



Published in final edited form as:

*J Am Geriatr Soc.* 2020 July ; 68(Suppl 2): S68–S73. doi:10.1111/jgs.16616.

## Statistical Considerations for Embedded Pragmatic Clinical Trials in People Living with Dementia

Heather G. Allore, PhD<sup>\*,†</sup>, Keith S. Goldfeld, DrPH<sup>‡</sup>, Roe Gutman, PhD<sup>§</sup>, Fan Li, PhD<sup>\*</sup>, Joan K. Monin, PhD<sup>¶</sup>, Monica Taljaard, PhD<sup>||,\*\*</sup>, Thomas G. Travison, PhD<sup>††,‡‡</sup>

<sup>\*</sup>Department of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut;

<sup>†</sup>Section of Geriatrics, Department of Internal Medicine, School of Medicine, Yale University, New Haven, Connecticut;

<sup>‡</sup>Division of Biostatistics, Department of Population Health, NYU Grossman School of Medicine, New York, New York;

<sup>§</sup>Department of Biostatistics, Brown University School of Public Health, Providence, Rhode Island;

<sup>¶</sup>Department of Social and Behavioral Sciences, School of Public Health, Yale University, New Haven, Connecticut;

<sup>||</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada;

<sup>\*\*</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada;

<sup>††</sup>Division of Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts;

<sup>‡‡</sup>Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts.

### Abstract

There is overwhelming need for nonpharmacological interventions to improve the health and well-being of people living with dementia (PLWD). The National Institute on Aging Imbedded Pragmatic Alzheimer's Disease (AD) and AD-Related Dementias Clinical Trials (IMPACT) Collaboratory supports clinical trials of such interventions embedded in healthcare systems. The embedded pragmatic clinical trial (ePCT) is ideally suited to testing the effectiveness of complex interventions in vulnerable populations at the point of care. These trials, however, are complex to conduct and interpret, and face challenges in efficiency (i.e., statistical power) and reproducibility. In addition, trials conducted among PLWD present specific statistical challenges, including difficulty in outcomes ascertainment from PLWD, necessitating reliance on reports by caregivers, and heterogeneity in measurements across different settings or populations. These and other challenges undercut the reliability of measurement, the feasibility of capturing outcomes using pragmatic designs, and the ability to validly estimate interventions' effectiveness in real-world settings. To address these challenges, the IMPACT Collaboratory has convened a Design and Statistics Core, the goals of which are: to support the design and conduct of ePCTs directed toward PLWD and their caregivers; to develop guidance for conducting embedded trials in this population; and to educate quantitative and clinical scientists in the design, conduct, and analysis of these trials. In this article, we discuss some of the contemporary methodological challenges in

---

Address correspondence to Heather G. Allore, PhD. Department of Biostatistics, School of Public Health, Yale University, 300 George St, Suite 775, New Haven, CT 06511. Heather.Allore@Yale.edu.

**Author Contributions:** All authors contributed to the conception and design of the work, contributed to manuscript drafts, critically reviewed the content, and approved the final version of the manuscript.

**Conflict of Interest:** None.

this area and develop a set of research priorities the Design and Statistics Core will undertake to meet these goals.

### Keywords

dementia; clinical trial; design; healthcare systems

---

## INTRODUCTION

To advance the general design and conduct of embedded pragmatic clinical trials (ePCTs),<sup>1</sup> the National Institutes of Health Healthcare Systems Collaboratory has generated a substantial knowledge base, noting the technical difficulties in embedding interventions in healthcare systems.<sup>2,3</sup> Complementary literature addresses some relevant challenges in the context of interventions tailored to specific morbidities<sup>4</sup>; incorporation of multiple healthcare systems and stakeholders<sup>5</sup>; measurement via administrative and billing data<sup>6-8</sup>; and joint measurement of participant and caregiver health outcomes.<sup>9</sup>

Despite these advances, important gaps remain in the methodology supporting ePCT trials. For example, novel designs, such as the stepped wedge<sup>10,11</sup> and other multiple-period cluster trials, are increasingly common, but lack a complete methodological foundation or consistent ethical conduct and reporting.<sup>12</sup> Designs testing multicomponent interventions and formally accounting for treatment heterogeneity are likewise understudied in this context, and there exists only limited guidance for developing pilot embedded trials.

