# Missing Data in Dyadic Modeling: Issues and Opportunities

Alexander M. Schoemann [1] John K. Sakaluk [2]

[1] East Carolina University

[2] University of Victoria

M3 2016

# Overview

- Dyadic data analyses
- Missing data handling
- Missing data challenges in dyadic data
- Planned missing data designs with dyadic data

# Dyadic Data

- Data collected from two individuals
    - Usually individuals have a social relationship
    - Usually measure the same variables on each member of the dyad
    - Responses within dyad are not independent

# Dyadic Data

- Distinguishability – whether dyad members can be "told apart"
  - AKA: Exchangeability
  - Distinguishable dyads (non-exchangeable): dyad where each member has a unique role
    - Heterosexual couples, parent and child, older and younger siblings
  - Indistinguishable dyads (exchangeable): dyad where both members have the same role
    - Homosexual couples, twins, friends, coworkers

# Data Structures for dyadic data

- Three structures for dyadic data from a standard design
  - Individual (long)
  - Dyad (wide)
  - Pairwise
- Choice of the data structure depends on the analysis technique and the type of dyad

# Individual data structure

- Each row represents an individual's score
  - There is a variable representing dyad membership
  - Between dyad variables are entered twice (once on each row)

# Individual data structure

```
##   d p x y z
## 1 1 1 5 9 3
## 2 1 2 2 8 3
## 3 2 1 6 3 7
## 4 2 2 4 6 7
## 5 3 1 3 6 5
## 6 3 2 9 7 5
```

# Dyad data structure

- Each row represents a dyad
  - Responses from different members of the dyad are in different variables

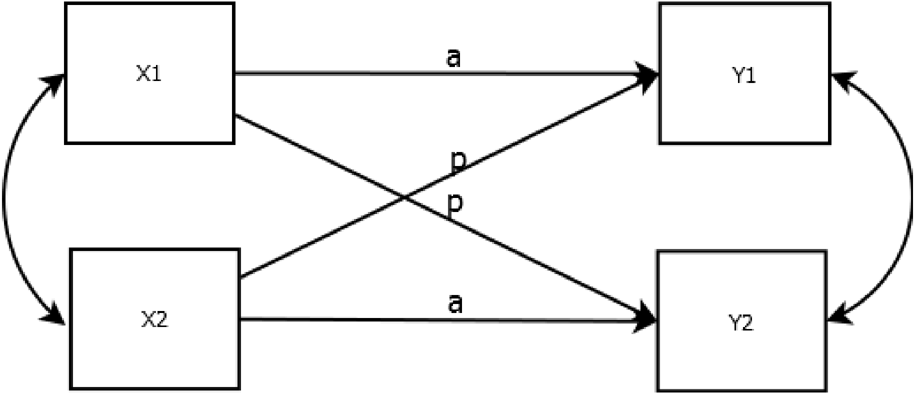
# Dyad data structure

```
##   d x.1 y.1 z.1 x.2 y.2 z.2
## 1 1   5   9   3   2   8   3
## 3 2   6   3   7   4   6   7
## 5 3   3   6   5   9   7   5
```

# Dyadic Data: Models

- Common models with dyads include:
  - Actor-Partner Interdependence Model (APIM)
  - Common Fate Model (CFM)
  - Social Relations Model (SRM)

# APIM



# APIM

- APIMs can be fit with SEM, MLM, or regression
- APIMs are fit differently with indistinguishable and distinguishable dyads
  - In the APIM for indistinguishable dyads the values of all parameters are constrained to be equal across individuals.
  - In the APIM for distinguishable dyads the values of all parameters are freely estimated across individuals.

# Missing data mechanisms

- Missing data comes in three "flavors"
  - Missing Completely at Random (MCAR)
    - No association between missingness and observed or unobserved variables
  - Missing at Random (MAR)
    - Association between missingness and observed variables, no association between missingness and unobserved variables
  - Missing Not at Random (MNAR or NMAR)
    - Association between missingness and unobserved variables

# Old methods of handling missing data

- Listwise deletion
- Pairwise deletion
- Mean imputation
- Regression imputation
- Others
    - Hot deck imputation, last observation carried forward, averaging available items

# Modern methods of handling missing data

- Full Information Maximum Likelihood (FIML)
- Multiple Imputation (MI)

# Missing data in dyadic data

- Rarely discussed by dyadic researchers
  - Missing data strategies only mentioned in about 30% of dyadic papers
  - Deletion strategies tend to dominate (followed by FIML)
- Dependence in dyadic data provides special challenges with missing data.
  - Techniques need to incorporate distinguishability when recovering missing data

# Patterns of missingness in dyadic data

- Two patterns of missingness
  1. Missing data by item
  2. Missing data by person
     - More on this later!

