What is ACCORDS?

ACCORDS is a ‘one-stop shop’ for pragmatic research:

- A multi-disciplinary, collaborative research environment to catalyze innovative and impactful research
- Strong methodological cores and programs, led by national experts
- Consultations & team-building for grant proposals
- Mentorship, training & support for junior faculty
- Extensive educational offerings, both locally and nationally

ACCORDS Upcoming Events

February 7, 2024 Bushnell Auditorium, Zoom	**Ethics, Challenges, & Messy Decisions in Shared Decision Making** Financial Toxicity and the Importance of Cost Discussions During Shared Decision Making *Presented by:* Mary Politi, PhD (Washington University in St. Louis)
February 26, 2024 Zoom	**Statistical Methods for Pragmatic Research** Latent Class Analysis: Assumptions and Extensions *Presented by:* Rashelle Musci, PhD (Johns Hopkins Bloomberg School of Public Health)
March 6, 2024 Bushnell Auditorium, Zoom	**Ethics, Challenges, & Messy Decisions in Shared Decision Making** Health Equity and Shared Decision Making *Presented by:* Channing Tate, PhD, MPH; Demetria Bolden, PhD, MBA; Lucinda Kohn, MD, MHS, Miria Kano, PhD
March 11, 2024 AHSB 2200/2201, Zoom	**Statistical Methods for Pragmatic Research** Pragmatic Statistical Learning: From Data to Interpretable Insights *Presented by:* Ryan Peterson, PhD & Kathryn Colborn, PhD

*all times 12-1pm MT unless otherwise noted

COPRH Con
Colorado Pragmatic Research in Health Conference

Innovations in Pragmatic Research Methods

From Data to Equity, Policy, and Sustainability

June 5 - 6, 2024 | 10am-3pm MT

ACCORDS
ADULT AND CHILD CENTER FOR OUTCOMES RESEARCH AND DELIVERY SCIENCE
UNIVERSITY OF COLORADO
CHILDREN'S HOSPITAL COLORADO

Registration is open now at www.COPRHCon.com

Jun Ying, PhD

Missing Data and Multiple Imputation

Missing Data and Multiple Imputation

Jun Ying, PhD

January 22nd, 2024

Outline

- Backgrounds
- Definition, Sources, Patterns and Mechanisms
- Prevention
- Diagnosis
- Traditional Methods
- Contemporary Methods
  - Maximum Likelihood Method
  - Multiple Imputation
  - Pattern Mixture Model
  - Sensitivity Analysis
- Summary
- Acknowledge

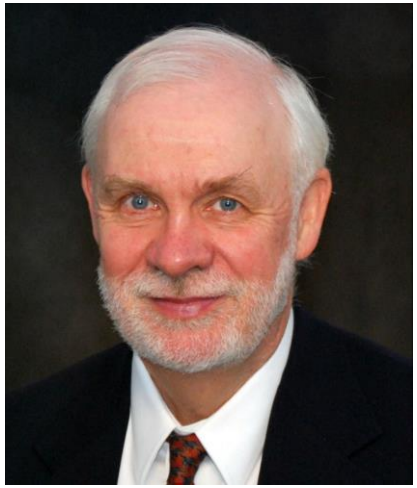
Background

- Early work (1920's - 1970's)
  - Karl Pearson, Ronald Aylmer Fisher, Egon Pearson
  - Imbalanced sample size due to incomplete data
- Theoretical development
  - Rubin 1976, Dempster 1977, Rubin 1987, and Little and Rubin 1987
  - Missing data as a field of research in statistics
- Application in clinical trials
  - EMA (European Medicines Agency) 2009
    - Guideline on missing data in confirmatory clinical trials
  - NAS (National Academy of Sciences) 2010
    - The prevention and treatment of missing data in clinical trials
    - FDA initiated and supported
    - Also published in NEJM 2012 Little *et al*
  - PCORI (Patient-Centered Outcomes Research Institute) 2012
    - Minimal Standards in Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research

