, for example, cluster randomized trials with dyadic or multivariate outcomes and longitudinal trials with multivariate outcomes. Practical tools including statistical software and tutorial-style manuscripts should accompany more theoretical work to promote the uptake of appropriate methods in practice.[38] Appendix 5 summarizes available references for multiple sources of clustering.

6. Journal editors and reviewers should require adherence to minimum reporting requirements16,17,26 and should especially insist that estimates of correlation are provided to inform the design of future studies. Investigators with access to routinely collected health administrative data should consider producing databases of correlation estimates for outcomes of potential interest in future trials and making them publicly available.[39]

Conclusions

There is a need and opportunity to improve the design, analysis, and reporting of pragmatic trials in dementia research. These trials often have multiple sources of clustering that need special consideration in the design, analysis, and reporting. While methods for longitudinal and cluster randomized designs are well-developed, accessible resources and new methods for dealing with multiple sources of clustering are required. Involvement of a statistician with expertise in longitudinal and clustered designs is recommended.

Acknowledgements

We thank Hayden Nix and Jennifer Zhang who contributed to the creation of the large database of pragmatic trials. We thank Dr. Spencer Hey who helped with database management aspects.

Funding

The author(s) disclosed receipt of the following financial support for the research of this article: This work was supported by the Canadian Institutes of Health Research through the Project Grant competition (competitive, peer-reviewed), award number PJT-153045; and by the National Institute of Aging (NIA) of the National Institutes of Health under Award Number U54AG063546, which funds NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix

Appendix

Appendix 1:

Search filters used in the identification of N=62 eligible trials

Search strategies for pragmatic trials in Ovid MEDLINE [1]	
#	**Search Statement**
	Trial design terms
1	(((pragmatic$ OR naturalistic OR real world OR real life OR unblinded OR unmasked OR cluster OR step$ wedge$ OR point of care OR factorial OR switchback OR switch back OR phase 4 OR phase IV) adj 10 (study OR trial)) OR (practical trial OR effectiveness trial OR ((cluster$ or communit$) adj2 randomi$))).tw.
	Trial attribute terms
2	(general practice$ OR primary care OR registry based OR health record$ OR medical record$ OR EHR OR EMR OR administrative data OR routinely collected data OR (communit$ adj2 intervention$) OR quality improvement OR implementation OR decision support OR health service$ OR health system$ OR comparative effectiveness OR CER OR usual care OR evidence based OR practice guideline$ OR (guideline$ adj1 recommend$) OR knowledge translation OR health technology assessment OR HTA OR cost effectiveness OR process evaluation OR economic evaluation OR patient oriented).tw.
	Limit to records likely to be RCTs
3	randomized controlled trial.pt. OR ((comparative effectiveness OR randomi?ed) adj10 (trial OR study)).ti.
4	(comment on OR phase 1 OR phase I OR phase 2 OR phase II OR non-randomi?ed OR quasi-randomi?ed OR pseudo-randomi?ed).ti. OR (clinical trial, phase I OR clinical trial, phase II OR systematic review OR meta-analysis OR review OR editorial).pt.
	Include records tagged as pragmatic trials
5	pragmatic clinical trial.pt.
	Sensitivity-maximizing search (combines trial design terms or attribute terms with RCT terms)
6	((1 OR 2) AND (3 NOT 4)) OR 5
7	exp Animals/ NOT Humans/
8	6 NOT 7

RCT=randomized controlled trial

[1] Taljaard M, McDonald S, Nicholls SG, Carroll K, Hey SP, Grimshaw JM, Fergusson DA, Zwarenstein M, McKenzie JE. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *J Clin Epidemiol.* 2020 Aug;124:75–84. doi: 10.1016/j.jclinepi.2020.05.003. Epub 2020 May 11. PMID: 32407765

Cochrane Dementia and Cognitive Improvement Group PubMed search filter to identify trials specifically focused on Alzheimer's and dementia disease[2]

((((((((Dementia[Mesh]) OR "Neurocognitive Disorders"[Mesh:NoExp]) OR dement*[tiab]) OR alzheimer*[tiab]) OR AD[tiab]) OR (("lewy bod*"[tiab] OR DLB[tiab] OR LBD[tiab] OR FTD[tiab] OR FTLD[tiab] OR "frontotemporal lobar degeneration"[tiab] OR "frontaltemporal dement*"[tiab]))) OR "cognit* impair*"[tiab]) OR ((cognit*[tiab] AND (disorder*[tiab] OR declin*[tiab] OR fail*[tiab] OR function*[tiab] OR degenerat*[tiab]

[2] **ALOIS A comprehensive, open-access register of dementia studies. Available at:** [https://alois.medsci.ox.ac.uk/about-alois]. Accessed: 4 December 2020.

OR deteriorat*[tiab]))) OR ((memory[tiab] AND (complain*[tiab] OR declin*[tiab] OR function*[tiab] OR disorder*[tiab])))

Airtable filter to identify trials in the elderly based on MeSH terms[3]

AND(IF(OR(FIND("Aged",{MeSH_Terms}),FIND("aged",{MeSH_Terms})),"True","False")="True",IF(OR(FIND("Middle",{MeSH_Terms}),FIND("middle",{MeSH_Terms})),"False","True")="True",IF(OR(FIND("Adult",{MeSH_Terms}),FIND("adult",{MeSH_Terms})),"False","True")="True")

Appendix 2: Data extraction form

Notes: Data abstraction pertains to **main study component** (the patient population on which the primary outcome evaluation is carried out), i.e., disregard sub-studies within the main trial, e.g., if a smaller group of patients are enrolled for more intensive follow-up.