Embedded pragmatic trials conducted among people living with dementia (PLWD) present additional statistical challenges meriting special focus. For example, ascertaining outcomes directly from PLWD may be challenging, necessitating reliance on reports by caregivers,<sup>13,14</sup> and robust methods accounting for the dyadic nature of these data are needed. Trials among PLWD also typically encounter the more general difficulties attending studies in aging,<sup>15</sup> including missing data<sup>16</sup>; irregular observation periods; interference from “competing” events, such as death or hospitalization; heterogeneity in measurements across different settings or populations; and the need for proxy responses for participants with limited communicative capacity.<sup>17</sup> These persistent challenges undercut the reliability of measurement and the feasibility of capturing outcomes using pragmatic designs. Methods to enhance the validity and ease of use of the electronic health record (EHR) are hence a persistent need.

The overarching goals of the Design and Statistics Core of the National Institute on Aging Imbedded Pragmatic Alzheimer’s Disease (AD) and AD-Related Dementias (ADRDs) Clinical Trials (IMPACT) Collaboratory are: to be a national resource supporting the design and conduct of ePCTs in healthcare systems testing interventions directed toward PLWD and their caregivers; to develop and disseminate guidance for conducting ePCTs in this vulnerable population; and to educate a cadre of quantitative and clinical scientists to design, conduct, and analyze these trials.

We elucidate and illustrate some of the contemporary methodological challenges in this area using the Collaboratory Demonstration Project Pragmatic Trial of Video Education in Nursing Homes (PROVEN).<sup>19,20</sup> PROVEN is a large parallel arm pragmatic cluster-randomized trial of advance care planning (ACP) video intervention among nursing home residents with advanced illness previously described. In brief, 359 facilities were randomized in approximately 1:2 ratio with intervention facilities trained to offer a video education program designed to address common ACP decisions to nursing home residents and their caregivers. PROVEN hypothesized that the ACP video program would reduce hospital transfer rates in long-stay nursing home patients with advanced illness.

## INTERVENTIONS DELIVERED TO PLWD AND THEIR CAREGIVERS

Trials in AD/ADRD often target both members of the PLWD/caregiver *dyad*, but it may be difficult to determine whether an intervention will obtain greater effectiveness by targeting the PLWD, the caregiver, or both. Studies in multiple illness contexts demonstrate that the measures of health recorded for both participants and caregivers are highly interrelated and influence each other over time.<sup>18,19</sup> In PROVEN, such measures could include evidence of advance directives for patients and caregivers. When these measures are the same for both dyad members, approaches, such as the Actor Partner Interdependence Model, explicitly consider the influence of the participant's and caregiver's health status on future outcomes.<sup>20</sup> This model can be used to estimate intervention effects,<sup>21</sup> but few, if any, ePCTs have exploited that capability in dementia, perhaps because of challenges in obtaining equivalent or complementary measurement in participants and caregivers. To enhance the utility of dyadic methods in ePCTs, some outstanding priorities are: to develop and refine such complementary measures; to develop algorithms leveraging EHR and administrative data that include linked dyads; and to develop methods for analyzing dyadic data in which measurements on participants are increasingly difficult to obtain with time, because of accelerated cognitive decline or other intercurrent complexities.

## MULTICOMPONENT INTERVENTIONS

ePCTs conducted in PLWD are commonly evaluating the effectiveness of a multicomponent behavioral intervention. Participants randomized to clusters in the experimental arm may receive all components, or only those suggested by their current health status, in a "standardly tailored" fashion.<sup>22</sup> For example, the intervention in PROVEN was composed of five ACP decisions, and clinicians could choose which video(s) to show and the administration tool (local tablet or online web link). This approach is attractive for use in an ePCT because it mimics clinical practice in its "pragmatic" delivery of intervention components to match clinical care needs, but portends some analytic challenges. For instance, tailoring of interventions may evolve with time. In addition, there is likely to be heterogeneity<sup>23</sup> in delivery of specific intervention components, and some patients may be exposed to more videos than others. Moreover, in other trials, participants may be indirectly exposed to unassigned components via interaction with the other dyad member. Outstanding methodological problems in this area include consideration of the influence of participant/caregiver interactions on treatment assignment and effect estimation; estimating the

heterogeneity of treatment effect with multicomponent interventions; and developing machinery to estimate sample sizes for tailored interventions (Table 1).