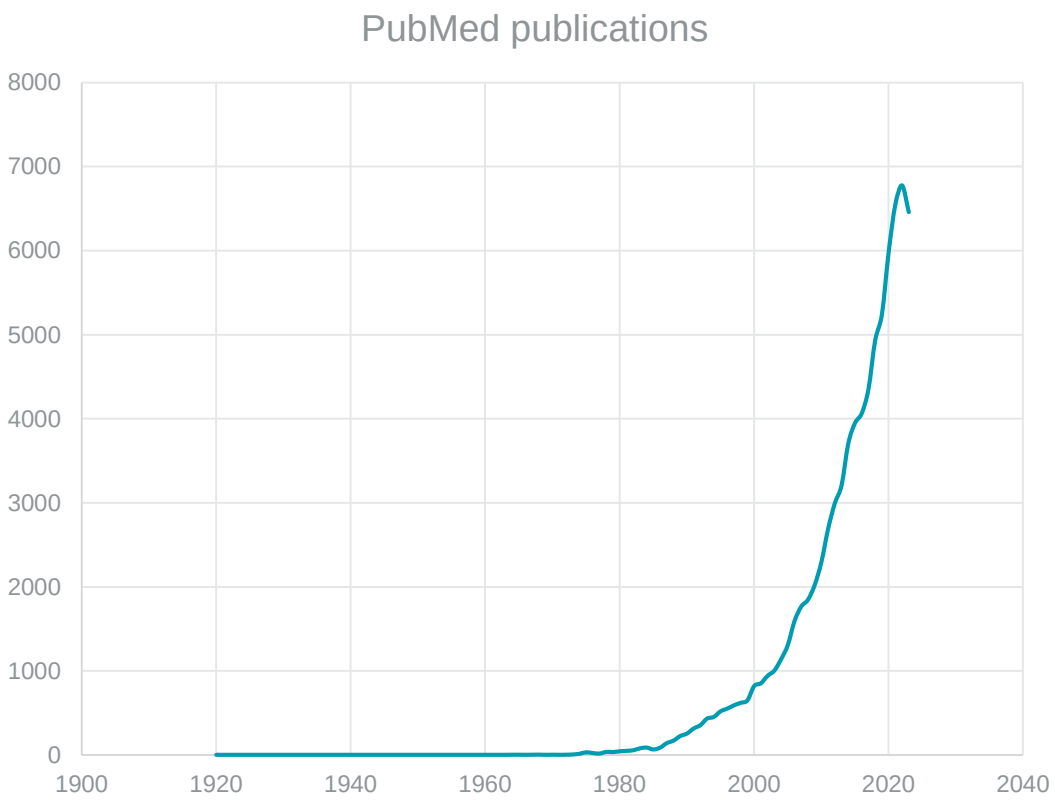
Background

- Donald B. Rubin, Professor of Statistics, Harvard University
  - Rubin, D.B. (1976). Inference and missing data. Biometrika, 63, 581-592.
  - Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

- Roderick J. Little, Professor of Biostatistics, University of Michigan
  - Little, R.J.A., & Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons
  - Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83, 1198-1202.



Background



Key Words	# of Publications in 2023
Casual Inference	20
Structure Equation Model	5,831
Missing Data	**6,459**
Multiple Comparisons	7,288
Bayesian	7,401
Machine Learning	33,076
Artificial Intelligence	36,967

Definition

- Missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest.
  - Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol. 2009;60:549–576

- Missing data *does not refer* to:
  - Observations with no data (*missing observations)*
  - Unobserved and/or unobservable variables