1. Reviewer name ☐

General Study Characteristics

2. Country of study recruitment (for identification of development status, use https://data.worldbank.org/country) (Select all that apply) (Please look up in the protocol where the study took place if this is not available in the report)
   a) Canada
   b) USA
   c) UK
   d) Other European Union (EU) country
   e) Australia or New Zealand
   f) Low or Middle Income Country (LMIC)
   g) Other developed

3. Setting.
   1 ☐ Primary care practices or primary care providers
   2 ☐ Hospitals or hospital wards
   3 ☐ Nursing homes or nursing home wards
   4 ☐ Communities or residential areas
   5 ☐ Adult day care centers
   6 ☐ Other (specify) ☐

4. Type(s) of **experimental** interventions (all components of study interventions) (Select all that apply)
   a) Educational interventions targeted at health professionals (e.g., distribution of educational materials, outreach visits, audit and feedback)
   b) Quality improvement interventions targeted at organization of health care or health services delivery (e.g., financial, shifting of professional roles, multi-disciplinary teams, integration of services, changes in setting or equipment, home visits by nurses)
   c) Patient health promotion, behavioural or educational intervention
   d) Patient therapeutic intervention (e.g., pharmacological or non-pharmacological clinical treatment – distinguish from indirect changes to patient therapies as a result of guideline adherence)
   e) Any type of caregiver intervention (an intervention on the spouse, adult child, family member, or care partner)
   f) Any type of intervention specifically targeting the patient-caregiver **dyad** (e.g., an intervention to improve communication between patient and their family member)
   g) Other (specify) _____

[3]Syntax applied in Airtable [https://airtable.com/product] to identify trials exclusively focused on the elderly population and may include people living with dementia as a defined subgroup. Syntax identifies trials tagged with MeSH terms that include "aged", but not "middle" or "adult".

5. Type of intervention administered in **control** arm (Note: disregard activities administered in all clusters prior to randomization e.g., to ensure similar levels of knowledge before starting the intervention. In multi-arm trials, choose the least intensive intervention as the "control" arm):

1	Not reported
2	No active intervention (e.g., usual care, treatment as usual)
3	Scaled down version of active intervention (includes some basic elements of active intervention) (e.g., one educational visit, printed guidelines only, augmented care)
4	Placebo or sham intervention (e.g., vitamin placebos, education on unrelated medical conditions)
5	Other active intervention (head to head comparison, comparative effectiveness research)
6	Other (specify)

6. Type(s) of **data collection** for primary and secondary outcomes (select all that apply):

a) Medical record or health record review (whether paper or electronic)
b) Routinely collected health administrative data (e.g., Medicare or Administrative Claims data)
c) Patient specimen collection (blood test) or mental or physical examination **not required for normal patient care** (provider might complete a form to assess the patient's cognitive decline)
d) Interviewer- or self-administered or caregiver administered patient questionnaires (telephone/postal/face-to-face/internet) (Questions are about the patient outcomes)
e) Interviewer- or self-administered caregiver questionnaires (questions are about the caregiver outcomes, i.e., spouse, adult child or other family member of care partner)
f) Health professional survey questionnaires or interviews (questions are about their own outcomes)
g) Other (specify if none of the above)

7. Total number of eligible **participants** in the trial: (As reported in flow diagram or in table describing baseline characteristics. NOTE: By "participant" we mean the lowest level of participant, i.e., patient/family caregiver – not providers. This information will be used to calculate the average cluster or center size of participants. If primary outcome is a dyadic outcome (i.e., Q7=3) or if couples are randomized, then count the number of pairs (couples). If trial is a CRT with cross-sectional design, then count total number of participants over pre and post phases.

8. Single center or multicenter trial? Note: A "center" is a site (e.g., facility, hospital, clinic, or community) participating in the trial by recruiting participants into the trial; it may also be involved in delivering treatments and collecting data. Often, centers are stratification factors in the randomization, or may be the units of randomization. A multicenter trial can may have a smaller number of large centers, or a large number of small centers.

Study Design Characteristics

9. Did the authors clearly identify a primary outcome for the trial? (NOTE: Trial report should have one or a limited number of clearly identified outcomes as "primary")

1	Yes: primary outcome (or co-primary outcomes) clearly identified
2	No: none or unclear (e.g., simply lists many outcomes without identifying the primary outcome)
3	Other (explain) (e.g., primary outcome identified in another publication)

10. What was the trial **primary outcome?** (**Note:** A primary or co-primary outcome must be identified for each trial. Use the following hierarchy: Primary outcomes stated by authors; if no primary outcomes specified, use outcome in sample size calculation; if sample size calculation not reported or reported for a sub-study only, use first outcome listed under 'Objectives'; if still unclear, refer trial to arbitration before proceeding.)

