Guidance for Proxy Data and Dyadic Analysis

Purpose of this Guidance Document

This document provides guidance on what dyadic data is, how it is different from proxy data, and what information can be gained by collecting and analyzing dyadic data compared to proxy data when conducting embedded pragmatic clinical trials (ePCTs) that include people living with dementia and care partners.

Dyadic Data

Data **about each member** of the dyad, ideally using the **same measure or equivalent versions of the same measure**, ideally at the **same time**, and ideally **self-reported**, e.g., self-reported quality of life for a person living with dementia and self-reported quality of life for their care partner. Dyadic analysis addresses the inference around and interpretation of these data as outcomes (dependent variables).

Proxy Data

Data reflecting **the experience of a person living with dementia as reported by the care partner as if they were the person living with dementia**, e.g., a care partner's perception of the quality of life of the person living with dementia.

Table of Contents

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

This work is supported by grant U54AG063546 from the National Institute on Aging

Constructs Often Measured for Both People Living with Dementia and Care Partners

Intervention studies of dyads have commonly assessed the following with self-report measures:

Psychological Health	Physical Health
Resources (e.g., Social Support)	Positive and Negative Coping Strategies

Analyzing Dyadic Data

Dyadic outcome data are collected under a by-design violation of the assumption of independence required for traditional data analysis similar to repeated measures or cluster sampling designs. Failure to accommodate this violation will produce biased measures of variation and specifically the estimates of standard errors necessary to construct confidence interval estimates of means and associations.

Methods that acknowledge and account for the dependence in outcomes between members of the dyad, such as **Multilevel Models (MLM)** and **Structural Equation Models (SEM)**, are commonly used for dyadic analysis and described further below (Kenny, Kashy, and Cooke 2006; Ledermann and Kenny 2017). **The Active Partner Independence Model** is an important and commonly employed theoretical framework to which these methods can be applied (APIM; Cook & Kenny, 2005).

Multilevel Modeling

- In MLM, the interdependence of dyadic outcomes is addressed by treating the dyad as a cluster, permitting a hierarchical correlation structure, with Individuals (Level-1) nested within Dyads (Level-2).
- In most longitudinal dyadic analysis, Time is also a Level-1 variable, as it typically does not vary within dyads, but only between dyads.
  - — Having time points that vary between members of the same dyad is uncommon; however, a 3-level model (Time Level-1, Individuals Level-2, Dyads Level-3) could be used if data were collected at different times for each dyad member, but the times must occur in the same range for each person (i.e., the times must overlap). With different time points within dyads, some models, such as a time-varying APIM, cannot be fit.
- Outcomes can be modeled with a shared mean corresponding to a dyad and a difference between dyad members' estimated means by coding dyad roles as -.5 and .5, or with an intercept (and slope if longitudinal) for each dyad member by creating a dummy code variable for each dyad member indicating which member each observation represents (Kenny, Kashy & Cook, 2006).

