Using a Dyadic Logistic Multilevel Model to Analyze Couple Data

Mariana A. Preciado[a], **Jennifer L. Krull**[b], **Andrew Hicks**[c], and **Jessica D. Gipson**[d]

Mariana A. Preciado: mariana.a.preciado@gmail.com; Jennifer L. Krull: krull@psych.ucla.edu; Andrew Hicks: hicks@hcp.med.harvard.edu; Jessica D. Gipson: jgipson@ucla.edu

[a]Director of Research and Evaluation, CollegeSpring. 800 S. Figueroa Street, Suite 760 Los Angeles, CA 90017 Phone: 213.550.2202

[b]Associate Professor, UCLA Department of Psychology, 4643 Franz Hall Los Angeles, CA 90095-1563. Phone: 310.206.2951

[c]Statistician, Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115 Phone: 617.432.3459

[d]Assistant Professor, UCLA Fielding School of Public Health, 650 Charles E. Young Drive South, CHS 46-071B, Los Angeles, CA 90095-1772. Phone: 310.794.7028

Abstract

There is growing recognition within the sexual and reproductive health field of the importance of incorporating both partners' perspectives when examining sexual and reproductive health behaviors. Yet, the analytical approaches to address couple data have not been readily integrated and utilized within the demographic and public health literature. This paper seeks to provide readers unfamiliar with analytical approaches to couple data an applied example of the use of dyadic logistic multilevel modeling, a useful approach to analyzing couple data to assess the individual, partner, and couple characteristics that are related to individuals' reproductively relevant beliefs, attitudes, and behaviors. The use of multilevel models in reproductive health research can help researchers develop a more comprehensive picture of the way in which individuals' reproductive health outcomes are situated in a larger relationship and cultural context.

Keywords

couples; reproductive health; multilevel modeling; methods

INTRODUCTION

There is increasing recognition of the importance of incorporating both partners' perspectives when examining sexual and reproductive health behaviors. The integration of

both partners' perspectives allows us to better predict sexual and reproductive health behaviors, facilitates the development of more appropriate and effective couple-level interventions, and provides critical insights into the powerful influences beyond the individual that also affect reproductive outcomes, such as those attributable to gender inequity and power differentials in relationships [1, 2]. Despite the notable benefits of couple or dyadic data, however, analytic approaches have not been well-integrated and utilized within the demographic and reproductive health literature.

Dyadic data is typically derived from surveys that elicit information from both members of a partnership pair (e.g., husband and wife) and, as such, offer a more holistic view of sexual and reproductive behaviors. Partners' beliefs, attitudes, and behaviors can have separate, and even interactive, effects on relationship functioning [3] and on sexual and reproductive outcomes [4–6]. The benefits of dyadic data, however, are coupled with additional considerations and potential complications. For example, partners' reports may be discrepant, requiring a decision regarding how these reports should be handled. In some cases, discrepant couple reports may be discarded, yet this strategy reduces statistical power for detecting effects and also prevents further exploration of *why* partner reports are discrepant. For example, these differences may reflect disparate perceptions of the relationship and/or differences in information available to each partner, as well as how that information may be processed and internalized given prevailing gender norms and roles [7].

Additionally, partners' reports are often dependent on each other, as partners are likely to influence one another and share a similar context. In addition to the conceptual issues, this dependency is statistically problematic because it creates a situation in which partners' error terms may be correlated, violating the assumption of independent errors in general linear models. Thus, researchers must utilize a strategy that accounts for this dependency, while simultaneously incorporating the unique reports of partners within the relationship.

Dyadic multilevel modeling provides a useful approach to analyzing couple data to assess the individual, partner, and couple characteristics that are related to beliefs, attitudes, and behaviors. This article is intended to give reproductive health researchers unfamiliar with multilevel modeling techniques an initial introduction from which to build their understanding and use of the approach. An applied example of the use of dyadic logistic multilevel modeling is provided, demonstrating the potential of dyadic logistic modeling and its utility in developing a more comprehensive picture of the way in which individuals' reproductive health outcomes are situated in a larger relationship and cultural context.