Definition

☹

ID	Y1	Y2	Y3
1	X	X	
2	X	X	
3	X	X	
4	X	X	
5	X	X	
6	X	X	
7	X	X	

☹

ID	Y1	Y2	Y3
1	X	X	X
2	X	X	X
3			
4	X	X	X
5	X	X	X
6	X	X	X
7	X	X	X

☺

ID	Y1	Y2	Y3
1	X	X	X
2	X		X
3	X	X	X
4	X	X	
5	X		X
6	X	X	X
7	X	X	X

Sources of Missing Data

- Clinical trials and longitudinal cohort studies
  - Missed appointments/follow-up visits
  - Dropout from study
- Survey Questionnaires and Instruments
  - Item non-response
    - *lack of knowledge required to answer question*
    - *question sensitive in nature or intrusive*
    - *burden on subject*
  - Structural “missing” data
    - *Voting preferences for ineligible voters*
- Others
  - Technical problems (equipment failure)

Patterns of Missing Data

- Pattern: Configuration of observed and missing values in the data set

Monotone Pattern			
Y1	Y2	Y3	Y4
X	X	X	X
X	X	X	X
X	X	X	X
X	X	X	.
X	X	.	.
X	.	.	.

Unit Non-Response Pattern			
Y1	Y2	Y3	Y4
X	X	X	X
X	X	X	X
X	X	X	X
X	X	X	X
X	X	.	.
X	X	.	.

General Pattern			
Y1	Y2	Y3	Y4
X	X	X	.
X	.	X	X
X	X	.	X
.	X	X	X
.	X	X	X
X	X	.	X

Planned Pattern			
Y1	Y2	Y3	Y4
X	.	X	X
X	.	X	X
X	X	.	X
X	X	.	X
X	X	X	.
X	X	X	.

Mechanisms (Types) of Missing Data

- Missing at Random (MAR)
  - The probability of missing data on a variable Y is related to the observed components of Y (Y^o) and other observed variables.
  - $P(R|Y^o)$
- Missing Completely at Random (MCAR)
  - The probability of missing data on a variable Y is not related any observed variables including itself.
  - P(R|.)
- Missing Not at Random (MNAR)
  - The probability of missing data on a variable Y is related to the missing components of Y (Y^m).
  - $P(R|Y^m)$

Mechanisms (Types) of Missing Data

Full Data	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	14
16	14
16	18

MCAR Pr(R|.)		
X	Y	R
4	9	0
6	19	0
7		1
8	6	0
9		1
9	11	0
10	12	0
14	14	0
14	16	0
15	14	0
16	14	0
16		1

MAR $P(R\|X, Y^0)$		
X	Y	R
4	9	0
6	19	0
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14	16	0
15		1
16		1
16		1

MNAR $P(R\|Y^m)$		
X	Y	R
4	9	0
6		1
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14		1
15	14	0
16	14	0
16		1

Consequences due to missing data

- **Biased results**
- Loss of power
- Complication in data handling, computation and analysis.

Mechanisms (Types) of Missing Data

Full Data	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	14
16	14
16	18

MCAR Pr(R|.)		
X	Y	R
4	9	0
6	19	0
7		1
8	6	0
9		1
9	11	0
10	12	0
14	14	0
14	16	0
15	14	0
16	14	0
16		1

MAR $P(R\|X, Y^0)$		
X	Y	R
4	9	9
6	19	19
7	14	14
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14	16	0
15		1
16		1
16		1

MNAR $P(R\|Y^m)$		
X	Y	R
4	9	0
6		1
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14		1
15	14	0
16	14	0
16		1

	Full Data	MCAR	MAR	MNAR
$\overline{Y}$	12.83	12.78	12	11.22

Preventing Missing Data

Table 1. Eight Ideas for Limiting Missing Data in the Design of Clinical Trials.