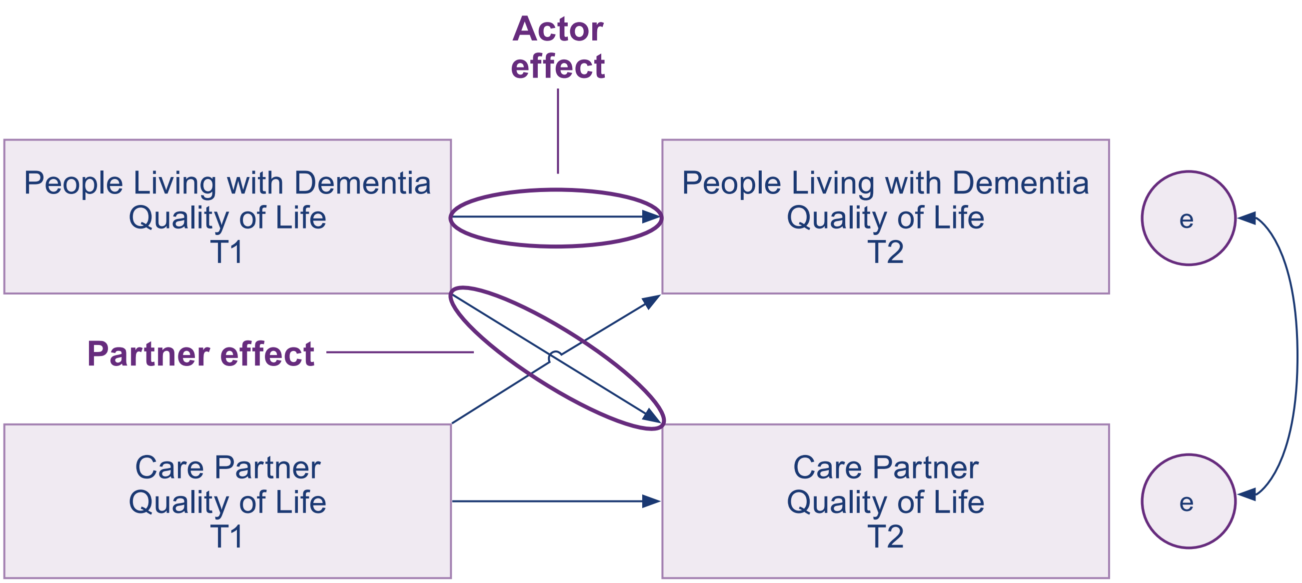
- Interventions are typically treated as Level-2 variables for individually randomized clinical trials (e.g. von Heymann-Horan et al. 2019) to permit estimation of average effects on people living with dementia and a care partner.
    - No clinical trials to date have used a cluster-randomized dyadic design; however, it should be possible to do so using a 3-level model. If cluster-randomization is used, the clusters will add another level of interdependence since outcomes are often correlated within organizations or institutions. In this case, a 3-level model may be used with Individuals (Level-1) nested in Dyads (Level-2), nested in Organizational Clusters (Level-3) with the intervention as a Level-3 variable as it varies between but not within organizations.

Structural Equation Modeling

- In SEM, interdependence of dyadic outcomes is addressed by allowing for correlated errors of each dyad member's outcomes, directly modeling the interdependence between members.
- SEM is a very flexible option for dyadic analysis and can be used to address most of the questions addressed with MLM, as well as more complicated path models.
- In clinical trials, interventions can be included as predictors in numerous ways, including multigroup analysis. (Kenny et al. 2006; Ledermann and Kenny 2017)

Actor-Partner Interdependence Model (APIM)

- The actor–partner interdependence model (APIM) is a model of dyadic relationships that integrates a conceptual view of interdependence with the appropriate statistical techniques for measuring and testing it. One can use MLM or SEM to use the APIM framework.
- The APIM describes the effects of dyad members' own predictors on their own outcomes, as well as the effects of partners on each other's outcomes. This model is the gold standard of dyadic analysis, and can be fit in numerous ways using MLM or SEM. (Gistelinck and Loeys 2020; Kenny et al. 2006; Kenny and Ledermann 2010)
- In clinical trials, treatment could be tested as a moderator of actor effects (e.g. effects of person living with dementia Time 1 quality of life (QOL) on their Time 2 QOL) or partner effects (e.g. effects of person living with dementia Time 1 QOL on care partner Time 2 QOL).



When Data Don't Require Dyadic Analysis

- **When data from a person living with dementia and their care partner are used as *independent variables***, they can simply be included as predictors in a model. In this case, you do not have to worry about interdependence of independent variables. For example, if you are interested in testing whether a person living with dementia or a care partner's adherence to an intervention predicts person living with dementia QOL only, you may include both adherence measures as independent variables in your regression model.
- **When you are interested in examining different *dependent variables* for people living with dementia and care partners**, you may perform analyses for each individual outcome.
- **When you have different measures of the same construct (e.g., depressive symptoms) and plan to use them as dependent variables for both members of the dyad**, these measures are interdependent. The first step often requires data harmonization to make the measures from people living with dementia and their care partners comparable. Then dyadic analysis using harmonized measures can be used to account for the interdependence. For example, if data reflecting depressive symptoms were collected using the PHQ-9 for people living with dementia and CES-D for care partners, then you could employ the clinical cutoffs for each measure to generate binary outcomes, use T-scores based on normative data, or standardize both measures and use z-scores.

Proxy Data

Why Obtain Proxy Data?

When self-reported data cannot be obtained, proxy data is the next best option. While there is increasing emphasis on including patient-reported outcomes when comparing interventions, some people, especially those living with dementia, may be unable to self-report or their self-reports may be considered unreliable. To avoid substantial missing data, proxy respondents can be asked to report on behalf of people living with dementia.