## CLUSTER-RANDOMIZED DESIGNS

Because of the importance of providers and care delivery to the health of PLWD, we often seek to evaluate the effectiveness of interventions on PLWD by assigning these interventions at the level of the facility or healthcare system.<sup>2</sup> Cluster-randomized trials, which assign intact groups—such as all PLWD being treated in a hospital, living in a nursing home, or seen by a particular primary care provider—are well suited to addressing such questions.<sup>24–26</sup>

In cluster trials, if we seek to analyze data at the level of the participant or caregiver—rather than aggregated by cluster—the analysis can become methodologically complex. This is because members of the same cluster share geographical, social, or other connections and their outcomes behavior may be similar, and this so-called *intracluster correlation* must be acknowledged in both design and analysis. In PROVEN, nursing homes have varying transfer rates, which may relate to different nursing homes' cultures. This feature of cluster randomization decreases the effective sample size (i.e., the amount of independent information obtained from the trial), sometimes to a dramatic degree. Thus, cluster trials require larger samples—sometimes much larger—than individually randomized trials.<sup>27</sup> They also require methods of analysis that account for the intracluster correlation.

With the increasing emphasis on ePCTs in recent years, novel cluster-randomized designs, such as the stepped wedge, have becoming increasingly popular.<sup>10</sup> An attractive feature of this design is that all clusters eventually receive the intervention: they start in the control condition, but then gradually cross over to the intervention in randomized order (Figure 1). This characteristic can help facilitate cluster recruitment,<sup>28</sup> and allows the researcher to directly observe within-cluster, and sometimes within-participant, effects of the intervention. By this mechanism, these designs tend to be more efficient than traditional parallel-arm cluster trials, obtaining greater power at a fixed sample size.<sup>29</sup>

A major methodological consideration in a stepped wedge trial is that the intervention effect is confounded with time by design—the control protocol always precedes the intervention protocol (Figure 1)—and failure to account for any secular trends in the analysis may therefore bias intervention effect estimates.<sup>11,30</sup> For example, if PROVEN had been designed as a stepped wedge, natural changes in the transfer rates over time could be confounded with the intervention effect. In addition, there remain substantial methodological challenges to the deployment of this design among people living with AD/ADRD. Methods for determining sample size requirements are not fully known for noncontinuous outcomes in certain settings, especially in the presence of complex correlation structures.<sup>31–34</sup> Furthermore, robust methods of analysis for stepped wedge trials with small numbers of clusters have not been fully developed.

Because ePCTs in AD/ADRD will often consider assessments from not only the PLWD but also their caregiver, it is critical to further develop models that accommodate measurements

on both (and their association with one another) in the context of the stepped wedge or other clustered design. A related area in need of development is the so-called “incomplete” stepped wedge, in which participants’ outcomes are captured in only a subset of all possible data collection periods.<sup>35,36</sup> In principle, this approach reduces participant data collection burden, but optimal design features for maintenance of validity and statistical power are yet to be investigated.

## ADHERENCE TO STUDY PROTOCOL

In randomized trials, the *intention-to-treat* strategy, requiring that all participants are analyzed as randomized, is commonly advocated in estimating the effectiveness of interventions.<sup>37–39</sup> This approach is relevant for embedded pragmatic trials because it mimics imperfect, “real-world” implementation.<sup>40</sup>

When trials experience nonadherence or differential loss to follow-up, this may produce attenuated intervention effect estimates.<sup>38</sup> This can lead researchers to conclude that an intervention is ineffective for all populations, whereas it may be effective for some, in particular, those participants who have sufficient reserve to complete it. Nonadherence with the assigned treatment arm and differential loss to follow-up (including that due to death) may be especially prevalent among PLWD, making adherence and trial participation challenging.<sup>41,42</sup> Moreover, in many ePCTs, the healthcare providers administer the interventions. However, these frontline providers that care for PLWD have limited time in each encounter and may not have the capacity to offer and apply additional experimental intervention, resulting in increasing nonadherence.<sup>43</sup> In PROVEN, nonadherence at the provider level corresponds to the rate at which nursing homes failed to offer the videos to all eligible residents. At the individual level, nonadherence corresponds to eligible residents and caregivers who decline watching the videos. Thus, ePCTs conducted on PLWD should emphasize methods that increase collection of outcomes on all participants and consider postrandomization auxiliary variables and proxy measures of outcomes at both the PLWD/caregiver and the provider levels.<sup>44,45</sup>