Target a population that is not adequately served by current treatments and hence has an incentive to remain in the study.
Include a run-in period in which all patients are assigned to the active treatment, after which only those who tolerated and adhered to the therapy undergo randomization.
Allow a flexible treatment regimen that accommodates individual differences in efficacy and side effects in order to reduce the dropout rate because of a lack of efficacy or tolerability.
Consider add-on designs, in which a study treatment is added to an existing treatment, typically with a different mechanism of action known to be effective in previous studies.
Shorten the follow-up period for the primary outcome.
Allow the use of rescue medications that are designated as components of a treatment regimen in the study protocol.
For assessment of long-term efficacy (which is associated with an increased dropout rate), consider a randomized withdrawal design, in which only participants who have already received a study treatment without dropping out undergo randomization to continue to receive the treatment or switch to placebo.
Avoid outcome measures that are likely to lead to substantial missing data. In some cases, it may be appropriate to consider the time until the use of a rescue treatment as an outcome measure or the discontinuation of a study treatment as a form of treatment failure.

Source: Little RJ, et al. The Prevention and Treatment of Missing Data in Clinical Trials. NEJM 36(14):1355-60 (2012)

Preventing Missing Data

Table 2. Eight Ideas for Limiting Missing Data in the Conduct of Clinical Trials.

Select investigators who have a good track record with respect to enrolling and following participants and collecting complete data in previous trials.
Set acceptable target rates for missing data and monitor the progress of the trial with respect to these targets.
Provide monetary and nonmonetary incentives to investigators and participants for completeness of data collection, as long as they meet rigorous ethical requirements.[15,16]
Limit the burden and inconvenience of data collection on the participants, and make the study experience as positive as possible.
Provide continued access to effective treatments after the trial, before treatment approval.
Train investigators and study staff that keeping participants in the trial until the end is important, regardless of whether they continue to receive the assigned treatment. Convey this information to study participants.
Collect information from participants regarding the likelihood that they will drop out, and use this information to attempt to reduce the incidence of dropout.
Keep contact information for participants up to date.

Source: Little RJ, et al. The Prevention and Treatment of Missing Data in Clinical Trials. NEJM 36(14):1355-60 (2012)

Diagnose of Missing Data Mechanisms

- Little's MCAR test and T-tests can be used to test if the data is MCAR or not.
- There is no statistical test to tell the difference between MAR and MNAR.

Diagnose of Missing Data Mechanisms: T-Test

X	Z	Y	R_Y
78	13		1
84	9		1
84	10		1
85	10		1
87			1
91	3		1
92	12		1
94	3		1
94	13		1
96			1
99	6	7	0
105	12	10	0
105	14	11	0
106	10	15	0
108		10	0
112	10	10	0
113	14	12	0
115	14	14	0
118	12	16	0
134	11	12	0

R_Y	X		Z	
	Mean	Std Dev	Mean	Std Dev
0	111.50	9.70	11.44	2.60
1	87.89	5.78	8.57	4.04
T score	6.35		1.73	
P value	**<.0001**		**0.11**	

Diagnose of Missing Data Mechanisms: Little's MCAR test

X	Z	Y	Pattern
78	13		2
84	9		2
84	10		2
85	10		2
87			1
91	3		2
92	12		2
94	3		2
94	13		2
96			1
99	6	7	4
105	12	10	4
105	14	11	4
106	10	15	4
108		10	3
112	10	10	4
113	14	12	4
115	14	14	4
118	12	16	4
134	11	12	4

Little's MCAR test:

- D_i^2= the distance between a pattern "center" i and the global "center";
- D^2 = the sum of all distances D_i^2
- D^2 ~ Chisq Distribution

D^2 = 14.63, df=5, P-value = 0.01

Missing Data Treatments

- Traditional Methods
  - Deleting Cases
  - Imputation Methods
- Contemporary Methods
  - Maximum Likelihood Estimation (MLE) Methods
  - Multiple Imputation (MI) Methods
  - Bayesian Method

Traditional Methods

- Case Deletion
  - Listwise
  - Pairwise
- Single Imputation
  - Mean Imputation
  - Regression (Expected Value) Imputation
  - Stochastic Regression Imputation
  - Hot Deck
  - Similar Response Pattern Imputation
  - Average of Available Items
  - Last Observation Carried Forward