Problems with Proxy Data

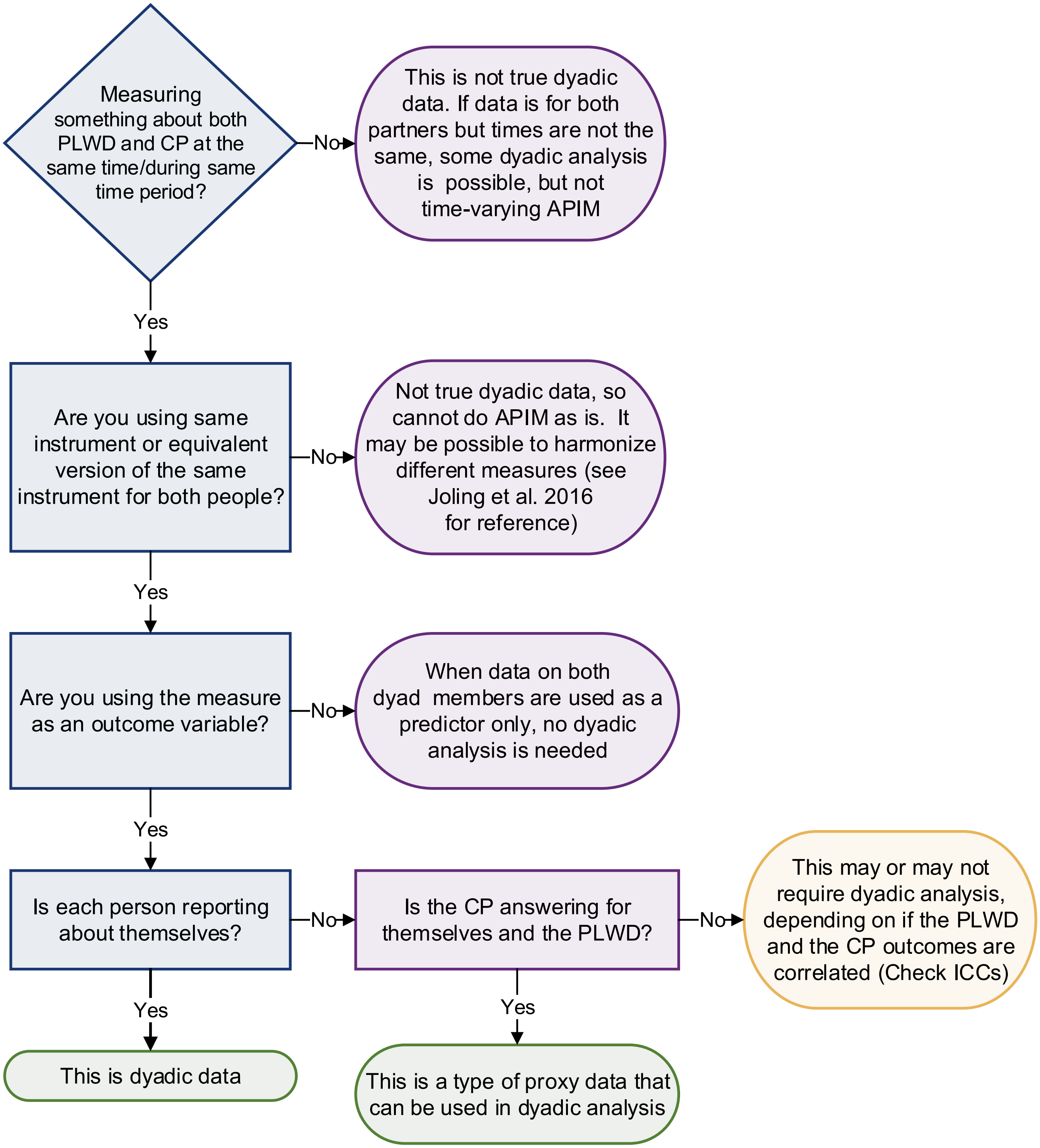
Though proxy data is far better than data that is entirely missing, there are problems with using proxy data that should be addressed. Some of these problems include:

- **Non-response** — Proxy responders may be more prone to non-response on certain items than people living with dementia.
- **Bias** — Proxy responders may view people living with dementia's experience differently than the people living with dementia view it
  - In trials, proxies' responses may be differential across treatment arms.
  - In cluster trials, some sites may observe more proxy responses than others.
  - Proxies may change over time or their interactions may change (e.g., adult child providing drop-in care to having the person living with dementia move in with the care partner).
- **Lack of scientifically sound approaches for analysis of self-report and proxy data**
  - In cluster pragmatic trials, proxies' responses may be differential across sites.
  - Transition from self-report to proxy over time.
  - Typically, studies do not have people living with dementia's self-report and proxy reports measured at the same time points, so it is difficult to determine whether proxies' reports differ from what the people living with dementia would have reported.