DYADIC LOGISTIC MULTILEVEL MODELING

Multilevel modeling is a statistical technique for the analysis of nested or clustered data, that is, individual observations within the same organizing unit. Since observations from within the same cluster maybe more similar to each other than randomly paired individual observations, the multilevel model includes an error structure that accounts for the dependence of the errors of observations from within the same cluster.

In the dyadic logistic multilevel model, there are two levels of analysis: the individual level, including both partners' individual observations (Level 1; identified by subscript i), and the

couple level (Level 2; identified by subscript *j*). At Level 1, individual respondents' outcomes and predictors are included in a single-level logistic regression equation specific to each couple. The logistic equation is used in cases in which the outcome is binary (0 = no outcome, 1 = outcome), transforming the binary outcome to the log odds of the outcome; this allows the formerly binary outcome to be analyzed using a linear regression model. Similar to logistic regression, predictors can be categorical or continuous and include main effects and interactions.

Level 1

$$\log\left[\frac{P}{1-P}\right]_{ij} = b_{0j} + b_{1j}X1_{ij} + b_{2j}X2_{ij} \quad (1)$$

As an example, let's say the outcome in equation 1 represents the log odds of an individual respondent (i) in a particular couple (j) reporting that they make the decision to use family planning jointly with their partner (1: joint decision, 0: not joint decision). Predictor $X1_{ij}$ is the respondent's age. Predictor $X2_{ij}$ is the respondent's *partner's* age.[5]

The b_{1j} coefficient is the expected increase in the log odds of reporting a joint-decision for every 1 unit increase in respondent's age, over and above the effect of partner age. The b_{2j} coefficient is the expected increase in the log odds of reporting a joint-decision for every 1 unit increase in partner age, over and above the effect of respondent age. At Level 1, the multilevel model is analogous to a logistic regression model performed for each couple.

As it is impossible to run this logistic regression model within each couple with only two observations, the multilevel model instead characterizes the distribution of such coefficients across couples. Two different types of parameters are thus estimated at Level 2 of the model: fixed effects (gammas or γ's) estimate the "average" Level 1 coefficients across couples, and variance components estimate the degree to which coefficients differ across couples. Specifically, each of the b coefficients at Level 1 are modeled as outcome variables at the couple level (Level 2) by an overall effect and, for the intercept only, a couple-specific error term. This couple-specific error term is known as a random effect because it allows the introduction of a random term (i.e., error term) at the couple level.

Level 2

$$b_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\text{var}(u_{0j}) = \tau_{00} u_{0j} \sim N(0, \tau)$$

$$b_{1j} = \gamma_{10} \quad (3)$$

---

[5]Respondent and partner ages are entered in their natural metric. Consequently, b_{0j} is the expected log odds of reporting a joint-decision for a respondent in couple j at 0 ("birth"). It may be more useful to center respondent and partner ages at the average age of the sample. For more information on centering, see (Kenny, Kashy and Cook 2006).

$$b_{2j}=\gamma_{20} \quad (4)$$

The equations at Level 2 demonstrate how the b coefficients at Level 1 are modeled. In equation 2, the expected log odds of reporting a joint decision for a given observation in which both partner and respondent age are 0 in couple j (b_{0j}) is modeled as a function of the overall log odds of reporting a joint decision at 0 for both respondent and partner age (γ_{00}) and an adjustment for each couple's deviation from the overall effect (u_{0j}). The variance of the one Level 2 random effect (u_{0j}) is referred to as τ_{00}. This represents the amount of variance by which individual couples' intercepts deviate from the overall intercept (γ_{00})[6]. Because individuals in the same couple share the same error term, the inclusion of the error also appropriately models within-couple similarity.