11. On whom was the trial **primary outcome** (selected above) measured? (e.g., who completed the questionnaire)
    - 1 Patient
    - 2 Caregiver
    - 3 Both patient and caregiver (i.e., a dyadic outcome)
    - 4 Bivariate or multivariate outcome measured on the same participant

12. Trial design (units of randomization, intervention, and observation):
    - 1 Individually randomized trial (patients or caregivers or patient-caregiver dyads are the units of randomization, intervention and observation)
    - 2 Cluster randomized (clusters are units of randomization; individuals are units of observation)
    - 3 Individually randomized group treatment trial (individuals are randomized to receive their treatment in a group setting, e.g., mindfulness based therapy delivered in groups)

13. Number of **clusters** randomized (if cluster randomized) or number of participating centers (if individually randomized) or number of groups (if individually randomized group treatment trial)

14. Trial design for comparing interventions:
    - 1 Parallel arm design
    - 2 Factorial design
    - 3 Cross-over design
    - 4 Stepped wedge design
    - 5 Other (specify)

15. Method of random allocation:
    - 1 Completely randomized (unrestricted randomization – this includes use of permuted blocks)
    - 2 Stratified or stratified permuted block design
    - 3 Pair-matched
    - 4 Other (specify) (e.g., covariate constrained, minimization or unclear)

16. Measurement schedule (based on primary outcome):
    - 1 One post-test only (single measurement of the primary outcome post-intervention or time to event outcome)
    - 2 One pre-test and one post-test (single measurement of primary outcome, both pre- and post)
    - 3 More than one post (2 or more repeated measurements of primary outcome post-intervention **only**)

4 Multiple pre and post (a mixture of multiple pre and post repeated measurements of primary outcome)
5 Other (specify) _____

17. If cluster randomized design (Q12=2) with repeated measures (Q16=2,3,4), indicate type of trial design at individual-level (focus on the primary outcome here):
1 Closed cohort design (same individuals measured repeatedly over time)
2 Cross-sectional design (different individuals measured at different time points)
3 Open cohort design (mixture of same and different individuals measured repeatedly)
4 Unclear or other (specify) _____

Sample size calculation methods

18. Sample size / power calculations presented for the primary outcome?
1 Yes (Sample size calculation is presented for the primary outcome)
2 No (No sample size calculation presented for the primary outcome
3 Other (specify) (e.g., sample size calculation is presented but for a different outcome, outcome not specified, or definition and scale does not match the primary outcome)

19. If cluster randomized trial or individually randomized group treatment trial (Q12=2 or 3): indicate whether sample size / power calculation accounted for clustering:
1 **Yes: Participant**-level accounting for ICC ("Sample size was based on a significant effect size of 0.5, incorporated an ICC of 0.05 and was based on enrollment of 4 patients per physician"; "Based on a mean (SD) number of admission days per resident enrolled, within cluster variance of 2 days and between-cluster variance of 3 days and 10 residents per nursing home". Usually will involve stating at least the average cluster size and the ICC/ design effect/ /within-and between-cluster variance or stated that accounting for clustering without reporting value of ICC.)
2 **Yes: Cluster**-level (Should be clear that cluster-level summary data are used for calculation e.g., "sample size was based on the hospital as the unit of analysis…assuming a rate of episiotomy of 42% at baseline, with a standard deviation of 15%, we need 18 hospitals to identify a decrease in episiotomy rate." Use of standard deviation in the case of proportions indicates that binary data was summarized at cluster-level and treated as continuous data for the purpose of sample size calculation.)
3 **No: Participant** -level without accounting for ICC (clear that individual-level data were used but no mention of clustering)
4 **Unclear** whether accounted for clustering (e.g., "sample size was calculated to give a power of 80% of detecting a difference of 1 SD at 5% significance in mean diagnosis concordance score"; "sample size of 500 participants would result in 80% power to detect a difference of 10 points between groups")
5 Other (specify) (e.g., based on intermediate level of clustering) _____

20. If repeated measures design (i.e., Q16=2,3,4) and sample size calculation was presented: Were repeated measures accounted for in sample size / power calculation?
1 No (Primary outcome has repeated measures but this was not accounted for in sample size calculation, e.g., sample size calculation is based on a simple t-test or chi-squared test post-intervention)
2 Yes (e.g., using a method based on ANCOVA, longitudinal method, cross-over design, stepped wedge design, or other repeated measures sample size method)
3 Other (specify) _____

21. If trial has a dyadic or multivariate primary outcome (Q11=3 or 4), was the **sample size calculation** based on the dyadic or multivariate outcome accounting for correlation in the dyadic or multivariate response?

1	Yes, sample size accounted for patient-caregiver (or multivariate) correlation
2	No, sample size calculation was conducted separate or did not account for patient-caregiver correlation
3	No sample size calculation for this outcome presented
4	Unclear or other

Methods of analysis

22. If cluster randomized trial or individually randomized group treatment trial (Q12=2 or 3), indicate whether **primary analysis** of **primary outcome** accounted for clustering:

1	**Yes**: Analysis was at **individual-level** accounting for ICC (e.g., using mixed-effects logistic regression, GEE taking account of clustering by physician, random effects for physician, hierarchical modeling, multi-level modeling)
2	**Yes**: Analysis was at **cluster-level** (clearly stated that analysis at cluster-level, e.g., "analyses performed using patient-level variables aggregated at the provider-level", analysis was based on hospital rates, t-test weighted by inverse variance etc.)
3	**No**: Analysis was at **individual**-level **not** accounting for ICC (e.g., multivariable regression analysis of patient-level data with no mention of clustering, or standard 2-sample test on patient-level data without mention clustering or stated that since ICCs were low, clustering was ignored in presentation of results)
4	**Unclear** whether at individual-level or cluster-level or whether accounted for clustering
5	Other (specify) (e.g., based on intermediate level of clustering, both individual-level and cluster-level analyses used for primary outcome analysis)

23. If repeated measures design (i.e., Q16=2,3,4): Were repeated measures used in primary analysis of the primary outcome?

1	Yes: analysis used repeated measures (e.g., using ANCOVA, longitudinal regression analysis, random effects regression of the repeated outcomes)
2	No: primary analysis did not use the repeated measures
3	Other (specify)

24. If trial has a dyadic or multivariate primary outcome (Q11=3 or 4), was the **analysis** based on the dyadic or multivariate outcome accounting for correlation in the dyadic or multivariate response?

