

***If data are going to be missing
anyway, why not plan for it?***

Kenneth L. Thompson, Ph.D.

March 31, 2016



MILLSAPS

COLLEGE

OUR REALITY

In our world, missing data is a reality.

ID	Race	Gender	Age	Major	Q1	Q 2	Q3	Q4	Q5	Q6	Q7
01	B		10	BUS	2	4	1	4	3	.	2
02	H	F		EDU	1	4	5	1	3	2	5
03	W	M	18	PHY	3	.	4	.	.	1	.
04	H		18	ANT	2	4	1	4	3	.	2
05	B	F	19		1	2	3	.	1	2	.
06		M			.	5	5	2	4	1	.
07	W	M	19	EDU	3	1	4	3	1	5	2
08	A	F	20		3	.	4	.	.	1	.
09	W		18	MTH	1	4	5	1	3	2	5
---	---	---	---	---	---	---	---	---	---	---	---



WHY CARE ABOUT MISSING DATA?

Why are the data missing?

Confusing items

Excessively long questionnaires

Is it advisable to ignore the missing data – how random is it?

Predictable Nonresponse pattern - MAR

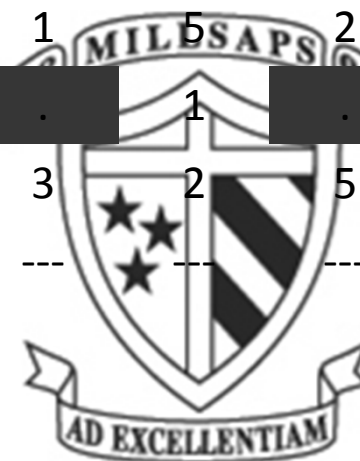
Unpredictable Nonresponse pattern – MCAR or MNAR



OUR REALITY

In our world, missing data is a reality.

ID	Race	Gender	Age	Major	Q1	Q 2	Q3	Q4	Q5	Q6	Q7
01	B		10	BUS	2	4	1	4	3	.	2
02	H	F		EDU	1	4	5	1	3	2	5
03	W	M	18	PHY	3	.	4	.	.	1	.
04	H		18	ANT	2	4	1	4	3	.	2
05	B	F	19		1	2	3	.	1	2	.
06		M			.	5	5	2	4	1	.
07	W	M	19	EDU	3	1	4	3	1	5	2
08	A	F	20		3	.	4	.	.	1	.
09	W		18	MTH	1	4	5	1	3	2	5
---	---	---	---	---	---	---	---	---	---	---	---



MISSING DATA CAN BE A PROBLEM!

Can decrease the power of your analysis

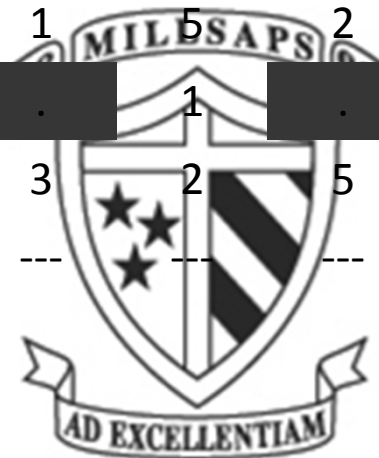
Observations are lost and reliability suffers.



OUR REALITY

In our world, missing data is a reality.

ID	Race	Gender	Age	Major	Q1	Q 2	Q3	Q4	Q5	Q6	Q7
01	B		10	BUS	2	4	1	4	3	.	2
02	H	F		EDU	1	4	5	1	3	2	5
03	W	M	18	PHY	3	.	4	.	.	1	.
04	H		18	ANT	2	4	1	4	3	.	2
05	B	F	19		1	2	3	.	1	2	.
06		M			.	5	5	2	4	1	.
07	W	M	19	EDU	3	1	4	3	1	5	2
08	A	F	20		3	.	4	.	.	1	.
09	W		18	MTH	1	4	5	1	3	2	5
---	---	---	---	---	---	---	---	---	---	---	---



OUR REALITY

In our world, missing data is a reality.