In equations 3 and 4, the effects of respondent age in couple j and partner age in couple j, respectively, are modeled as an overall fixed effect across couples with no corresponding random effect. In multilevel models with larger group sizes, more than one effect can be allowed to vary randomly at Level 2. However, the group size of the dyadic logistic multilevel model only allows for the estimation of one random effect at Level 2.[7] Thus, this model is set up with the assumption that the effects of respondent and partner age are the same across couples.[8]

At Level 2, the first γ subscript indicates which Level 2 equation contains the coefficient. For example, in equation 2, the first subscript of the γ coefficient (0) indicates that that coefficient belongs to the 0 equation – in other words, the equation for the intercept (b_{0j}). The second γ coefficient indicates the position in the equation. In equation 2, the second γ subscript (0) indicates that the coefficient is in the 0 position – in other words, the intercept position within the L2 equation.

Finally, couple predictors (i.e., those predictors that vary across couples but not across individuals) can be added to the Level 2 equations. For instance, a predictor W_j, representing the absolute difference between the partners' ages, can be entered to equation 2 as a predictor of the couple intercept.

Level 2

$$b_{0j}=\gamma_{00}+\gamma_{01}W_{j}+u_{0j} \quad (5)$$

In equation 5, γ_{01} is the main effect of age difference, or, more specifically, the expected increase in the log odds of reporting a joint decision for every 1 unit increase in the absolute difference between the partners' ages, controlling for respondent and partner age. The

---

[6]This allows the model to capture the degree to which relationship partners' responses are similar to each other. However, it does not allow for situations in which partners' responses are *dissimilar*. For an alternate parameterization of the Level 2 variance term, see Kenny, Kashy, & Cook, 2006, chapter 4.
[7]The use of additional random effects with such a small group size could prevent model estimation and create biased estimates of the error terms (Hox and Maas 2006, Newsom 2002, Raudenbush and Bryk 2002).
[8]Violation of this assumption (e.g., if the effects of respondent and partner age do vary across couples) will not bias significance tests of coefficients in this model (Kenny, Kashy and Cook 2006).

variance of the random effect (τ_{00}) is now the residual variance in intercepts after accounting for the absolute difference between partner ages.

While it is useful to conceptualize the multilevel model as a series of equations at two different levels, in actuality, the model is estimated as a single combined equation. For the example given above, algebraic substitution of the Level 2 expressions for the coefficients into the Level 1 model gives:

Single-Equation Form

$$\log\left[\frac{P}{1-P}\right]_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X1_{ij} + \gamma_{20}X2_{ij} + u_{0j} \quad 6)$$

The single-equation form of the model makes clear that, in our example, the multilevel model actually estimates the coefficients across couples (the γs) and then includes one error term to model the deviation of individual couples' intercept coefficients from the overall intercept.

Single Level and Cross-Level Interactions

As stated previously, any type of individual predictor can be entered into the Level 1 equation (see equation 1). This includes interactions among multiple predictors. For instance, if we add a third predictor into the Level 1 equation – let's say the respondent's income ($X3_{ij}$), we could examine the interaction of respondent's age and respondent's income by entering that interaction of the Level 1 equation as follows.

Level 1

$$\log\left[\frac{P}{1-P}\right]_{ij} = b_{0j} + bX1_{ij} + b_{2j}X2_{ij} + b_{3j}X3_{ij} + b_{4j}X1_{ij}X3_{ij} \quad (7)$$

The b_{4j} coefficient would have been modeled at Level 2 as an overall interaction coefficient.

Level 2

$$b_{4j} = \gamma_{30} \quad (8)$$

The single-equation form for the full model would have then included an interaction term, specifying a Level 1 interaction.

Single-Equation Form

$$\log\left[\frac{P}{1-P}\right]_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X1_{ij} + \gamma_{20}X2_{ij} + \gamma_{30}X3_{ij} + \gamma_{40}X1_{ij}X2_{ij} + u_{0j} \quad (9)$$

Likewise, had we wanted to examine the interaction of two couple-level predictors, absolute age difference (W1) and relationship duration (W2), we could have entered the interaction on the Level 2 equation for b_{0j} (equation 10) and algebraically combined the equations as indicated in equation 11.[9]

