Fatherhood and Marriage Local Evaluation & Cross-Site Project

Evaluation Technical Assistance Brief

OPRE report #2021-23

JANUARY 2021

Analyzing data on paired partners in program evaluation: Strategies to overcome common challenges

Many healthy marriage and responsible fatherhood (HMRF) programs serve paired partners simultaneously, such as couples seeking to improve their relationship or coparents raising a child together. For example, this is often the case when the focus of the HMRF program is to strengthen the quality of the couple's relationship, and not to change individuals' attitudes, behavior, and decision making about relationships in general (Stanley et al. 2019). To understand how effective such pair-centric programs are in affecting relevant outcomes, it is typically necessary to collect data from both partners.

Evaluators face four key challenges in correctly analyzing data from paired partners:

1. Statistical tests that don't adjust for the interdependence of partners' outcomes will overstate how likely it is that an impact is statistically different than zero; as a result, an evaluation can erroneously conclude that a program had an impact on an outcome.
2. Analyzing partners' data separately can give a false sense of the differences between partners' outcomes or any gender differences in the results if no formal test for these differences is conducted.
3. Impact estimates for some outcomes can be biased if a program affects the number and type of pairs for whom these outcomes can be observed.
4. When some follow-up data are missing for one partner in a pair, dropping that pair (and others with the same issue) from the analysis can yield biased results.

In the rest of this brief, we discuss these challenges in depth and offer some strategies researchers can consider for addressing them. Our discussion assumes that participants enroll in a program and enter the evaluation as a pair (for example, that an experimental evaluation uses random assignment of paired partners to either program or control groups, not individual members of the pair). The strategies discussed below have strengths and limitations, so in choosing one that is appropriate, researchers should think carefully about the circumstances of their evaluation.

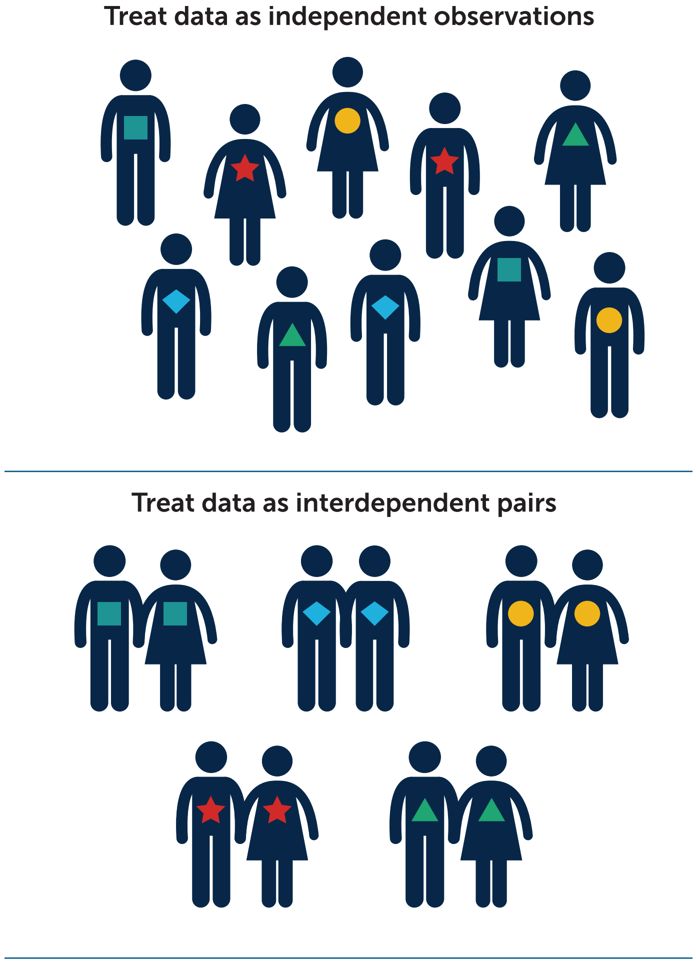
Remember that partners' outcomes could be related

What is the challenge?

Analyzing paired partners is different from analyzing individuals because partners in a pair are not independent of each other—they are connected, or interdependent. Usually, an individual's outcome is more like the outcome of their romantic partner than that of another random person in the sample. For example, one person's report of conflict in the pair's relationship is linked to their partner's rating of conflict—if one partner reports frequent relationship conflict, we would expect their partner to do so also. Less frequently, partner's reports are negatively related—especially when they are measured as a share relative to their partner (for example, the share of household chores done by each partner). Because partners' outcomes are not completely independent, analysis should not treat paired partners the same as two independent sample members.