Traditional Methods: Case Deletion

Listwise

ID	Y1	Y2	Y3	Y4
1	X	X	X	.
2	X	.	X	X
3	X	X	.	X
4	X	X	X	X
5	X	X	X	X
6	X	X	X.	X

ID	Y1	Y2	Y3	Y4
4	X	X	X	X
5	X	X	X	X
6	X	X	X.	X

Pairwise

ID	Y1	Y2	Y3	Y4
1	X	X	X	.
2	X	.	X	X
3	X	X	.	X
4	X	X	X	X
5	X	X	X	X
6	X	X	X.	X

ID	Y1	Y2
1	X	X
3	**X**	X
4	X	X
5	X	X
6	X	X

ID	Y1	Y3
1	X	**X**
2	**X**	X
4	X	X
5	X	X
6	X	X.

ID	Y3	Y4
2	X	X
4	X	X
5	X	X
6	X.	X

Traditional Methods: Case Deletion

- Possibly Biased
- Reduced Power
- Inconsistent Results

Traditional Methods: Mean Imputation

Full Data	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	14
16	14
16	18

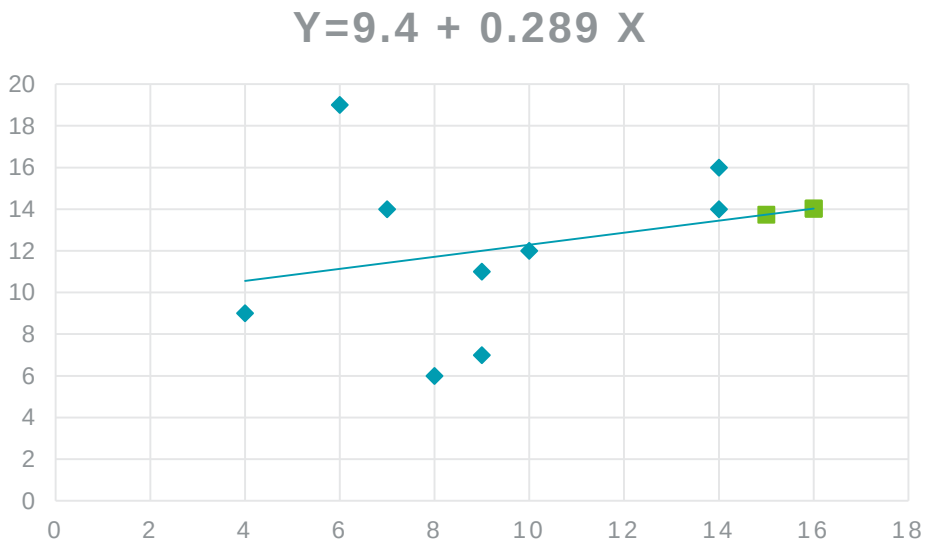
MAR	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	.
16	.
16	.

Imputed	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	12
16	12
16	12

	Full Data	MAR	Imputed
$\overline{Y}$	12.83	12	12
Std	4.04	4.24	3.62

Traditional Methods: Regression Imputation

MAR	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	.
16	.
16	.



Imputed2	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	13.74
16	14.02
16	14.02