Level 2

$$b_{0j} = \gamma_{00} + \gamma_{01}W1_j + \gamma_{02}W2_j + \gamma_{03}W1_jW2_j + u_{0j} \quad \text{10)}$$

Single-Equation Form

$$\log\left[\frac{P}{1-P}\right]_{ij} = \gamma_{00} + \gamma_{01}W1_j + \gamma_{02}W2_j + \gamma_{03}W1_jW2_j + \gamma_{10}X1_{ij} + \gamma_{20}X2_{ij} + \gamma_{30}X1_{ij}X2_{ij} + u_{0j} \quad \text{(11)}$$

However, cross-level interactions, or interactions between two or more variables represented at different levels of the model, are represented differently. For example, had we wanted to examine the interaction of absolute age difference (W) and respondent age (X1), we would have needed to enter absolute age difference on not only the equation for b_{0j} (the intercept for group j), but also the equation for b_{1j} (the coefficient for respondent age in group j).

Level 2

$$b_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad \text{(12)}$$

$$b_{1j} = \gamma_{10} + \gamma_{11}W_j \quad \text{(13)}$$

Thus, when the Level 1 and Level 2 equations are combined algebraically, both the lower-order effect of absolute age difference (from the b_{0j} equation) and its interaction with respondent age (from the b_{1j} equation) are present in the single-equation form of the model.

Single-Equation Form

$$\log\left[\frac{P}{1-P}\right]_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X1_{ij} + \gamma_{11}X1_{ij}W_j + \gamma_{20}X2_{ij} + u_{0j} \quad \text{(14)}$$

In equation 14, the coefficient γ_{11} represents the effect of the cross-level interaction of respondent age and absolute age difference, over and above the effect of partner age. In other words, this represents the expected change in the association between respondent age and the log odds of reporting a joint decision for each one unit increase in absolute age difference between partners, controlling for partner age.

RESEARCH EXAMPLE

Sample

We used data from the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS), part of an ongoing cohort study of Filipino women who have been followed, along with their children, since 1983–84 [13, 14]. The analytic sample includes data from the CLHNS

---

[9]It should be noted that a cross-product term formed by multiplying two Level 1 predictors representing respondent and partner measures of the same construct (e.g., X1ij and X2ij representing respondent and partner age in equation (8) multiplied to produce the cross-product X1ijX2ij will be identical for both members of the couple and thus will form a Level 2 interaction.

children (aged 20 – 22 years in 2005) and their opposite-sex romantic partners (398 couples; n = 796 individuals).

Dependent Variable

The dependent variable in this example represents the log odds of an individual respondent (i) in a particular couple (j) reporting that they confide in their partner when they have a problem (0 =does not confide with partner, 1 = confides with partner). Across the sample, 65% of respondents indicated that they confided in their partner.

Partners reports of confiding in each other were positively correlated ($r = .07$, $p =.03$). This correlation highlights the importance of using an analytic technique that can account for the relationship between respondent and partner responses. However, the small size of the correlation indicates a notable lack of agreement between partners, further highlighting the importance of including both partner responses in the analysis. Thus, a multilevel approach is ideal.

Independent Variables

For the purpose of this example, characteristics used to predict whether a respondent confides in their partner were respondent's age, partner's age, and the absolute difference in age between the respondent and partner. The average age across the sample was 22 years (21 for females; 23 for males) and the average age difference was three years. All three age variables were centered on the mean so that a "0" value for each variable corresponded to the mean (for more information on centering, see Kenny, Kashy, & Cook, 2006). To control for gendered differences in responses and possible differences in predictor associations by gender, sex was included in the model and interacted with each of the other predictors. Sex is coded as a dummy variable, with 1 representing males and 0 representing females.