The potential outcomes framework of causal inference<sup>46</sup> is useful in conceptualizing the potential influences of noncompliance and loss to follow-up.<sup>47</sup> Under this framework, one estimable quantity of interest is the average effect among individuals who would adhere to both the intervention and control.<sup>48</sup> This idea has been extended to incorporate information on posttreatment variables (e.g., levels of adherence) to refine comparisons between treatment groups and to different randomization settings.<sup>49,50</sup> Generally, maximizing adherence through study design is the preferred approach. However, when adherence is low, specialized analysis and data collection are required.<sup>51</sup>

Important areas of further methodological research (Table 1) include development of adherence measures for PLWD and the integration of such measures in trials; addressing heterogeneity of adherence and follow-up at the participant and cluster level; development of methods that adjust for both participants’ and caregivers’ adherence; and development of postintervention measures and strategies for multicomponent interventions. Because attrition and nonadherence occur postrandomization, statistical and practical assumptions are

necessary to develop proper inferences. Thus, the development of methods to examine the sensitivity and robustness of results derived from ePCTs is an important area of further research.

## RELIABLE AND RESPONSIVE OUTCOMES

Clinical trials increasingly emphasize person-centered outcomes, such as quality of life, functionality, mood, and caregiver burden.<sup>52</sup> EHR outcomes capture is not structured to solicit outcomes directly from a participant, but rather relies on secondary data sources; as a result, the outcomes are more focused on process measures, like hospitalizations, emergency department visits, or location of death.

Outcomes for trials in dementia should be both *reliable* and *responsive*.<sup>53,54</sup> Reliability reflects consistency of measurement across different raters in the same period or consistency of measurement across limited time for the same participant or caregiver. Responsiveness indicates sensitivity to amelioration or exacerbation in the underlying construct, such as depression or stress. PROVEN used a process measure as the outcome that meets these definitions. However, an outcome that has floor or ceiling effects, such as the Decision Making Involvement Scale,<sup>55</sup> could be non-responsive for the portion of the cohort at the relevant end of that construct's distribution.

In ePCTs, outcomes are likely to be derived from the EHR. When the EHR uses abbreviated versions or subsets of validated scales, reliability and responsiveness cannot be assumed to be inherited from the parent scales. Researchers may need to conduct independent assessments of these modified instruments to ensure that the measured outcomes are appropriate goals, and/or develop transformations that allow estimation of the full scale from the measured portion. Important outstanding problems in this area include integrating person-centered outcomes into the EHR while minimizing participant burden and developing methods to evaluate the reliability and responsiveness of these measures so that they can be used confidently in ePCTs.

## DESIGNING AND PILOTING PRAGMATIC INTERVENTIONS

In its mission to transform the delivery, quality, and outcomes of care provided to PLWD and their caregivers by accelerating the testing and adoption of evidence-based interventions within healthcare systems, the IMPACT Collaboratory funds 1-year pilot studies of ePCTs. The Design and Statistics Core will play a major role in supporting applicants in the design and statistical analysis of their pilot applications and funded studies, although much more work needs to be done to understand how to optimally design pilots to ensure the successful progression to a full-scale ePCT.

Designing pilot studies for pragmatic interventions can be challenging, especially when applying multicomponent interventions, and the future trial will use a complex design, such as cluster randomization. The U.K. Medical Research Council<sup>56</sup> provides guidelines for the development and evaluation of complex interventions. The intervention development phase focuses on developing the intervention to the point where it can be reasonably expected to have a meaningful effect. This phase often includes systematic reviews of the existing



evidence base and development of a theoretical basis for the mechanism by which the intervention operates, possibly using interviews with key stakeholders, such as healthcare system managers, care providers, PLWD, and their caregivers. In the pilot testing phase, the focus is on conducting a smaller-scale study in preparation for embarking on a full-scale study (i.e., fully powered trial) of the intervention. This may not necessitate randomization. The main goal is to determine whether and how to proceed with the full-scale ePCT, and not to estimate the intervention's effectiveness.<sup>57-60</sup> As such, a pilot trial for an ePCT should consider areas of uncertainty specific to the full-scale evaluation of the intervention in conditions that resemble usual clinical care. For example, in PROVEN, there may be uncertainty regarding: the acceptability of the intervention in the skilled nursing facilities; the success of planned PLWD and caregiver identification and recruitment procedures to ensure representativeness; the sites' ability to deliver the video intervention as planned; PLWD and caregivers' anticipated adherence to the intended intervention in usual care conditions; availability and completeness of routinely collected data for outcome assessment; and the extent to which the trial can be conducted without interfering in real-world care.