	Full Data	MAR	Imputed2
$\overline{Y}$	12.83	12	12.48
Std	4.04	4.24	3.72

Contemporary Methods: MLE

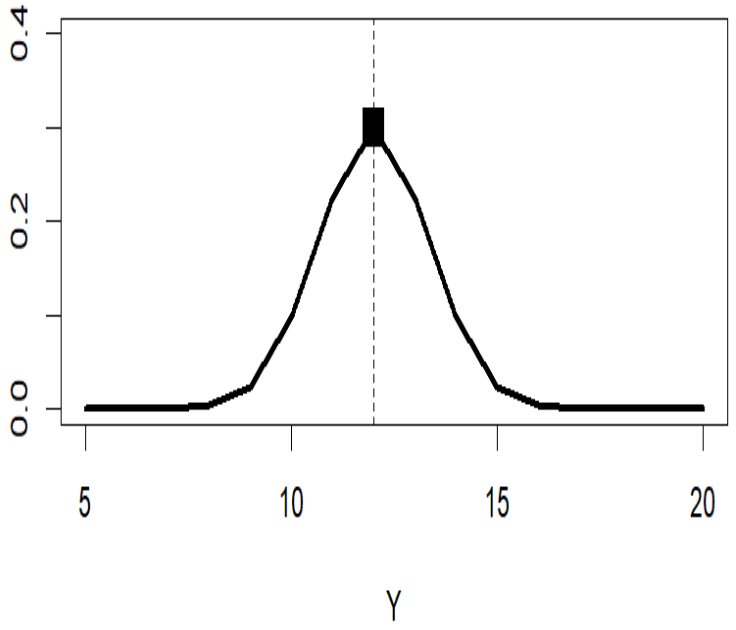
- (Traditional) Maximum Likelihood Method
  - Available in most software packages
  - Using listwise deletion for missing data
  - **Unbiased only under MCAR**
  - Some packages used to handle MAR under special conditions (such as longitudinal data or clustered data)
- FIML (full information maximum likelihood) method
  - Considered a “gold standard” MLE method
  - Performed well under MAR, or MCAR
  - Widely available in structure question model (SEM) software packages
    - MPLUS, LISREL, SAS Calis, SPSS, R lavaan
- EM Algorithm
  - Performed well under MAR, or MCAR
  - Available in most SEM software packages

MLE Method

- Traditional MLE uses listwise deletion of missing data
- Unbiased if MCAR, otherwise not.

MCAR Pr(R|.)	
X	Y
4	9
6	19
7	
8	6
9	
9	11
10	12
14	14
14	16
15	14
16	14
16	

MCAR Pr(R|.)	
X	Y
4	9
6	19
8	6
9	11
10	12
14	14
14	16
15	14
16	14



FIML

- MLE computed from both full cases and incomplete cases
- Performed well under MAR, or MCAR

Mean of X=10.67 Mean of Y = 11.5		
X	Y	LL
4	9	-6.08
6	19	-7.39
7	14	-5.37
8	6	-5.58
9	7	-5.23
9	11	-4.74
10	12	-4.67
14	14	-5.09
14	16	-5.42
15		-2.92
16		-3.23
16		-3.23

X	Y	LL
10.66667	11.5	**-58.9589**
10.66667	12	-58.8055
10.66667	**12.5**	**-58.7819**
10.66667	13	-58.8881
11	11.5	-59.0201
11	12	-58.849
11	12.5	-58.8077
11	13	-58.8962

EM Algorithm

- E Step: Fill the missing cells with expected values
- M Step: Estimate MLE under the “full data”, and calculate the LL
- Iterate the steps until no more improvement of LL.