The statistical approach chosen to analyze paired partners' data must take into account that the partners' outcomes are not independent of each other. Most statistical tests assume "independence of observations"—that is, that individuals or reports in the sample are unrelated to one another. For example, most statistical tests would treat the 10 individuals in the left-hand panel of Figure 1 as 10 independent observations (N = 10), when they are in fact 5 sets of paired partners, as can be seen in the right-hand panel. If statistical tests treat the sample as 10 individuals, the results from tests of statistical significance will be incorrect. Specifically, the statistical test is more likely to calculate *p*-values that are too small (suggesting statistical significance) when partners' reports are positively correlated and *p*-values that are too large when partners' reports are negatively correlated. Thus, any analysis in which data from paired partners are treated as independent is susceptible to bias. An evaluation could overestimate (or underestimate) the impacts of a program by concluding an effect was statistically significant (or not) based on an incorrect *p*-value (Kenny et al. 2006).

Figure 1. Analysis should account for partners' outcomes being interdependent



Strategies to address the challenge

To determine the most appropriate statistical approach, it is important to consider the type of outcome you are analyzing. The first type of outcomes is inherent characteristics of the pair, not of the partners who make up the pair. These are best defined as pair-level variables. For example, whether a romantic couple is married or not is a pair-level characteristic. Even if both paired partners are asked about their marital status, there is no concern about interdependence in their responses, because there is only one "true" value for this pair-level variable.[1] For such outcomes, it makes sense to construct and analyze data at the pair level—for example, by measuring a pair's marital status as an indicator variable for whether both partners report they are married. For this reason, we will not discuss this type of outcome further in this brief.

The second type of outcome is one that reflects each partner's perception or assessment of a shared aspect of their relationship, such as relationship conflict. Such outcomes are not independent. A third type of outcome captures individual-level characteristics, such as risk of depression. Individual-level characteristics might still not be independent for paired partners. For example, one partner's risk of depression could influence their partner's. There are two possible approaches to analyzing the second and third types of outcomes, which we describe below.

One approach is to construct and analyze outcome measures at the pair level. Intuitively, this could be appealing for the type of outcome in which each partner is rating the same aspect of the paired partners' relationship. For example, evaluators can measure relationship conflict by taking the average of each partner's report of relationship conflict; such a measure would mechanically account for each partners' report of relationship conflict not being independent of the other's. When an outcome is measured at the pair level, the correct sample size to be used for traditional statistical tests (for example, chi-square or t-tests) is the number of pairs—not the number of individuals. This is because both paired partners would be assigned as a unit to either receive HMRF services (for example, by being randomly assigned to the program group) or not, so it is imperative for standard errors and *p*-values to reflect the number of units assigned and the potential correlation of outcomes within units.

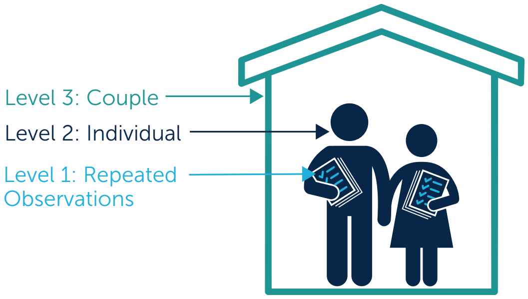
The other approach is for evaluators to analyze data at the individual level and use statistical models that account for the interdependency of partners' reports. This approach may be appealing for outcomes that tap individuals' characteristics more than they do the pairs' characteristics. An example is an individual's risk of depression, which could be modeled as an individual-level variable while still accounting for partners' reports not being independent of each other. Next, we review three common methods that take that interdependence into account.

Repeated-measures analysis of variance (ANOVA) is most commonly used when observations are not independent of one another. If we consider each pair to be a group, the researcher can specify within-group factors that can vary within each pair, such as observations from different points in time or the genders of partners, as well as between-group factors that differ from one pair to the next, such as marital status. In an impact evaluation, paired partners' program group assignment must be specified as a between-group factor in a repeated-measures ANOVA.