# Missing data challenges: Distinguishable dyads

- Any missing data technique must allow parameters to differ across individuals
  - For FIML: specify the model with parameters allowed to differ across individuals
  - For MI: technique depends on the data structure

# Missing data challenges: Distinguishable dyads MI

- With dyad (wide) data structure: imputation with an unconstrained model
- With individual (long) data structure:
  - MI must take into account nested data structure
  - Data needs to include variables representing the interaction between individual and each variable in the model

# Missing data challenges: Indistinguishable dyads

- Any missing data technique must constrain parameters to equality across individuals
  - For FIML: specify the model with parameters constrained to equality across individuals
  - For MI: technique depends on the data structure

# Missing data challenges: Indistinguishable dyads MI

- With dyad (wide) data structure:
  - Imputation model must include equality constraints on means, variances, and covariances across individuals in the dyad
  - Limited software support
- With individual (long) data structure:
  - MI must take into account nested data structure

# Missing data challenges: Indistinguishable dyads

- What if we pretend data are distinguishable?
  - Impute with an unconstrained model in a dyad format
- Monte Carlo simulation study
  - Population values from Kenny, Kashy, & cook (2006)
  - Actor effect = -.591
  - Partner effect = .888
  - n dyads = 20
  - 20% MCAR missing
  - 20 imputations

# Missing data challenges: Simulation results

- Parameter estimates and bias

| Parameter | listwise | FIML | MI |
|---|---|---|---|
| a | -.596 (0.8) | -.597 (0.9) | -.448 (31.8) |
| p | .894 (0.6) | .885 (-.3) | .700 (-26.9) |

- Power

| Parameter | listwise | FIML | MI |
|---|---|---|---|
| a | .398 | .506 | .286 |
| p | .636 | .810 | .620 |

# Missing data opportunity: Planned missing data

- Missing data does not have to be a problem!
- Two types of planned missing data designs:
  - Time based planned missing data designs
    - Control participant entry into the study (e.g., cohort sequential design)
  - Participant based planned missing data designs
    - Randomly assign participants to receive only a subset of items

# Missing data opportunity: Planned missing data

- For dyadic data both planned missing data designs can be used
  - Participant based designs to reduce questionnaire length (e.g. 3-Forms planned missing data designs)
  - Time based planned missing data designs (e.g. control when dyads are measured in a longitudinal study)

# Missing data opportunity: Planned missing data

- A third type of planned missing data design is possible with dyadic data: dyad based planned missing
  - Some dyads have data from both members
  - Some dyads have data only collected from one dyad member
- This design can lead to cost savings/power increases compared to assessing all dyad members

# Missing data opportunity: Dyadic planned missing data

| Dyad | Person 1 | Person 2 |
|---|---|---|
| 1 | X | X |
| 2 | X | X |
| 3 | X | O |
| 4 | O | X |

# Missing data opportunity: Dyadic planned missing data

- Missing data in this designs can be assigned or naturally occurring
  - When missingness is assigned (dyads are randomly assigned to have 1 or 2 members measures) missingness is MCAR
  - When missingness is natural (only 1 member of a dyad responds) missingness is (probably) MAR or MNAR
    - Researchers should measure dyad/partner variables related to non-response

# Missing data opportunity: Dyadic planned missing data

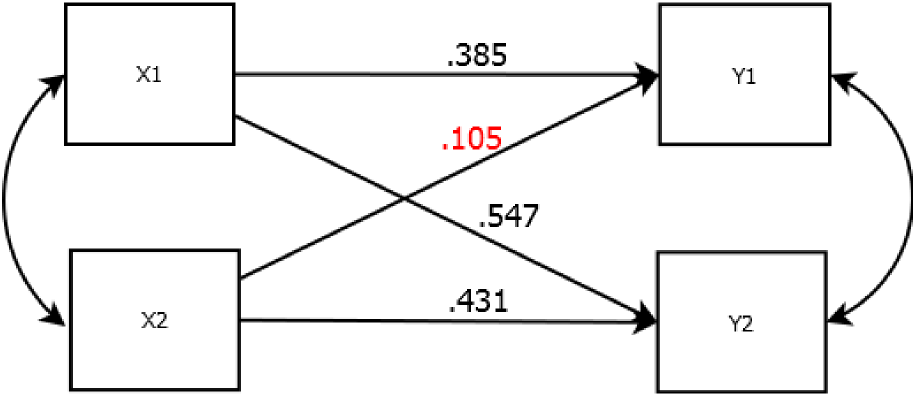
- Missing data should (if possible) be balanced across individuals
  - Equal missing for both members in distinguishable dyads
  - Missingness equally distributed across members for indistinguishable dyads
    - Data management