Data Organization and Analytic Strategy

All analyses were conducted in Stata/MP 13.1 using the *melogit* command — Multilevel mixed-effects logistic regression. We initially tested interactions of all predictors with respondent sex; however, respondent's sex by respondent's age was the only significant interaction. Therefore, we ran a final combined logistic model regressing confiding in one's partner on respondent's age ($X1$), partner's age ($X2$), respondent sex ($X3$), an interaction between respondent's age and sex ($X1X3$) and the difference between respondent's and partner's ages (W) as specified in the following single-equation form of the model:

<u>Single-Equation Model</u>

$$y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X1_{ij} + \gamma_{20} X2_{ij} + \gamma_{30} X3_{ij} + \gamma_{40} (X1_{ij} X3_{ij}) + u_{0j} \quad (15)$$

RESULTS

The likelihood ratio test comparing the multilevel model against a single level ordinary linear regression model containing the same predictors was marginally significant, chi-square = 1.67, $p = 0.09$, indicating that variance of u_{0j} (i.e., τ_{00}) was marginally greater than

0. This test further affirmed the importance of using a multilevel model to analyze this data to avoid bias in significance testing [15, 16].[10]

The effect of partner's age was significant, over and above the effects of the other predictors included in the model (Table 1). Specifically, a one-year increase in the partner's age was related to a 0.10 increase in the log-odds of confiding in one's partner (γ_{10}=0.10, p=.04). By exponentiating the coefficients (i.e., e^ γ_{20}), the coefficient can be converted to an odds ratio, offering a multiplicative interpretation. Specifically, the results indicate that the expected odds of confiding in one's partner are 1.11 times greater with each one year increase in partner's age (OR_{10}=1.11, p=.04).

The effect of the difference between partners' ages was also significant, over and above the effects of the other predictors. Specifically, each additional year difference in the age between the two partners corresponded to a 0.12 decrease in the log odds of confiding in one's partner (γ_{01}=−0.12, p=.03), or, alternatively, the expected odds of confiding in one's partner decreases by a factor of 0.88 (OR_{01}=−0.89, p=.03). In other words, the greater the difference between the two partners in age, the less likely the respondent is to confide in their partner.

Sex was not statistically significant in predicting whether or not the respondent confided in their partner. However, the interaction of respondent sex and age was significant in the multivariate model (γ_{40} =0.21, p=.014; OR_{40}=1.23, p=.014). We, therefore, examined the effect of partner age on confiding in the partner separately by respondent sex (i.e., simple effects).

For male respondents, their age was related to a 0.24 increase in the log odds of confiding in their partner (γ_{simple} =0.24, p<.001). In terms of odds ratios, a one year increase in age corresponded to a 1.27 increase in the expected odds of confiding in their partner (OR_{simple} =1.27, p<.001). In other words, older male respondents were more likely to respond that they confided in their partner. On the other hand, for female respondents, age was not significantly related to their reports of confiding in their partner (γ_{simple} =0.04, p=.62; OR_{simple} =1.04, p=.58). This difference between the importance of age for male and female respondents suggests age has a much larger effect and is only statistically significant when the respondent is male.

DISCUSSION

Multilevel modeling offers a flexible framework, with which researchers can analyze couples data to assess the individual, partner, and couple characteristics that are related to individuals' reproductively relevant beliefs, attitudes, and behaviors. Using the methods outlined in the first part of this paper, we were able to show that an individual's age, sex,

---

[10]While a significant likelihood ratio test clearly indicates a need for a multilevel model (as opposed to a single-level regression model), a non-significant result should be interpreted with caution, as power of this test may be an issue when sample sizes are small to moderate, and simulation studies of clustered data indicate that even a very small degree of within-group similarity can lead to downwardly biased standard errors and overly liberal inferences (e.g., Barcikowski 1981). In the case of dyadic data, relationship partners' data is very likely to be dependent to at least some degree, and even if that dependence is not detected by the likelihood ratio test, it can nonetheless bias results. In the case of dyadic data, therefore, it is almost always advisable to use a multilevel model to account for error dependence simply because of the strong likelihood of dependence inherent in this data structure.

partner's age, and age difference between respondents and partners each have an important and simultaneously unique relationship with the individual's report of confiding in their partner. Had we not included both partners' data in these analyses, we could not have detected the different relationships between one's own characteristics and the outcome and one's *partner's* characteristics and the outcome.