Pilot designs should identify primary and secondary feasibility objectives. Qualitative, quantitative, and mixed methods should be considered when assessing feasibility. Outcomes may or may not include those to be used in the larger trial; however, hypothesis testing for effectiveness or efficacy is not recommended. The treatment difference observed in a pilot should not dictate the effect size that the future trial is powered to detect, although it may provide limited evidence of whether obtaining clinically important thresholds in a larger trial is conceivable. A pilot can usefully provide an estimated standard deviation, mean, or proportion in the control arm to inform the sample size calculation for the future trial, although there could be substantial uncertainty about these estimates if the pilot sample size is small. Confidence intervals around these parameters can be helpful to inform the choice of more conservative values to avoid underpowering the main study.

Although pilot ePCTs will have modest sample sizes, the sample size should nevertheless be justified, and this may be driven by the primary feasibility objective. Pilot trials are seldom of sufficient size to reliably estimate parameters required for the design of cluster-randomized trials, including variances and the intraclass correlations of outcomes measurement. When routinely collected EHR or administrative data are available on a larger number of sites and participants, researchers should consider leveraging these resources to calculate intraclass correlation coefficients to inform the design of the larger trial.

A final methodological consideration is prespecified criteria to determine whether or how to proceed with a full-scale ePCT. Analysis of pilot trials should set these criteria at design, and their interpretation should address whether progression criteria were met. For example, a pilot trial in advance of PROVEN may have specified a minimum level of adherence to be achieved in delivering the Advance Care Planning Video Program to residents, such that the intervention can plausibly achieve a difference that would affect decision-making.

The manner in which findings from the pilot will influence the design of the larger trial should likewise be considered during planning of the pilot trial.

Investigators conducting a pilot ePCT should register the trial and consider publishing its protocol. Robust guidelines for designing pilot embedded pragmatic trials are a critical need in this area and are currently in development (Table 1).

## CONCLUSION

There is overwhelming need for nonpharmacological interventions to improve the health and well-being of PLWD and their caregivers. The ePCT is ideally suited to testing the effectiveness of complex interventions in vulnerable populations, but this applicability comes at nontrivial cost to operability, efficiency, interpretability, and reproducibility. The IMPACT Collaboratory Design and Statistics Core is focused on biostatistical methods to design, conduct, and analyze embedded pragmatic trials among PLWD; to develop and promulgate best practices in this area; and to provide guidance to researchers planning or conducting such trials. The research priorities identified above are tailored to supporting the IMPACT Collaboratory's mandate to build the nation's capacity to conduct ePCTs within healthcare systems for PLWD and their caregivers.

## ACKNOWLEDGMENTS

**Financial Disclosure:** This work was supported by the National Institute on Aging (NIA) of the National Institutes of Health under Award U54AG063546, which funds NIA Imbedded Pragmatic Alzheimer's Disease (AD) and AD-Related Dementias Clinical Trials Collaboratory. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Sponsor's Role:** The sponsor had no role in designing, drafting, or editing the manuscript.

## REFERENCES

1. Mitchell SL, Mor V, Harrison J, McCarthy EP. Embedded Pragmatic Trials in Dementia Care: Realizing the Vision of the NIA IMPACT Collaboratory. *J Am Geriatr Soc.* 2020;68(S2):S1–S7. [PubMed: 32589280]
2. Cook AJ, Delong E, Murray DM, Vollmer WM, Heagerty PJ. Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core. *Clin Trials.* 2016;13(5):504–512. 10.1177/1740774516646578. [PubMed: 27179253]
3. Weinfurt KP, Hernandez AF, Coronado GD, et al. Pragmatic clinical trials embedded in healthcare systems: generalizable lessons from the NIH Collaboratory. *BMC Med Res Methodol.* 2017;17(1):144 10.1186/s12874-017-0420-7. [PubMed: 28923013]
4. Bruvik FK, Allore HG, Ranhoff AH, Engedal K. The effect of psychosocial support intervention on depression in patients with dementia and their family caregivers: an assessor-blinded randomized controlled trial. *Dement Geriatr Cogn Dis Extra.* 2013;3(1):386–397. 10.1159/000355912. [PubMed: 24348500]
5. Ivers NM, Desveaux L, Presseau J, et al. Testing feedback message framing and comparators to address prescribing of high-risk medications in nursing homes: protocol for a pragmatic, factorial, cluster-randomized trial. *Implement Sci.* 2017;12(1):86 10.1186/s13012-017-0615-7. [PubMed: 28705208]
6. Gravenstein S, Davidson HE, Han LF, et al. Feasibility of a cluster-randomized influenza vaccination trial in U.S. nursing homes: lessons learned. *Hum Vaccin Immunother.* 2017;14(3):736–743. 10.1080/21645515.2017.1398872. [PubMed: 29090982]
7. Gravenstein S, Davidson HE, Taljaard M, et al. Comparative effectiveness of high-dose versus standard-dose influenza vaccination on numbers of US nursing home residents admitted to hospital:



- a cluster-randomised trial. *Lancet Respir Med*. 2017;5(9):738–746. 10.1016/S2213-2600(17)30235-7. [PubMed: 28736045]
8. Mor V, Volandes AE, Gutman R, Gatsonis C, Mitchell SL. PRagmatic trial of video education in nursing homes (PROVEN): the design and rationale for a pragmatic cluster randomized trial in the nursing home setting. *Clin Trials*. 2017;14(2):140–151. 10.1177/1740774516685298. [PubMed: 28068789]
  9. Monin JK, Schulz R, Kershaw TS. Caregiving spouses' attachment orientations and the physical and psychological health of individuals with Alzheimer's disease. *Aging Ment Health*. 2013;17(4):508–516. 10.1080/13607863.2012.747080. [PubMed: 23215899]
  10. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182–191. 10.1016/j.cct.2006.05.007. [PubMed: 16829207]
  11. Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. 2015;350:h2925 10.1136/bmj.h2925. [PubMed: 26055828]
  12. Hemming K, Taljaard M, Marshall T, Goldstein CE, Weijer C. Stepped-wedge trials should be classified as research for the purpose of ethical review. *Clin Trials*. 2019;16(6):580–588. 10.1177/1740774519873322. [PubMed: 31818147]
  13. Reamy AM, Kim K, Zarit SH, Whitlatch CJ. Understanding discrepancy in perceptions of values: individuals with mild to moderate dementia and their family caregivers. *Gerontologist*. 2011;51(4):473–483. 10.1093/geront/gnr010. [PubMed: 21383111]
  14. Schulz R, Cook TB, Beach SR, et al. Magnitude and causes of bias among family caregivers rating Alzheimer's disease patients. *Am J Geriatr Psychiatry*. 2013;21(1):14–25. 10.1016/j.jagp.2012.10.002. [PubMed: 23290199]
  15. 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):1386–1392. 10.1111/j.1532-5415.2010.02926.x. [PubMed: 20533963]
  16. Hardy SE, Allore H, Studenski SA. Missing data: a special challenge in aging research. *J Am Geriatr Soc*. 2009;57(4):722–729. 10.1111/j.1532-5415.2008.02168.x. [PubMed: 19220562]
  17. Roydhouse JK, Gutman R, Keating NL, Mor V, Wilson IB. Proxy and patient reports of health-related quality of life in a national cancer survey. *Health Qual Life Outcomes*. 2018;16(1):6 10.1186/s12955-017-0823-5. [PubMed: 29304818]
  18. Lyons KS, Lee CS. The theory of dyadic illness management. *J Fam Nurs*. 2018;24(1):8–28. 10.1177/1074840717745669. [PubMed: 29353528]
  19. Kenny DA, Kashy DA, Cook WL, Simpson JA. *Dyadic Data Analysis*. 1st ed New York, NY: The Guilford Press; 2006.
  20. Ledermann T, Kenny DA. Analyzing dyadic data with multilevel modeling versus structural equation modeling: a tale of two methods. *J Fam Psychol*. 2017;31(4):442–452. 10.1037/fam0000290. [PubMed: 28165269]
  21. Braithwaite SR, Fincham FD. Computer-based dissemination: a randomized clinical trial of ePREP using the actor partner interdependence model. *Behav Res Ther*. 2011;49(2):126–131. 10.1016/j.brat.2010.11.002. [PubMed: 21159326]
  22. Allore HG, Tinetti ME, Gill TM, Peduzzi PN. Experimental designs for multicomponent interventions among persons with multifactorial geriatric syndromes. *Clin Trials*. 2005;2(1):13–21. 10.1191/1740774505cn067oa. [PubMed: 16279575]
  23. Allore H, McAvay G, Vaz Fragoso CA, Murphy TE. Individualized absolute risk calculations for persons with multiple chronic conditions: embracing heterogeneity, causality, and competing events. *Int J Stat Med Res*. 2016; 5(1):48–55. 10.6000/1929-6029.2016.05.01.5. [PubMed: 27076862]
  24. Esserman D, Allore HG, Trivison TG. The method of randomization for cluster-randomized trials: challenges of including patients with multiple chronic conditions. *Int J Stat Med Res*. 2016;5(1):2–7. [PubMed: 27478520]
  25. Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part I-design. *Am J Public Health*. 2017;107(6):907–915. 10.2105/AJPH.2017.303706. [PubMed: 28426295]