1	Yes, analysis accounted for patient-caregiver (or multivariate) correlation
2	No, analysis was conducted separate or did not account for patient-caregiver (or multivariate) correlation
3	Unclear or other

25. Was the ICC or other parameters related to correlations within clusters and within individuals reported? (Note: this is referring to estimates produced in the analysis - not referring to estimates used for sample size calculation):

	Yes	No or NA
a) Intracluster correlation (ICC or within-period ICC)	1	2

b) Correlation over time (e.g., cluster autocorrelation coefficient CAC, Pearson correlation with pre-intervention measurement, other over-time correlation) 1 2
c) Correlation between patient and caregiver (or multivariate) response 1 2

26. If any covariates were used in the randomization (i.e., Q15=2,3,4) did the primary outcome analysis adjust for those covariates (e.g., centers adjusted as either fixed or random effects)
1 No (Primary outcome analysis did not adjust for any factors used in randomization)
2 Yes (primary outcome analysis adjusted for all randomization factors)
3 Yes (primary outcome analysis adjusted for some randomization factors but not all)
4 Other or unclear (specify)

27. What was the method of primary outcome analysis? (Please provide a basic description paying attention to type of model, how they accommodated any clustering, and how they accommodated repeated measures, e.g., Linear mixed effect regression with baseline measure of outcome included as covariate and clusters specified as random effects)

28. Does the trial have missing outcomes, attrition, or drop-out for the primary outcome? (Note: choose "No" only when it is explicitly stated there was no attrition, or when it is clear from the design or flow diagram that there was no attrition. If it not clearly stated and there could plausibly have been attrition, then pick unclear).
1 Yes
2 No
3 Unclear or other

29. If Q28 = Yes, was any method used to account for missing outcomes in the primary analysis of the primary outcome?
1 No missing data method reported (i.e., complete case analysis)
2 Regression adjustment for covariates stated to be associated with missingness
3 Single imputation
4 Multiple imputation
5 Inverse probability weighting
6 Unclear or other

30. If Q28 = Yes, select any other methods used to account for missing outcomes in **other analyses** of the primary outcome (e.g., in sensitivity analysis)? (Select all that apply)
1 No other missing data methods reported for primary outcome
2 Regression adjustment for covariates stated to be associated with missingness
3 Single imputation
4 Multiple imputation
5 Inverse probability weighting
6 Unclear or other (e.g., complete case analysis)

31. Did the primary outcome analysis show a statistically significant effect?
1 Yes
2 No
3 Unclear or other (specify)

32. Any comments about the trial or the data extraction? (1)

Appendix

Appendix 3:

Journals where N=62 eligible trials were published

Journal title	Number of trials
Age & Ageing	1
Aging & Mental Health	2
Alzheimer's Research & Therapy	1
American Journal of Geriatric Psychiatry	6
American Journal of Physical Medicine & Rehabilitation	1

Journal title	Number of trials
American Journal of Psychiatry	1
Annals of Internal Medicine	1
Applied Nursing Research	1
Biological Research for Nursing	1
BMC Medicine	1
BMJ	1
BMJ Open	1
British Journal of Psychiatry	1
Clinical Interventions In Aging	1
Dementia & Geriatric Cognitive Disorders	1
Deutsches Arzteblatt International	1
Emergency Medicine Journal	1
Gerontologist	2
Health Technology Assessment (Winchester, England)	1
International Journal of Geriatric Psychiatry	4
International Journal of Nursing Studies	2
International Psychogeriatrics	3
JAMA Internal Medicine	2
JAMA Psychiatry	1
Journal of Aging & Health	1
Journal of Comparative Effectiveness Research	1
Journal of Medical Internet Research	1
Journal of Neurology, Neurosurgery & Psychiatry	1
Journal of the American Geriatrics Society	7
Journal of the American Medical Directors Association	2
Journals of Gerontology Series B-Psychological Sciences & Social Sciences	1
Palliative Medicine	2
PLoS Medicine / Public Library of Science	3
PLoS ONE [Electronic Resource]	3
The Lancet. Psychiatry	1
Zeitschrift fur Gerontologie und Geriatrie	1

Appendix

Appendix 4:

Post-hoc analysis examining the extent to which clustering was accounted for in sample size and analysis according to number and sources of clustering

	Accounted for all sources of clustering?	
Sources of clustering	**Sample size calculation**	**Analysis**
No sources of clustering (N=3)	**NA**	**NA**
One source of clustering (N=12)	**4 (33.3%)**	**9 (75.0%)**
Repeated measures only (N=8)	2 (25%)	8 (100%)
Cluster randomization only (N=3)	2 (66.7%)	1 (33.3%)
Group treatment only (N=1)	0	0
Two sources of clustering (N=38)	**4 (10.5%)**	**9 (23.7%)**
Repeated measures and multivariate/dyadic outcome (N=4)	0	0
Cluster randomization and repeated measures (N=25)	3 (12.0%)	8 (32.0%)
Cluster randomization and multivariate/dyadic outcome (N=1)	0	0
Group treatment and repeated measures (N=8)	1 (12.5%)	1 (12.5%)
Three sources of clustering (N=9)	**0**	**0**
Cluster randomization and repeated measures and multivariate/dyadic outcome (N=9)	0	0

NA: Not applicable

Appendix 5: Key references describing methods for trials with multiple sources of clustering

Cluster randomized trials with repeated measures on the same cluster and/or the same individual over time:

- Murray DM, Hannan PJ, Wolfinger RD et al. Analysis of data from group-randomized trials with repeat observations on the same groups. *Statistics in Medicine* 1998; 17: 1581–1600.
- Ukoumunne OC, Thompson SG. Analysis of cluster randomized trials with repeated cross-sectional binary measurements. *Statistics in Medicine*. 2001.,15;20(3):417–33.
- Liu, A., Shih, W.J. and Gehan, E, Sample size and power determination for clustered repeated measurements. *Statist. Med 2002.*, 21: 1787–1801. https://doi.org/10.1002/sim.1154
- Localio AR, Berlin JA, Have TR. Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: bias and coverage of frequentist and Bayesian methods. *Statistics in Medicine*. 2006., 30;25(16):2720–36.