Multilevel modeling (MLM) is a flexible approach that allows the investigator to account for several levels of interdependence, sometimes called nesting. For example, as shown in Figure 2, couples' data can be conceptualized as three levels—repeated measures over time (Level 1) that are reported by an

[1] Most aspects of relationship status fit this description. Additional examples include, whether the paired partners are living together, are engaged, or have a child together.

Figure 2. Data nested at three levels



individual (Level 2) who is part of a couple (Level 3). When there are only two time points included in analyses, as is often the case in HMRF evaluations, data can be analyzed using two-level models with individual characteristics such as age and baseline measure of an outcome modeled at Level 1 and couple characteristics, such as marital status and program group assignment, modeled at Level 2. MLM also allows evaluations to examine the independent contributions of each partner, which is useful when primary research questions have to do with partners' influence on each other.[2] For example, MLM can be used to examine whether an individual's mental health impacts not only their own marital satisfaction but their partner's.

Structural equation modeling (SEM) is another approach to model, estimate, and test a network of relationships between variables. It allows investigators to model pair-level variables assessed at the individual level.[3] For example, although each partner reports the frequency of relationship conflict individually, it can be understood as being a characteristic of the relationship. SEM also allows examination of actor-partner interdependence models.

Program evaluation typically focuses on the differences *between* pairs. For example, a random assignment study will examine the difference in outcomes between pairs who are randomly assigned to a program group and those assigned to a control group. In this case, both approaches—either analyzing outcomes at the pair level or using statistical models that account for partners' reports not being independent—will typically yield similar impact estimates that are valid for the sample of pairs being studied (Schochet 2013).[4] Ideally, evaluators should develop analysis protocols that pre-specify both a benchmark analytic approach and sensitivity checks using alternative approaches.

Think carefully about partner-specific effects

What is the challenge?

When the results of data analysis are presented separately for men and women,[5] readers tend to look for gender differences, and they could erroneously infer gender differences in program impacts that do not really exist. For example, a study might reveal that a program significantly reduces men's reports of relationship conflict and has no significant impact on women's reports of relationship conflict. If these findings are presented separately, readers could conclude that the program has greater impacts on relationship conflict for men than it does for women. However, if the difference between male and female partners was not tested directly, it would be wrong to conclude that gender moderates a program's impacts.

[2] These models, referred to as actor-partner interdependence models, are described in Campbell & Kashy (2002).
[3] These models, referred to as common-fate models, are described in Ledermann, T., & Macho, S. (2014).
[4] One technical consideration is the extent to which the evaluation sample is considered to represent a broader population. When evaluation findings are intended to reflect only the sample of pairs in the study, analyzing outcomes at the pair level is appropriate (Schochet 2013). For example, if an HMRE program serves a convenience sample of pairs who choose to sign up for the program (rather than drawing a random sample from a broader population of pairs), then evaluation findings represent only the pairs in the study. However, when an evaluation sample is considered to represent a broader population, the underlying assumptions of MLM are more appropriate than analyzing outcomes at the pair level (Moerbeek et al. 2003, Schochet 2013).
[5] Analysis of data from same-gender couples (or other types of pairs who may not be able to be consistently distinguished by gender) require additional statistical considerations. These analyses can be conducted with MLM and SEM using similar models to those described above. For more information, see Kashy, Donnellan, Burt, & McGue (2008).

Strategies to address the challenge

There might be important gender differences in the impacts of HMRF programs, but they should be hypothesized before they are tested, and only examined with forethought and justification. Pre-specifying an analysis that explores differences in impacts by gender will reduce the chance of reporting or interpreting spurious gender differences in impacts. If a test for gender differences is warranted, then the statistical models described above can be used to test whether impacts differed by gender.

- Repeated-measures ANOVA of a pair's data will require gender to be specified as a within-group factor, so gender will be included as a moderator of program impacts automatically—a test that is not warranted unless a gender difference is hypothesized.
- Multilevel modeling and SEM approaches can be used to analyze pairs' data without unnecessarily testing for gender differences. For example, a researcher can apply SEM for both distinguishable and non-distinguishable pairs (that is, not taking gender into account). In a multilevel model, individuals can be nested within pairs, and gender can optionally be added as a moderator of program impacts at the individual level.