E-step #1

X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	10
16	10
16	10

M-step #1

Mean Y	-2LL
11.5	63.4

E-step #2

X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	11.41
16	11.39
16	11.39

M-step #2

Mean Y	-2LL
11.8	62.7

E-step #3

X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	12.23
16	12.32
16	12.32

M-step #3

Mean Y	-2LL
12.1	62.5

E-step #4

X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	12.76
16	12.92
16	12.92

MLE Example

MAR	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	
16	
16	

Stats	Full Data	FIML	EM	MLE (listwise)
Mean of Y	12.83	12.48	12.48	12

Contemporary Methods: MI

- First proposed by Rubin (1977) and elaborated in his (1987) book
- Assuming MAR
- Available in many software packages:
  - SAS MI, R Mice (Amelia, missForest, mi), MPLUS, SPSS
- Three Phase Approach:
  - **Imputation:** Impute missing values using an appropriate model that incorporates random variation. Do this M times producing M "complete" data sets.
  - **Analysis:** Estimate the unknown parameter θ_i (e.g. mean, slope) and its standard error SE(θ_i) on each data set using standard complete-data methods
  - **Pooling:** combine estimations from all datasets to find the final or pooled estimates of unknow parameters.
    - $\theta^* = \frac{\sum_{i=1}^{M} \theta_i}{M}$
    - $Se(\theta^*) = \sqrt{V(\theta^*)} = \sqrt{V_W + \frac{M+1}{M} V_B}$
    - Where $V_W = \frac{1}{M}\sum_{i=1}^{M} Se(\theta_i)^2$, and $V_B = \frac{1}{M-1}\sum_{i=1}^{M}(\theta_i - \theta^*)^2$

Imputation Method: FCS

- FCS (Fully conditional specification)

Example:

$$Y_4 = \alpha_4 + \beta_{41}Y_1 + \beta_{42}Y_2 + \beta_{43}Y_3$$
$$Y_3 = \alpha_3 + \beta_{31}Y_1 + \beta_{32}Y_2$$
$$Y_2 = \alpha_2 + \beta_{21}Y_1$$
$$Y_1 = \alpha_1 + \beta_{11}Y_4$$

General Pattern			
Y1	Y2	Y3	Y4
X	X	X	.
X	.	X	X
X	X	.	X
.	X	X	X
.	X	X	X
X	X	.	X

Imputation Method: MCMC

- MCMC (Bayesian Markov chain Monte Carlo)

Example:

1. Likelihood function:
   $Y \sim MVN\ (\mu, \Sigma)$

2. Prior distribution:
   $\mu \sim MVN\ (\mu_0, \Sigma_0)$
   $\Sigma \sim W^{-1}(P_0, \Lambda_0)$

3. Posterior distribution:
   $\mu|. \sim MVN\ (\mu^*, \Sigma^*)$
   $\Sigma|. \sim W^{-1}\ (P^*, \Lambda^*)$
   $Y^m \sim MVN\ (\mu|., \Sigma|)$

General Pattern			
Y1	Y2	Y3	Y4
X	X	X	.
X	.	X	X
X	X	.	X
.	X	X	X
.	X	X	X
X	X	.	X

Imputation Method: Monotone

- Monotone (e.g. dropout in longitudinal studies)

Example:

Step1: Y1 → Y2
Step2: Y1, Y2 → Y3
Step3: Y1, Y2, Y3 → Y4

Monotone Pattern			
Y1	Y2	Y3	Y4
X	X	X	X
X	X	X	X
X	X	X	X
X	X	X	.
X	X	.	.
X	.	.	.

MI Example (using FCS)

MAR	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	
16	
16	

MI #1	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	16.4
16	16.1
16	15.9

MI #2	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	17.2
16	11.5
16	20.7

MI #3	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	14.8
16	11.5
16	7.8

MI #4	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	27.4
16	13.6
16	20.6

MI #5	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	12.3
16	17.5
16	11.7

Stats	MI #1	MI #2	MI #3	MI #4	MI #5	Pooled
Mean	13.03	13.12	11.84	14.13	12.46	**12.92**
SE	1.18	1.33	1.13	1.75	1.14	**1.62**

MI Example

MAR	
X	Y
4	9
6	19
7	14
8	6
9	7
9	11
10	12
14	14
14	16
15	
16	
16	

Stats	Full Data	MI (FCS)	MI (MCMC)	FIML	EM	MLE (listwise)
Mean of Y	12.83	12.92	12.78	12.48	12.48	12