- Teerenstra S, Lu B, Preisser JS et al. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics* 2010; 66(4): 1230–1237.
- Hooper R, Teerenstra S, de Hoop E et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* 2016; 35(26): 4718–4728.
- Hemming K, Kasza J, Hooper R, Forbes A, Taljaard M. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator, *International Journal of Epidemiology*, Volume 49, Issue 3, June 2020, Pages 979–995, https://doi.org/10.1093/ije/dyz237
- Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*. 2020 Jul 6:962280220932962. Doi: 10.1177/0962280220932962. Epub ahead of print. PMID: 32631142.

Clustered dyadic data:

- Moerbeek, M., & Teerenstra, S. (2016). Power analysis of trials with multilevel data. Boca Raton, FL: CRC Press
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). Dyadic data analysis. New York, NY: Guilford Press

Clustered multivariate outcomes:

- Turner RM, Omar RZ, Thompson SG. Modelling multivariate outcomes in hierarchical data, with application to cluster randomised trials. *Biom J*. 2006 Jun;48(3):333–45. Doi: 10.1002/bimj.200310147. PMID: 16845899.
- Li D, Cao J, Zhang S. Power analysis for cluster randomized trials with multiple binary co-primary endpoints. *Biometrics*. 2019 Dec 24.

Multivariate longitudinal data:

- Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res*. 2014;23(1):42–59. doi:10.1177/0962280212445834

References

1. Baier RR, Mitchell SL, Jutkowitz E, et al. Identifying and Supporting Nonpharmacological Dementia Interventions Ready for Pragmatic Trials: Results From an Expert Workshop. J Am Med Dir Assoc 2018;19(7):560–562. [PubMed: 29656839]
2. Zwarenstein M ‘Pragmatic’ and ‘explanatory’ attitudes to randomised trials. Journal of the Royal Society of Medicine 2017; 110(5), 208–218. [PubMed: 28504072]
3. Mitchell SL, Mor V, Harrison J, et al. Embedded Pragmatic Trials in Dementia Care: Realizing the Vision of the NIA IMPACT Collaboratory. J Am Geriatr Soc 2020; 68: S1–S7.

4. Allore HG, Goldfeld KS, Gutman R, et al. Statistical Considerations for Embedded Pragmatic Clinical Trials in People Living with Dementia. J Am Geriatr Soc 2020; 68: S68–S73. [PubMed: 32589276]
5. Andridge RR, Shoben AB, Muller KE, et al. Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. Stat Med 2014;33(13):2178–90. [PubMed: 24399701]
6. Donner A and Klar N. Design and Analysis of Cluster Randomization Trials in Health Research New York. New York: Oxford University Press, 2000.
7. Fitzmaurice GM, Laird NM and Ware JH. Applied Longitudinal Analysis, Second Edition, by: Hoboken, NJ: Wiley, 2011, ISBN 978–0-470–38027-7, xxv + 701 pp.
8. Hooper R and Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. BMJ 2015; 350: h2925. [PubMed: 26055828]
9. Reed RG, Butler EA and Kenny DA. Dyadic Models for the Study of Health. Soc Personal Psychol Compass 2013; 7: 228–245.
10. Kenny DA. The effect of nonindependence on significance testing in dyadic research. Personal Relationships 1995; 2: 67–75.
11. Hardy SE, Allore H and Studenski SA. Missing data: a special challenge in aging research. J Am Geriatr Soc 2009; 57(4):722–729. [PubMed: 19220562]
12. Van Ness PH, Charpentier PA, Ip EH, et al. Gerontologic biostatistics: the statistical challenges of clinical research with older study participants. J Am Geriatr Soc 2010; 58(7):1386–92. [PubMed: 20533963]
13. Turner RM, Omar RZ and Thompson SG. Modelling multivariate outcomes in hierarchical data, with application to cluster randomised trials. Biom J 2006; 48(3): 333–345. [PubMed: 16845899]
14. Li D, Cao J and Zhang S. Power analysis for cluster randomized trials with multiple binary co-primary endpoints. Biometrics 2020; 76(4): 1064–1074. [PubMed: 31872435]
15. Food and Drug Administration. Guidance for industry: Alzheimer's disease: developing drugs for the treatment of early stage disease. U.S. Department of Health and Human Services Food and Drug Administration, Rockville, MD, USA, 2013.
16. Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 Statement: extension to cluster randomised trials. BMJ 2012; 345: e5661. [PubMed: 22951546]
17. Hemming K, Taljaard M, McKenzie JE, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 Statement with explanation and elaboration. BMJ 2018; 363: k1614. [PubMed: 30413417]
18. Kahan Brennan C and Morris Tim P. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis BMJ 2012; 345: e5840 [PubMed: 22983531]
19. Taljaard M, Weijer C, Grimshaw JM, et al. Developing a framework for the ethical design and conduct of pragmatic trials in healthcare: a mixed methods research protocol. Trials 2018; 19: 525. [PubMed: 30261933]
20. Taljaard M, McDonald S, Nicholls SG, et al. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. J Clin Epidemiol 2020; 124: 75–84. [PubMed: 32407765]
21. Nicholls SG, Carroll K, Hey SP, et al. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. J Clin Epidemiol 2021; 137: 45–57. [PubMed: 33789151]
22. ALOIS A comprehensive, open-access register of dementia studies, https://alois.medsci.ox.ac.uk/about-alois. (Accessed 4 December 2020).
23. Diaz-Ordaz K, Froud R, Sheehan B, et al. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. BMC Med Res Methodol 2013; 13: 127. [PubMed: 24148859]
24. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: Designing trials that are fit for purpose BMJ 2015, 350: h2147. [PubMed: 25956159]
25. Nicholls SG, Zwarenstein M, Hey SP, et al. The importance of decision intent within descriptions of pragmatic trials. J Clin Epidemiol 2020; 125: 30–37. [PubMed: 32422248]

26. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008; 337: a2390. [PubMed: 19001484]
27. NIA IMPACT Collaboratory: Transforming Dementia Care, https://impactcollaboratory.org/ (Accessed 13 December 2020).
28. Lee EJ, Patel A, Acedillo RR, et al. Cultivating Innovative Pragmatic Cluster-Randomized Registry Trials Embedded in Hemodialysis Care: Workshop Proceedings From 2018. Can J Kidney Health Dis 2019; 6:2054358119894394, e collection 2019. doi:10.1177/2054358119894394 [PubMed: 31903190]
29. Murray DM, Taljaard M, Turner EL, et al. Essential Ingredients and Innovations in the Design and Analysis of Group-Randomized Trials. Annu Rev Public Health 2020; 41: 1–19. [PubMed: 31869281]
30. Xiong C, Yu K, Gao F, et al. Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. Clin Trials 2005; 2(5): 387–93. [PubMed: 16317808]
31. Sozu T, Sugimoto T and Hamasaki T. Sample Size Determination in Superiority Clinical Trials With Multiple Co-Primary Correlated Endpoints. J Biopharm Stat 2011; 21(4): 650–668. [PubMed: 21516562]
32. Lafaye de Micheaux P, Liquet B, Marque S, et al. Power and sample size determination in clinical trials with multiple primary continuous correlated endpoints. J Biopharm Stat 2014; 24(2): 378–397. [PubMed: 24605975]
33. Moulton LH. Covariate-based constrained randomization of group-randomized trials. Clin Trials 2004; 1(3): 297–305. [PubMed: 16279255]
34. Li F, Lokhnygina Y, Murray DM, et al. An evaluation of constrained randomization for the design and analysis of group-randomized trials. Stat Med 2016, 35(10): 1565–1579. [PubMed: 26598212]
35. Hossain A, Diaz-Ordaz K and Bartlett JW. Missing continuous outcomes under covariate dependent missingness in cluster randomised trials. Stat Methods Med Res 2016; 26: 1543–1562. [PubMed: 27177885]
36. Hossain A, DiazOrdaz K and Bartlett JW. Missing binary outcomes under covariate-dependent missingness in cluster randomised trials. Stat Med 2017; 36: 3092–3109. [PubMed: 28557022]
37. Turner EL, Yao L, Li F, et al. (2020). Properties and pitfalls of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes under covariate-dependent missingness. Stat Methods Med Res 2020; 29(5): 1338–1353. [PubMed: 31293199]
38. Hemming K, Kasza J, Hooper R, et al. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. Int J Epidemiol 2020; 49(3): 979–995. [PubMed: 32087011]
39. Korevaar E, Kasza J, Taljaard M, et al. Intra-cluster correlations from the CLustered OUtcome Dataset bank to inform the design of longitudinal cluster trials. Clin Trials. 2021 Jun 4:17407745211020852. doi: 10.1177/17407745211020852. Epub ahead of print.

Table 1.

Potential sources of clustering and consequences of ignoring the intracluster correlation in the sample size calculation and analysis

Type of clustering	Explanation	Implications for calculated sample size when correlation is ignored[a]	Implications for analysis when correlation is ignored[a]
Cluster randomized trial	Observations from multiple individuals belonging to the same cluster are usually positively correlated	Too small	Increased risk of Type I error
Individually randomized group treatment trial	Multiple observations from individuals receiving treatment in the same group or by the same interventionist are usually positively correlated	Too small	Increased risk of Type I error
Individually randomized parallel arm trial with repeated measures on the same individual after intervention (treatment is between-subject effect)	Multiple repeated measures on the same individual are usually positively correlated	Too small[b]	Increased risk of Type I error[b]
Individually randomized cross-over trial (treatment is within-subject effect)	Multiple repeated measures on the same individual are usually positively correlated	Too large[b]	Increased risk of Type II error[b]
Individually randomized parallel arm trial with repeated measures on the same individual before and after intervention	Multiple repeated measures on the same individual are usually positively correlated	Either too large or too small[c]	Increased risk of Type I or Type II error[c]
Dyadic outcome with both members of the pair allocated to the same intervention (treatment is a between-dyads effect)	Measurements on two individuals in a dyad (e.g., patient-caregiver) may be positively or negatively correlated	Too small if correlation is positive[d] Too large if correlation is negative[d]	Increased risk of Type I error if correlation is positive[d] Increased risk of Type II error if correlation is negative[d]
Multivariate or co-primary endpoints when the trial is designed to evaluate a joint effect on all the endpoints	Multiple components of the multivariate outcome are usually positively correlated. Power decreases as the number of endpoints being evaluated increases. To maintain nominal power, the sample size should be increased. Accounting for the correlation can lessen the increase in the sample size.	Too large	Increased risk of Type II error

[a]Note: we consider here a superiority trial design with a continuous or binary endpoint and a single source of clustering; we further consider an analysis involving all measurements but ignoring the intracluster correlations.