Remember that some outcome measures exclude some people

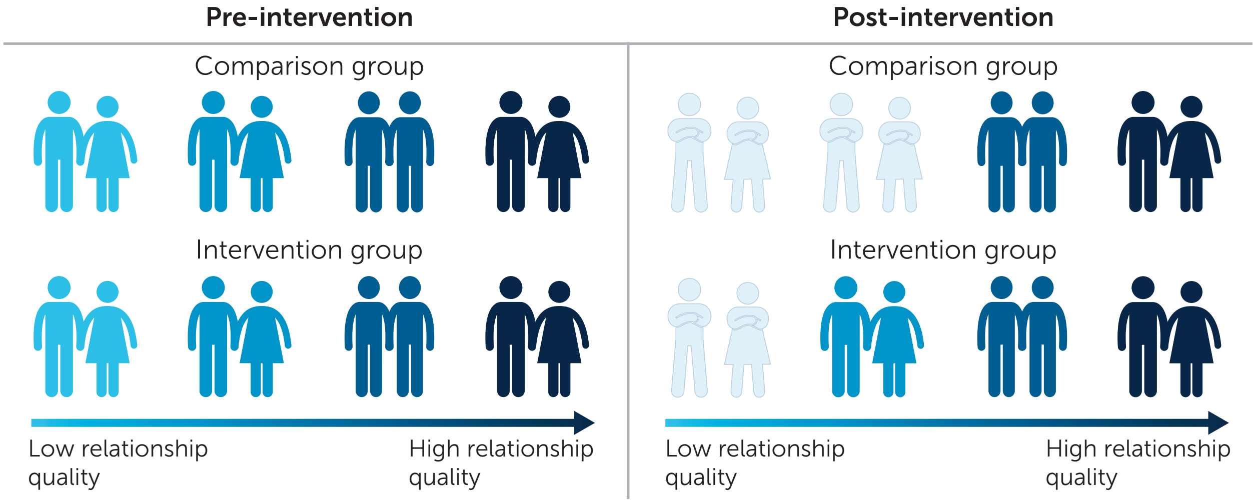
What is the challenge?

Certain outcomes cannot be defined for all sample members. For example, relationship quality can only be defined for people who are in a relationship. This phenomenon, known as truncation, is particularly problematic if our ability to observe the outcome depends on something that can be directly affected by the program. For example, whether we observe relationship quality depends on the pair's relationship status—if a program affects the number or type of pairs who stay together, it also affects the pairs for whom we can observe relationship quality.

Truncation can bias estimates of a program's impact on affected outcomes. Consider the example of an HM program designed to affect a range of aspects of couples' relationships, such as commitment and quality. In this example, the actual effect of the program is to encourage couples to be more committed to their relationships but the program does not change the underlying relationship quality. Thus, the program causes some couples whose relationship quality is poor to stay together when they would not have done so in the absence of the program. In this case, relationship quality will not change from baseline to follow-up for any couple. However, when we examine only couples for whom relationship quality is observable at follow-up—that is, those who remain together at follow-up—we will find that the average quality of relationships in the treatment is lower than the average quality of relationships in the control group. This is because couples with lower quality relationships are more likely to stay together if they are in the treatment group and not in the control group (see Figure 3). The finding of lower quality could, however, lead us to erroneously conclude that the program had a negative impact on relationship quality, when in fact, the program only changed the *types* of couples for whom relationship quality could be observed at follow-up.

Truncation is more common in certain circumstances. In the example of relationship quality, truncation is more common for populations where fewer couples tend to have stable relationships. It is also more common with longer follow-up periods, because there is more time for couples to break up. The risk of bias from truncation is greatest when the program impacts the characteristic that determines whether an outcome can be observed.

Figure 3. Some outcomes cannot be observed if a relationship ends



Strategies to address the challenge

When selecting outcomes for a study of program impacts, evaluators must carefully consider the sample for which outcomes are defined—especially if the target population, length of follow-up, or other factors suggest that there is a risk of truncation. Ideally, truncated outcomes should not be designated as key tests of program effectiveness.

If truncated measures cannot be avoided, evaluators can assess the risk of bias by treating truncation as a type of sample attrition, because—as with other types of attrition—truncated outcomes are not available for some sample members. Evaluators can assess the risk of bias from any type of attrition by using established guidelines from systematic evidence reviews like the Administration for Children and Families' [Pathways to Work Evidence Clearinghouse] or the Department of Education's [What Works Clearinghouse]. Studies should not report findings for which the assessment indicates a high risk of bias due to truncation and should provide clear caveats for findings with moderate risk of bias.