Approaches to Addressing Proxy Data

- **Replacement** — assume proxy response is the person living with dementia's response
  - — Easy, but could result in biased estimates. In addition, it assumes Missing Completely at Random (MCAR)
- **Regression adjustments** — Create a single column, y_i^{comb}, that includes the responses of people living with dementia for those that responded and proxy responses for people living with dementia that did not respond. Perform analysis using a generalized linear model adjustments with link function g^{-1} (Wolinsky et al. 2012)
  - — $g(E(y_i^{comb}|X_i, I_i)) = \beta_0 + \beta_1 X_i + \gamma I_i^{proxy}$
  - — $g(E(y_i^{comb}|X_i, I_i)) = \beta_0 + \beta_1 X_i + \gamma_1 I_i^{proxy} + \gamma_2 X_i I_i^{proxy}$ (Reither et al, 2009; Wu et al, 2013)
  - — Easily implemented, but relies on correctness of the model, such that different conclusions may be reached with different models. This procedure assumes Missing at Random (MAR).
- **Propensity scores** — For every person living with dementia with proxy response, identify similar people living with dementia who self-reported. Because describing similarity on many covariates can be complex, a possible solution is to calculate the probability of observing proxy response given the covariates (propensity score). Propensity scores can be used for weighting (Elliot et al, 2008), matching (Ellis et al, 2003; Li et al 2015), or imputation (Roydhouse et al. 2020).
  - — Better than regression adjustment as it uses observed covariates to find similar people living with dementia and methods are in common statistical software, but it may result in poor operating characteristics if people living with dementia's covariates are not highly correlated with the outcome of interest. These methods assume MAR. Another concern arises if covariates are proxy reported. (Roydhouse et al, 2020)
- **Bayesian equating** — Uses Rank preserving or Bayesian non-parametric models to equate "tests" of people living with dementia and proxies (Burgette & Reiter, 2012; Karabatsos & Walker, 2009).
  - — Imputation using flexible modeling is performed once and can be used in multiple analyses. These methods assume MAR and were developed for a single continuous outcome.
  - — It is computationally complex.
- **Equating with Bayesian Item Response Theory (IRT)** — Uses graded response modeling with predictive mean matching (Gu & Gutman, 2017), or a multivariate ordinal probit model (Gu & Gutman, 2019).
  - — Imputation is performed using IRT modeling once and can be used in multiple analyses. Allows for multiple correlated ordinal/binary outcomes. It assumes MAR and requires specialized computations.

Do I Have Dyadic Data?



Dyadic Analysis References

Kenny DA. David A. Kenny [Internet]. 2014 [cited 2024 May 1]. Available from: https://davidakenny.net/

Atkins DC. Using multilevel models to analyze couple and family treatment data: basic and advanced issues. *J Fam Psychol*. 2005 Mar;19(1):98-110. doi: 10.1037/0893-3200.19.1.98. PMID: 15796656.

Cook WL, Kenny DA. The Actor–Partner Interdependence Model: A model of bidirectional effects in developmental studies. *Int J Behav Dev*. 2005 Mar;29(2):101-109. doi: 10.1080/01650250444000405.

Kenny DA, Kashy DA, Cook WL. Dyadic Data Analysis. New York, NY: Guilford Press; 2006.

Kenny DA, Ledermann T. Detecting, measuring, and testing dyadic patterns in the actor-partner interdependence model. *J Fam Psychol*. 2010 Jun;24(3):359-66. doi: 10.1037/a0019651. PMID: 20545409.

Joling KJ, Windle G, Dröes RM, Meiland F, van Hout HP, MacNeil Vroomen J, van de Ven PM, Moniz-Cook E, Woods B. Factors of Resilience in Informal Caregivers of People with Dementia from Integrative International Data Analysis. *Dement Geriatr Cogn Disord*. 2016;42(3-4):198-214. doi: 10.1159/000449131. Epub 2016 Sep 27. PMID: 27669054.