[b]Assumes all measurements are analyzed, ignoring the correlation. Failing to utilize available repeated measures (i.e., basing the sample size calculation or analysis on a single measurement per subject) has the opposite effect: it means that the sample size may be larger than required and the analysis may be statistically inefficient.

[c]Depending on the number of repeated measurements and the strength of the within-subject correlation.

[d]Assumes all measurements are analyzed, ignoring the correlation. Failing to utilize the available pairwise measurements (i.e., basing the sample size calculation or analysis on a single member of the dyad when the observations are positively correlated) has the opposite effect: the sample size may be larger than required and the analysis may be statistically inefficient..

Table 2.

General characteristics of trials included in the review (N=62)

Characteristic	Frequency (%)[a]
Country [b]	
Canada	2 (3.2%)
United States of America	13 (21.0%)
United Kingdom	14 (22.6%)
European Union	17 (27.4%)
Australia or New Zealand	5 (8.1%)
Low and Middle-Income Country	4 (6.5%)
Other	9 (14.5%)
Setting	
Primary care	8 (12.9%)
Hospital care	6 (9.7%)
Nursing homes	28 (45.2%)
Communities	15 (24.2%)
Other	5 (8.1%)
Intervention [b]	
Educational intervention targeting health professionals	19 (30.6%)
Quality improvement targeting organization/health care system	17 (27.4%)
Patient non-pharmacological intervention	29 (46.8%)
Patient pharmacological intervention	3 (4.8%)
Any intervention targeting caregiver only	12 (19.4%)
Any intervention targeting the patient-caregiver dyad	8 (12.9%)
Control arm(s)	
No active intervention (usual care)	40 (64.5%)
Scaled down version of active intervention	14 (22.5%)
Placebo or sham intervention	3 (4.8%)
Other active intervention (i.e., head to head comparison)	3 (4.8%)
Other[c]	2 (3.2%)
Data collection [b]	
Review of medical records	21 (33.9%)
Routinely collected health administrative data	2 (3.2%)
Mental or physical examination not required for normal care	33 (53.2%)
Patient-focused questionnaires completed by patient and/or caregiver	45 (72.6%)
Caregiver-focused questionnaires	22 (35.5%)
Health professional questionnaires	4 (6.5%)
Direct observation	3 (4.8%)

Characteristic	Frequency (%)[a]
Single center or multicenter trial?	
Single center	5 (8.1%)
Multicenter	56 (90.0%)
Unclear	1 (1.6%)
Number of individuals (or dyads) randomized	
Mean (Standard deviation)	574 (1958)
Median (Q1 to Q3)	267 (140 to 402)
Range	101 to 15,574
Journal impact factor	
Median (Q1 to Q3)	3.6 (2.8 to 5.0)
Range	0.9 to 27.6

[a] Entries are frequency (%) unless otherwise indicated

[b] A trial can belong to multiple categories; thus, numbers don't add up to 100%

[c] One study had two control arms: sham and usual care; another study had two control arms: scaled down version of active intervention and usual care

Table 3.

Trial design features (N=62) based on the identified primary trial outcome

Characteristic	Frequency (%)[a]
Type of design (unit of randomization)	
Individually randomized	15 (24.2%)
Cluster randomized	38 (61.3%)
Individually randomized group treatment trial	9 (14.5%)
Trial design for comparing interventions	
Parallel arm	55 (88.7%)
Factorial	4 (6.5%)
Cross-over	1 (1.6%)
Stepped wedge	2 (3.2%)
Method of random allocation	
All trials **(N=62)**	
Completely randomized	24 (38.7%)
Stratified	27 (43.5%)
Pair-matched	9 (14.5%)
Stratified and minimization	2 (3.2%)
Cluster randomized trials **(N=38)**	
Completely randomized	15 (39.5%)
Stratified	13 (34.2%)
Pair-matched	9 (23.7%)
Stratified and minimization	1 (2.6%)
Primary or co-primary outcome(s) clearly identified?	
Yes	50 (80.6%)
No or unclear	12 (19.4%)
Type of primary outcome variable [b]	
Continuous	50 (80.6%)
Binary	9 (14.5%)
Time to event	3 (4.8%)
Nature of the primary outcome [b]	
Univariate outcome measured on the patient	39 (62.9%)
Univariate outcome measured on the caregiver	6 (9.7%)
Bivariate outcome measured on both the patient and caregiver	5 (8.1%)
Multivariate outcome measured on the same participant	12 (19.4%)
Measurement schedule for the primary outcome [b]	
One post-intervention measurement only	8 (12.9%)

Characteristic	Frequency (%)[a]
One pre- and one post-intervention measurement	12 (19.4%)
More than one post-intervention measurement only	4 (6.5%)
More than one pre- and post-intervention measurements	38 (61.3%)
If cluster randomized design with repeated measures (N=34), type of trial design at individual-level	
Closed cohort	32 (94.1%)
Cross-sectional	1 (2.9%)
Open cohort	1 (2.9%)
Size of cluster randomized trials (N=38)	
Number of clusters randomized	
Not reported	1
Mean (Standard deviation)	31 (31)
Median (Q1 to Q3)	22 (15 to 34)
Range	2 to 168
Cluster size (number of individuals or dyads per cluster)	
Mean (Standard deviation)	43 (11)
Median (Q1 to Q3)	14 (8 to 29)
Range	3 to 708

[a]Entries are frequency (%) unless otherwise indicated

[b]If primary outcome was not identified in the report, reviewers nevertheless attempted to identify a “primary” outcome, using the outcome used in sample size calculation, or if sample size calculation was not reported, the first outcome listed under ‘Objectives’.