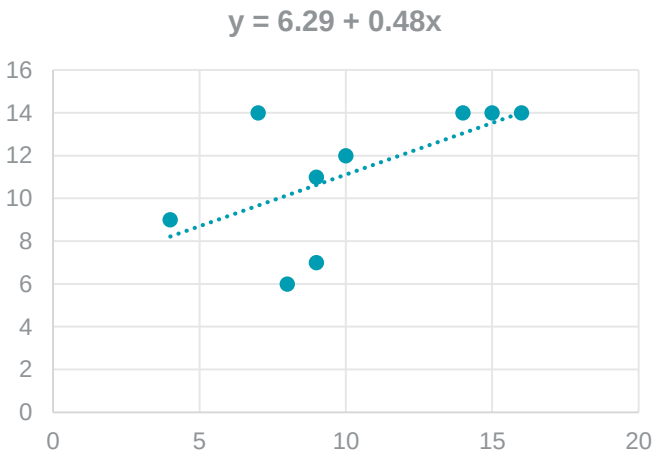
PMM (Pattern Mixture Model)

- Used to handle MNAR data
- Procedures:
  - Find patterns of missing data
  - Estimate unknown parameters (such as mean, slope, etc) within the pattern
  - Obtain a pooled estimate from pattern level estimates and weighted by distributions of the patterns

PMM Example

MNAR		
X	Y	R
4	9	0
6		1
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14		1
15	14	0
16	14	0
16		1

MNAR		
X	Y	R
4	9	0
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
15	14	0
16	14	0



MNAR		
X	Y	R
6	9.18	1
14	13.05	1
16	14.01	1

Stats	R=0	R=1	Pooled
Mean of Y	11.22	12.08	11.44 =11.22x75%+12.08x25%

MNAR Example

MNAR		
X	Y	R
4	9	0
6		1
7	14	0
8	6	0
9	7	0
9	11	0
10	12	0
14	14	0
14		1
15	14	0
16	14	0
16		1

Stats	Full Data	PMM	MI (FCS)	FIML	MLE (listwise)
Mean of Y	12.83	11.58	11.68	11.44	11.22

Sensitivity Analysis

- In practice, one can not tell the difference between MAR and MNAR data.
- MI or MLE methods perform well under MAR but not MNAR.
- To reduce the risk of biasness due to misidentifying the missing mechanism, one should consider sensitivity analysis.
- Rubin (1987) suggested to inflate (deflate) the imputed values by 20%.
- The idea is to inspect if the findings still hold under some “extreme” conditions.
- Similarly, in analyses based upon MCAR assumption, one needs to consider a sensitivity analysis under MAR.

Sensitivity Analysis Example

MNAR	
X	Y
4	9
6	
7	14
8	6
9	7
9	11
10	12
14	14
14	
15	14
16	14
16	

PMM	
X	Y
4	9
6	9.18
7	14
8	6
9	7
9	11
10	12
14	14
14	13.05
15	14
16	14
16	14.01

20% Inflation	
X	Y
4	9
6	11.02
7	14
8	6
9	7
9	11
10	12
14	14
14	15.65
15	14
16	14
16	16.81

Stats	Full Data	MLE	PMM	Sensitivity (20% inflation)
Mean of Y	12.83	11.22	11.44	12.04

Summary

- Missing data problems exist in many studies including clinical trials, observational studies and survey studies.
- Ignoring missing data could cause biased results
- Mishandling missing data could also have the consequences of biasness and inflated Type I error.
- Multiple imputation is currently the most widely used method in handling missing data problems.
- Still, since there is no gold standard in this research area, one should always be careful to choose appropriate methods, and use a sensitivity analysis to check the robustness of findings.

Acknowledgement

- Hermine Brunner, MD, Cincinnati Children's Hospital Medical Center
- Carl Fichtenbaum, MD, University of Cincinnati, College of Medicine
- Mario Shootman, PhD, University of Arkansas for Medical Sciences
- Xia Wang, PhD, University of Cincinnati, Department of Mathematics
- Changchun Xie, PhD, University of Cincinnati, College of Medicine
- Nanhua Zhang, PhD, Cincinnati Children's Hospital Medical Center
- Rong Zhou, PhD, MedPace