26. Turner EL, Prague M, Gallis JA, Li F, Murray DM. Review of recent methodological developments in group-randomized trials: part 2-analysis. *Am J Public Health*. 2017;107(7):1078–1086. 10.2105/AJPH.2017.303707. [PubMed: 28520480]
27. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol*. 2006;35(5):1292–1300. 10.1093/ije/dyl1129. [PubMed: 16943232]
28. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391 10.1136/bmj.h391. [PubMed: 25662947]
29. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. *J Clin Epidemiol*. 2016;69: 137–146. 10.1016/j.jclinepi.2015.08.015. [PubMed: 26344808]
30. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials*. 2017;18(1):101 10.1186/s13063-017-1833-7. [PubMed: 28259174]
31. Kasza J, Hemming K, Hooper R, Matthews J, Forbes AB. Impact of nonuniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*. 2019;28(3):703–716. 10.1177/0962280217734981. [PubMed: 29027505]
32. Li F, Forbes AB, Turner EL, Preisser JS. Power and sample size requirements for GEE analyses of cluster randomized crossover trials. *Stat Med*. 2019;38(4):636–649. 10.1002/sim.7995. [PubMed: 30298551]
33. Li F, Turner EL, Preisser JS. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics*. 2018;74(4):1450–1458. 10.1111/biom.12918. [PubMed: 29921006]
34. Design Li F. and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure. *Stat Med*. 2020;39(4): 438–455. 10.1002/sim.8415. [PubMed: 31797438]
35. Kasza J, Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics*. 2019;75(1):144–152. 10.1111/biom.12959. [PubMed: 30051909]
36. Kasza J, Taljaard M, Forbes AB. Information content of stepped-wedge designs when treatment effect heterogeneity and/or implementation periods are present. *Stat Med*. 2019;38(23):4686–4701. 10.1002/sim.8327. [PubMed: 31321806]
37. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med*. 1991;10(10): 1595–1605. 10.1002/sim.4780101011. [PubMed: 1947515]
38. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther*. 1995;57(1):6–15. 10.1016/0009-9236(95)90260-0. [PubMed: 7828382]
39. U.S. Food and Drug Administration. Research C for DE and E9 Statistical Principles for Clinical Trials. U.S. Food and Drug Administration <http://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9-statistical-principles-clinical-trials>. Published April 5, 2019. Accessed February 5, 2020.
40. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48–55. 10.1177/1740774511420743. [PubMed: 21948059]
41. Coley N, Gardette V, Toulza O, et al. Predictive factors of attrition in a cohort of Alzheimer disease patients: the REAL.FR study. *Neuroepidemiology*. 2008;31(2):69–79. 10.1159/000144087. [PubMed: 18622142]
42. National Research Council (US) Panel on Handling Missing Data in Clinical Trials. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: National Academies Press (US); 2010 <http://www.ncbi.nlm.nih.gov/books/NBK209904/>. Accessed February 5, 2020.
43. Goy E, Kansagara D, Freeman M. A Systematic Evidence Review of Interventions for Non-Professional Caregivers of Individuals with Dementia. Washington, DC: Department of Veterans Affairs (US); 2010 <http://www.ncbi.nlm.nih.gov/books/NBK49194/>. Accessed February 5, 2020.
44. Mustillo S, Kwon S. Auxiliary variables in multiple imputation when data are missing not at random. *J Math Sociol*. 2015;39(2):73–91. 10.1080/0022250X.2013.877898.