Collect and use as much data as possible

What is the challenge?

Follow-up data on one partner in a pair can sometimes be missing. In this case, dropping the pair from the impact analysis is not recommended. The sample of pairs with data available on both partners might not be representative of all the pairs who enrolled in the study (Barton et al. 2020).[6] For example, it might be more difficult to locate and interview both partners in pairs with less stable relationships than it is to locate both partners when they are married or living together. In this case, follow-up survey data for pairs in which both partners completed the survey are unlikely to represent the outcomes of all pairs in the study, which might lead to biased estimates of program impacts. Making use of data for pairs in which only one partner responded increases what we can learn about pairs in the study, particularly if available data can inform an educated guess about the missing data.

[6] Barton et al. (2020) studied a relationship education program and found that, compared with individuals whose partners completed the enrollment form, individuals whose partners did not participate showed higher levels of break-up potential, physical aggression, negative communication, psychological distress, and anger, on average.

Strategies to address the challenge

First, every possible step should be taken to maximize survey response rates so data are available for both partners in most pairs in the study.[7] Preventing missing data is crucial because addressing missing data almost always requires assumptions which cannot be fully verified.

Next, evaluators should select an analytic approach that can handle pairs in which only one member is a respondent. Multilevel modeling approaches (described above) are one option because they use estimation methods that account for missing data. Alternatively, statistical methods such as imputation can be used to fill in missing data for an individual.[8] For example, if data are missing on an individual's relationship commitment at follow-up, we can impute a value using combined information from the baseline—such as their initial relationship commitment—and from the follow-up—such as the responding partner's reported relationship commitment. Notably, using only baseline data to impute missing follow-up data is not recommended. For example, it is not advisable to impute follow-up outcomes when neither partner responded to the survey.

Conclusion

This brief describes the benefit of collecting data from both partners for evaluations of programs serving paired partners, and strategies researchers can use to address common challenges to analyzing these data. In designing such evaluations, researchers should carefully weigh the strengths and limitations of these strategies in the context of their evaluation. A thoughtful plan for collecting and analyzing data from both partners of participating pairs can improve our understanding of whether and how programs serving pairs are changing the lives of participants.

[7] Resources describing strategies for achieving survey response rate targets are available from the Fatherhood Research and Practice Network, University of Kansas Community Toolbox, Corporation for National and Community Service, and Kellogg Foundation.
[8] Examples of this approach are described in the ACF-sponsored Building Strong Families and Parents and Children Together studies.

References

Barton, A.W., S.G. Hatch, and B.D. Doss. “If You Host It Online, Who Will (and Will Not) Come? Individual and Partner Enrollment in a Web-Based Intervention for Distressed Couples.” *Prevention Science: the Official Journal of the Society for Prevention Research*, April 2020.

Campbell, L., & D. A. Kashy. Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide. *Personal Relationships*, 9(3), 327-342.

Kashy, D. A., M. B. Donnellan, S. A. Burt, and M. McGue. “Growth curve models for indistinguishable dyads using multilevel modeling and structural equation modeling: The case of adolescent twins’ conflict with their mothers. *Developmental Psychology*, March 2008.

Kenny, D. A., D. Kashy, and W. Cook. *Dyadic Data Analysis*. New York: Guilford Press, 2006.

Ledermann, T., and S. Macho. Analyzing change at the dyadic level: The common fate growth model. *Journal of Family Psychology*, 28, 204-213.

Schochet, P. Z. “Estimators for Clustered Education RCTs Using the Neyman Model for Causal Inference.” *Journal of Educational and Behavioral Statistics*, vol. 38, no. 3, 2013, pp. 219–238.

This brief was written by Ankita Patnaik, Galena Rhoades, Quinn Moore, and Brian Doss, under contract with OPRE, ACF, DHHS (#HHSP2332009564WC/HHSP23337050T). OPRE Project Officers: Katie Pahigiannis and Pooja Gupta Curtin. Mathematica Project Director: Sarah Avellar.

 Suggested citation: Patnaik, A., G. Rhoades, Q. Moore, and B. Doss. (2021). Analyzing data on paired partners in program evaluation: Strategies to overcome common challenges. Evaluation technical assistance brief. (OPRE Report 2021-23). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services (https://www.acf.hhs.gov/opre).