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) adj10 (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] 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")

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

[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".

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) _____

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 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
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 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