Table 4.

Conduct and reporting of sample size or power calculations (N=62)

Characteristic	Frequency (%)
Sample size / power calculations presented?	
Yes[a]	50 (80.6%)
No	12 (19.4%)
Did sample size / power calculation account for clustering?	
Cluster randomized trial **(N=38)**	
Yes: Participant-level accounting for clustering	20 (52.6%)
Yes: Cluster-level	2 (5.3%)
No: Participant -level without accounting for clustering	5 (13.2%)
Unclear whether accounted for clustering	1 (2.6%)
No sample size calculation presented	10 (26.3%)
Individually randomized group treatment trial **(N=9)**	
Yes: Participant-level accounting for clustering	3 (33.3%)
No: Participant -level without accounting for clustering	6 (66.7%)
No sample size calculation presented	0
Did sample size/ power calculation account for repeated measures? (N=54)	
Yes	9 (16.7%)
No	33 (61.1%)
No sample size calculation presented	12 (22.2%)
Was correlation in dyadic or multivariate outcome accounted for in sample size calculation? (N=17)	
Yes	0
No	12 (70.6%)
No sample size calculation presented	5 (29.4%)

[a]Includes one trial presenting a sample size calculation based on an unspecified outcome

Table 5.

Methods of analysis and reporting of correlation coefficients (N=62)

Characteristic	Frequency (%)
Did primary analysis of primary outcome account for clustering?	
Cluster randomized trials **(N=38)**	
Yes: Analysis was at individual-level accounting for clustering	20 (52.6%)
Yes: Analysis was at cluster-level	3 (7.9%)
Yes: Other[a]	2 (5.3%)
No: Analysis was at individual-level not accounting for clustering	11 (28.9%)
No: Clustering effect was small; proceeded without accounting for clustering	1 (2.6%)
Unclear whether accounted for clustering	1 (2.6%)
Individually randomized group treatment trial **(N=9)**	
Yes: Analysis was at individual-level accounting for clustering	1 (11.1%)
No: Analysis was at individual-level not accounting for clustering	7 (77.8%)
No: Clustering effect was small; proceeded without accounting for clustering	1 (11.1%)
Were repeated measures utilized in primary analysis of the primary outcome? (N=54)	
Yes	35 (64.8%)
No	19 (35.2%)
Was correlation in dyadic or multivariate outcome accounted for in analysis? (N=17)	
Yes	1 (5.9%)
No	16 (94.1%)
Reported any correlation coefficients?	
Trials with potential intracluster correlation (N=47)	17 (36.2%)
Trials with potential correlation over time (N=54)	0
Trials with potential correlation between multivariate outcomes (N=17)	0
Methods of analysis	
Simple method (e.g., t-test, chi-squared test)	9 (14.5%)
Generalized Estimating Equations (GEE)	11 (17.7%)
Mixed-effects regression	22 (35.5%)
Fixed-effects regression	16 (25.8%)
Other (e.g., structural equation modeling, MANCOVA, two-stage method)	4 (6.5%)

[a]One trial conducted analyses at both individual-level accounting for clustering and cluster-level; one trial used a two-stage method: in stage one ANCOVA was used within each cluster-pair; in stage two random effects meta-analysis was used

Table 6.

Covariate adjustment and missing data treatment (N=62)

Characteristic	Frequency (%)
If covariates were used in the randomization, did primary outcome analysis adjust for those covariates? (N=38)	
Yes, for all relevant covariates	14 (36.8%)
Yes, but not for all relevant covariates	4 (10.5%)
No	18 (47.4%)
Unclear[a]	2 (3.2%)
Does the trial have missing outcomes or attrition for the primary outcome?	
Yes	59 (95.2%)
No	2 (3.2%)
Unclear	1 (1.6%)
Method used to account for missing data in the primary analysis of the primary outcome (N=59)	
None	34 (57.6%)
Regression adjustment for covariates stated to be associated with missingness	6 (9.7%)
Single imputation	9 (14.5%)
Multiple imputation	9 (14.5%)
Other[b]	1 (1.7%)
Method used to account for missing data in sensitivity analyses of the primary outcome (N=59)	
None	52 (88.1%)
Regression adjustment for covariates stated to be associated with missingness	3 (5.1%)
Single imputation	2 (3.4%)
Multiple imputation	1 (1.7%)
Other[c]	1 (1.7%)
Any method used to account for missing data in primary or sensitivity analyses for the primary outcome?	
Yes	27 (45.8%)
No	32 (54.2%)
Did the trial report a statistically significant effect for the primary outcome(s)?	
Yes	30 (48.4%)
No	29 (46.8%)
Mixed results[d]	3 (4.8%)

[a]One trial presented both adjusted and unadjusted analyses without clearly stating the primary analysis; one trial used stratification but did not define the stratification variables

[b]Primary outcome assessed in “modified intent to treat population” consisting only of those residents with available data at 8 months

[c]Single imputation for primary analysis followed by complete case analysis as sensitivity analysis

[d]One trial did not present results for the overall outcome score, but only for the individual components; one three-arm trial had two co-primary outcomes with mixed results; one trial had three co-primary outcomes with mixed results