45. Cornish RP, Macleod J, Carpenter JR, Tilling K. Multiple imputation using linked proxy outcome data resulted in important bias reduction and efficiency gains: a simulation study. *Emerg Themes Epidemiol.* 2017;14:14. 10.1186/s12982-017-0068-0. [PubMed: 29270206]
46. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat.* 1978;6(1):34–58. 10.1214/aos/1176344064.
47. Little R, Kang S. Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Stat Med.* 2015;34(16):2381–2390. 10.1002/sim.6352. [PubMed: 25363683]
48. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996;91(434):444–455. 10.2307/2291629.
49. Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika.* 1999;86(2):365–379. 10.1093/biomet/86.2.365.
50. Forastiere L, Mealli F, VanderWeele TJ. Identification and estimation of causal mechanisms in clustered encouragement designs: disentangling bed nets using Bayesian principal stratification. *J Am Stat Assoc.* 2016;111(514): 510–525. 10.1080/01621459.2015.1125788. [PubMed: 28008210]
51. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med.* 2017;377(14):1391–1398. 10.1056/NEJMs1605385. [PubMed: 28976864]
52. Moniz-Cook E, Vernooij-Dassen M, Woods R, et al. A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging Ment Health.* 2008;12(1):14–29. 10.1080/13607860801919850. [PubMed: 18297476]
53. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998;2(14):iiv, 1–74.
54. Holden SK, Jones WE, Baker KA, Boersma IM, Kluger BM. Outcome measures for Parkinson's disease dementia: a systematic review. *Mov Disord Clin Pract.* 2016;3(1):9–18. 10.1002/mdc3.12225. [PubMed: 26998505]
55. Feinberg LF, Whitlatch CJ. Decision-making for persons with cognitive impairment and their family caregivers. *Am J Alzheimers Dis Other Demen.* 2002;17(4):237–244. 10.1177/153331750201700406. [PubMed: 12184513]
56. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ.* 2008;337:a1655. 10.1136/bmj.a1655. [PubMed: 18824488]
57. Eldridge SM, Lancaster GA, Campbell MJ, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLoS One.* 2016;11(3):e0150205. 10.1371/journal.pone.0150205. [PubMed: 26978655]
58. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract.* 2004;10(2):307–312. 10.1111/j.2002.384.doc.x. [PubMed: 15189396]
59. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol.* 2010;10:1. 10.1186/1471-2288-10-1. [PubMed: 20053272]
60. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? a review of current practice and editorial policy. *BMC Med Res Methodol.* 2010;10:67. 10.1186/1471-2288-10-67. [PubMed: 20637084]

	Period 1	Period 2	Period 3	Period 4	Period 5
Cluster 1					
Cluster 2					
Cluster 3					
Cluster 4					
Cluster 5					
Cluster 6					
Cluster 7					
Cluster 8					

**Figure 1.** Depiction of a stepped wedge cluster-randomized trial, with eight clusters and five time periods. White cells (lower left) denote time periods for which the control condition is administered, whereas blue cells (upper right) denote periods for which the intervention is delivered. Clusters, which might be collections of people living with dementia in specific settings or affiliated with specific care providers, range from two to four periods in their duration of exposure to intervention.

**Selected Research Priorities to Address Contemporary Challenges in Designing and Analyzing Embedded Pragmatic Trials Among People with Dementia and Their Caregivers**

**Table 1.**

<b>Challenge</b>	<b>Research priorities</b>
Evaluating multicomponent interventions	Methods that consider the influence of participant/caregiver interactions on treatment assignment and effect estimation and clarify population heterogeneity in treatment effects Methods or simulations informing sample size computations
Testing interventions on PLWD/caregiver dyads	Measures capturing complementary information from PLWD and caregivers Methods to clarify the influence of PLWD/caregiver interactions on treatment assignment and intervention effects Algorithms to identify caregivers in the electronic health record and link PLWD and caregiver data Models to estimate intervention effects in the context of secular declines in measurement quality
Developing methods and tools for contemporary cluster designs	Reliable, responsive, and reproducible measurement protocols leveraging the electronic health record and administrative data Sample size tools for noncontinuous outcomes and complex correlation structures Valid and efficient designs that reduce participant burden
Capturing protocol adherence in estimating effectiveness	Measurement for, and models incorporating the effects of, adherence for participants and caregivers Methods to address heterogeneity of adherence and follow-up at the participant and facility level Methods or simulations supporting sensitivity analyses, capturing the effects of adherence and attrition
Creating robust standards for embedded trials	Guidelines for design and reporting of embedded pilot trials

Abbreviation: PLWD, people living with dementia.