While there are alternatives to multilevel modeling for the analysis of couple data, including structural equation modeling (e.g., [11, 17]) and bivariate probit/logit regression models (e.g., [18]), the multilevel modeling framework offers a flexible, intuitive approach that can help researchers develop a more comprehensive picture of the way in which individuals' reproductive health outcomes are situated in a larger relationship and cultural context.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

References

1. Becker S. Couples and reproductive health: A review of couple studies. Studies in Family Planning. 1996; 27(6):291–306. [PubMed: 8986028]

2. Blanc AK. The effect of power in sexual relationships on sexual and reproductive health: An examination of the evidence. Studies in Family Planning. 2001; 32(3):189–213. [PubMed: 11677692]

3. Kashy, DA.; Kenny, DA. The analysis of data from dyads and groups, in Handbook of research methods in social psychology. Reis, HT.; Judd, CM., editors. New York: Cambridge University Press; 2000. p. 451-477.

4. Sipsma H. Pregnancy desire among a sample of young couples who are expecting a baby. Perspectives on Sexual and Reproductive Health. 2012; 44(4):244–251. [PubMed: 23231332]

5. Gipson JD, Hindin MJ. The effect of husbands' and wives' fertility preferences on the likelihood of a subsequent pregnancy, Bangladesh 1998–2003. Population Studies-a Journal of Demography. 2009; 63(2):135–146.

6. Mason KO, Smith HL. Husbands' versus wives' fertility goals and use of contraception: The influence of gender context in five Asian countries. Demography. 2000; 37(3):299–311. [PubMed: 10953805]

7. Neff LA, Karney BR. To know you is to love you: The implications of global adoration and specific accuracy for marital relationships. Journal of Personality and Social Psychology. 2005; 88(3):480–497. [PubMed: 15740441]

8. Acock Alan, C. A Gentle Introduction to Stata. 2nd ed.. College Station, TX: Stata Press; 2008.

9. Kenny, DA.; Kashy, DA.; Cook, WL. Dyadic data analysis. New York: The Guilford Press; 2006.

10. Hox, J.; Maas, C. Multilevel models for multimethod measures, in Multimethod Measurement in Psychology. Eid, M.; Diener, E., editors. Washington, DC.: American Psychological Association; 2006. p. 269-282.

11. Newsom JT. A multilevel structural equation model for dyadic data. Structural Equation Modeling. 2002; 9(3):431–447.

12. Raudenbush, SW.; Bryk, AS. Advanced Quantitative Techniques in the Social Sciences. 2nd ed.. Thousand Oaks: Sage Publications; 2002. Hierarchical Linear Models: Applications and Data Analysis Methods.

13. Feranil AB, Gultiano SA, Adair LS. The Cebu Longitudinal Health and Nutrition Survey: Two Decades Later. Asia-Pacific Population Journal. 2008; 23(3):39–54.

14. Adair LS, Popkin BM. The Cebu Longitudinal Health and Nutrition Survey: History and Major Contributions of the Project. Philippine Quarterly of Culture and Society. 2001; 29:5–37.

15. Long, JS. Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage Publications; 1997.

16. Barcikowski RS. Statistical Power with Group Mean as the Unit of Analysis. Journal of Educational Statistics. 1981; 6(3):267–285.

17. Wendorf CA. Comparisons of structural equation modeling and hierarchical linear modeling approaches to couples' data. Structural Equation Modeling. 2002; 9(1):126–140.

18. Allendorf K. Couples' reports of women's autonomy and health-care use in Nepal. Studies in Family Planning. 2007; 38(1):35–46. [PubMed: 17385381]

Table 1

Dyadic logistic multilevel model of whether or not the respondent confides in their partner.

	Logistic Multilevel Model
Intercept	1.03*** (0.20)
Respondent's age	0.04 (0.07)
Partner's age	0.10* (0.05)
Age difference	−0.12* (0.05)
Sex	−0.08 (0.19)
Respondent's age * sex	0.21* (0.09)
τ_{00}	0.34

***
p 0.001;

*
p 0.05;