# Missing data opportunity: Dyadic planned missing data

- Example power analysis with MCAR planned missing data
  - Total budget $10000
    - Dyads: $50 per dyad
    - Singles: $10 per person
    - With no planned missing n = 200 dyads

# Missing data opportunity: Dyadic planned missing data

- Population model



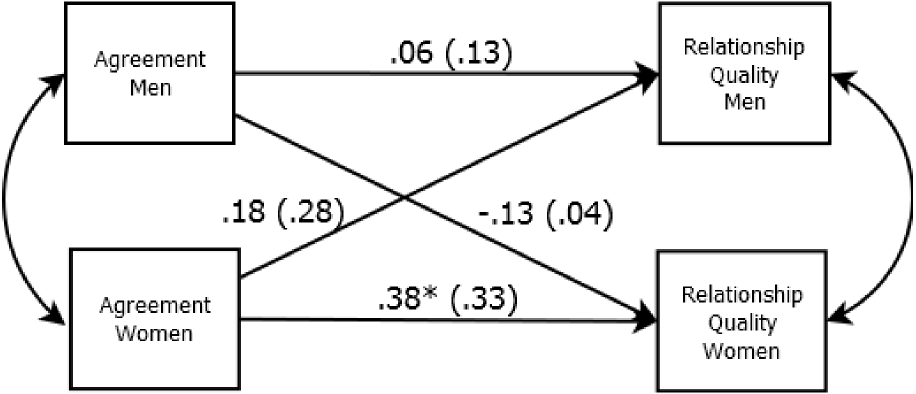
# Missing data opportunity: Dyadic planned missing data

| N dyad | N individuals | Power |
|---|---|---|
| 200 | 0 | .637 |
| 190 | 50 | .643 |
| 176 | 120 | .667 |
| 111 | 445 | .552 |

# Missing data opportunity: Dyadic planned missing data

- Example: Relationship norms and relationship quality
  - Heterosexual couples surveyed on endorsement of relationship norms and relationship quality
    - Focus on agreement about relationship norms
  - 25 heterosexual couples, 11 individuals in relationships without partner data (evenly split between men and women)

# Missing data opportunity: Dyadic planned missing data




Values in parentheses use only complete dyads

# Future directions

- Provide guidance on dyadic planned missing data designs
  - Ratio of dyads to singles
  - Distribution of singles across dyad members
- Determine optimal methods for MI with indistinguishable dyads
  - Impact of MAR mechanisms

# Thank you!

- Slides from today at:

`http://MARlab.org/Supplemental_Materials/`

- email: `schoemanna@ecu.edu`

# References

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). The analysis of dyadic data. *New York: Guilford.*

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological methods, 11*, 127.

# Simulation population values I

```
## This is lavaan 0.5-20

## lavaan is BETA software! Please report any bugs.

##              lhs op          rhs label    est
## 1  SATISFACTION  ~    ACT_HOUSE     a -0.591
## 2  SATISFACTION  ~   PART_HOUSE     p  0.888
## 3        PSATIS  ~    ACT_HOUSE     p  0.888
## 4        PSATIS  ~   PART_HOUSE     a -0.591
## 5  SATISFACTION ~~ SATISFACTION    v2  2.382
## 6        PSATIS ~~       PSATIS    v2  2.382
## 7     ACT_HOUSE ~~    ACT_HOUSE    v1  1.060
## 8    PART_HOUSE ~~   PART_HOUSE    v1  1.060
## 9     ACT_HOUSE ~~   PART_HOUSE        0.417
## 10 SATISFACTION ~1               int1  4.791
```

# Simulation population values II

```
## 11       PSATIS ~1              int1  4.791
## 12    ACT_HOUSE ~1              int2  1.630
## 13   PART_HOUSE ~1              int2  1.630
## 14 SATISFACTION ~~      PSATIS         0.812
```