Ledermann T, Kenny DA. Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. *J Fam Psychol*. 2017 Jun;31(4):442-452. doi: 10.1037/fam0000290. Epub 2017 Feb 6. PMID: 28165269.

Lyons KS, Sayer AG. Using multilevel modeling in caregiving research. *Aging & Mental Health*. 2010 Oct 19;9(3):189–195. doi: 10.1080/13607860500089831.

Lyons KS, Lee CS. The Theory of Dyadic Illness Management. *J Fam Nurs*. 2018 Feb;24(1):8-28. doi: 10.1177/1074840717745669. Epub 2018 Jan 20. PMID: 29353528.

Lyons KS, Lee CS. Understanding the family care dyad: A comparison of two multilevel models. *Eur J Cardiovasc Nurs*. 2020 Feb 10;19(2):178-184. doi: 10.1177/1474515120902368.

Lyons KS, Lee CS. A Multilevel Modeling Approach to Examine Incongruent Illness Appraisals in Family Care Dyads Over Time. *J Fam Nurs*. 2020;26(3):229-239. doi: 10.1177/1074840720944439.

Proxy Data References

Burgette LF, Reiter JP. Nonparametric Bayesian Multiple Imputation for Missing Data Due to Mid-Study Switching of Measurement Methods. *J Am Stat Assoc*. 2012 Jul 24;107(498):439–449. doi: 10.1080/01621459.2011.643713.

Elliott MN, Beckett MK, Chong K, Hambarsoomians K, Hays RD. How do proxy responses and proxy-assisted responses differ from what Medicare beneficiaries might have reported about their health care? *Health Serv Res*. 2008 Jun;43(3):833-48. doi: 10.1111/j.1475-6773.2007.00820.x. PMID: 18454770; PMCID: PMC2442242.

Ellis BH, Bannister WM, Cox JK, Fowler BM, Shannon ED, Drachman D, Adams RW, Giordano LA. Utilization of the propensity score method: an exploratory comparison of proxy-completed to self-completed responses in the Medicare Health Outcomes Survey. *Health Qual Life Outcomes*. 2003 Sep 18;1:47. doi: 10.1186/1477-7525-1-47. PMID: 14570594; PMCID: PMC222919.

Gu C, Gutman R. Combining item response theory with multiple imputation to equate health assessment questionnaires. *Biometrics*. 2017 Sep;73(3):990-998. doi: 10.1111/biom.12638. Epub 2016 Dec 9. PMID: 27936287.

Gu C, Gutman R. Development of a common patient assessment scale across the continuum of care: A nested multiple imputation approach. *Ann Appl Stat*. 2019 Mar;13(1): 466-491. doi: 10.1214/18-AOAS1202.

Karabatsos G, Walker SG. A Bayesian Nonparametric Approach to Test Equating. *Psychometrika*. 2008 Dec 20;74:211–232. doi: 10.1007/s11336-008-9096-6.

Li M, Harris I, Lu ZK. Differences in proxy-reported and patient-reported outcomes: assessing health and functional status among medicare beneficiaries. *BMC Med Res Methodol*. 2015 Aug 12;15:62. doi: 10.1186/s12874-015-0053-7. PMID: 26264727; PMCID: PMC4534114.

Reither EN, Utz RL. A procedure to correct proxy-reported weight in the National Health Interview Survey, 1976–2002. *Popul Health Metrics*. 2009 Jan 6;7:2. doi: 10.1186/1478-7954-7-2.

Roydhouse JK, Gutman R, Keating NL, Mor V, Wilson IB. Propensity scores for proxy reports of care experience and quality: are they useful? *Health Serv Outcomes Res Method*. 2020 Mar;20:40–59. doi: 10.1007/s10742-019-00205-4.

Wolinsky FD, Jones MP, Wehby GL. Gathering data from older adults via proxy respondents: research challenges. *J Comp Eff Res*. 2012 Nov;1(6):467-70. doi: 10.2217/cer.12.54. PMID: 24236462.

Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Maria Glymour M. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Dis Assoc Disord*. 2013 Jul-Sep;27(3):207-12. doi: 10.1097/WAD.0b013e31826cfe90. PMID: 22992720; PMCID: PMC3731387.

Citation: Newkirk K, Monin J. Guidance for Proxy Data and Dyadic Analysis. NIA IMPACT Collaboratory; 2024. doi: https://doi.org/10.58234/23678174

This document was prepared with the support of the IMPACT Collaboratory by grant U54AG063546 from the National Institute on Aging.