ID	Race	Gender	Age	Major	Q1	Q 2	Q3	Q4	Q5	Q6	Q7
02	H	F		EDU	1	4	5	1	3	2	5
07	W	M	19	EDU	3	1	4	3	1	5	2
09	W		18	MTH	1	4	5	1	3	2	5
---	---	---	---	---	---	---	---	---	---	---	---



CAN'T KEEP DOING BUSINESS THE SAME WAY

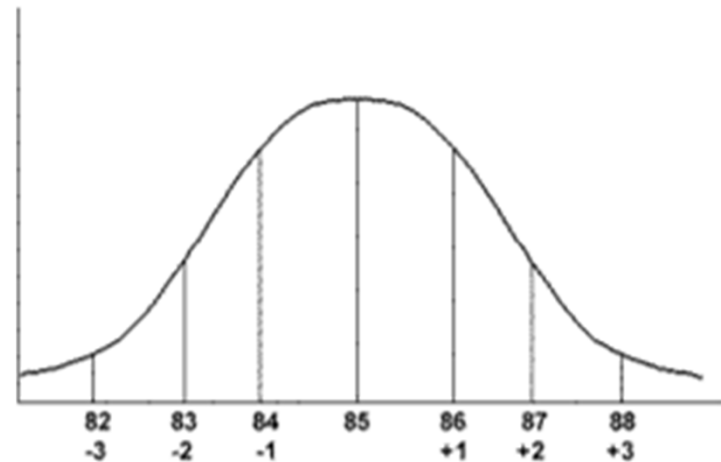
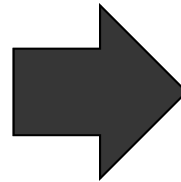
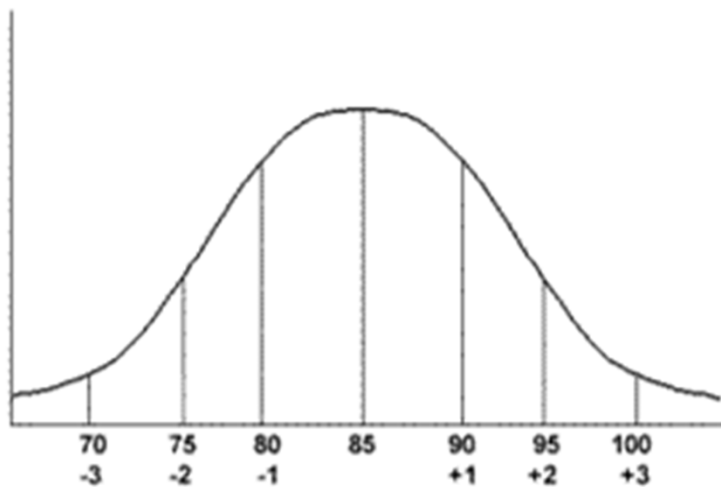
Deletion reduces power and generalizability

ID	Race	Gender	Age	Major	Q1	Q 2	Q3	Q4	Q5	Q6	Q7
07	W	M	19	EDU	3	1	4	3	1	5	2
---	---	---	---	---	---	---	---	---	---	---	---



CAN'T KEEP DOING BUSINESS THE SAME WAY

Substitution can bias results



SO, WHAT CAN WE DO?

Reduce the length of the instrument

Correct confusing items

But, you can also think outside the box by purposefully creating missing data leading to

- a shorter instrument

- reduced fatigue

- increased sample size



3-FORM PLANNED MISSING DATA DESIGN

Graham, Hofer, & Piccinin, 1994

Items split into 4 sets (X, A, B, C)

A participants get X plus 2 additional sets (XAB, XAC, XBC)

The 3-Form Design, With X Set

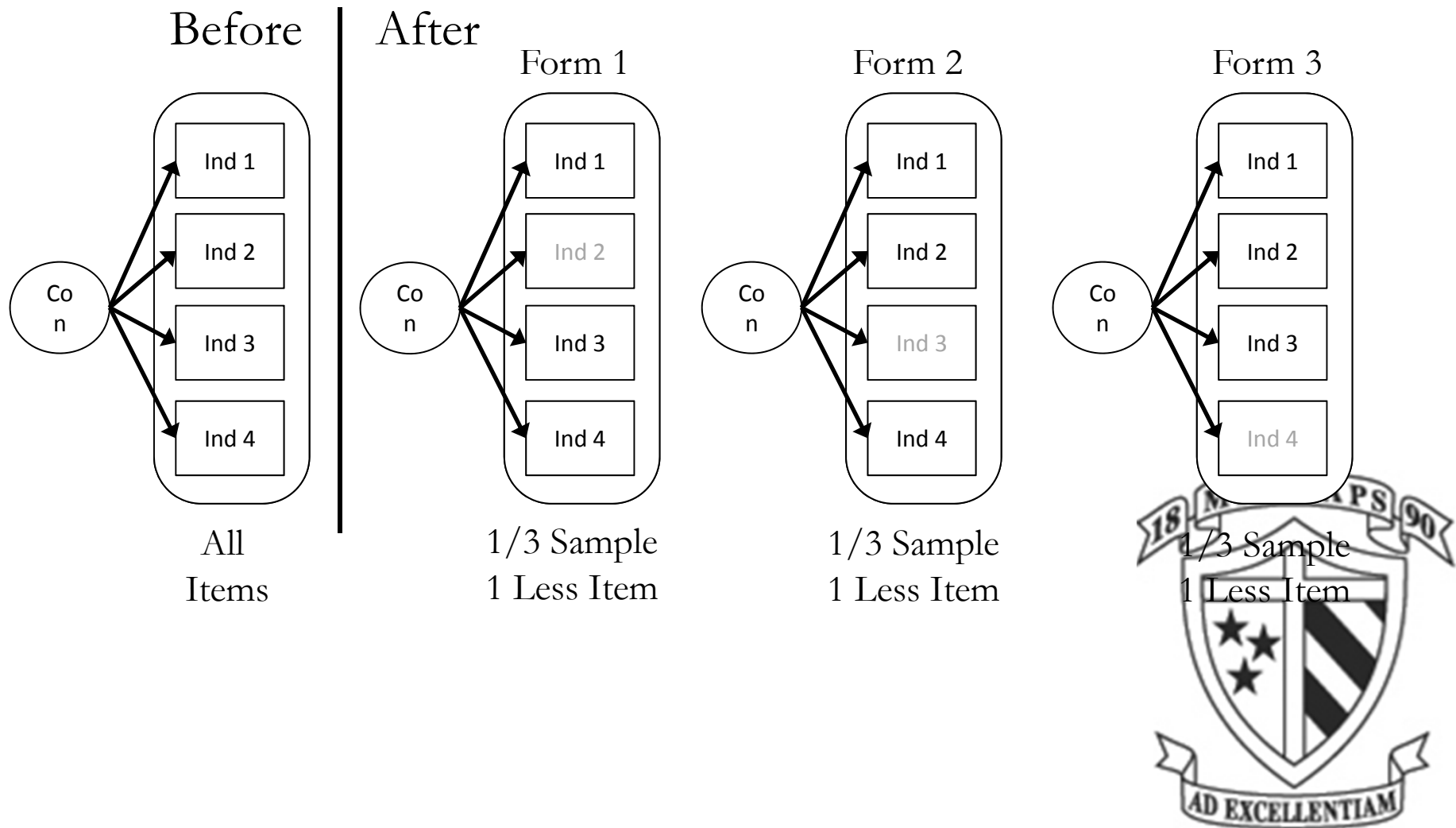
Form	Item set			
	X	A	B	C
1	1	1	1	0
2	1	1	0	1
3	1	0	1	1

Note. 1 = questions asked; 0 = questions not asked.

From Graham et al. (2006)



HOW CAN WE PLAN MISSING DATA?



HOW TO USE 3-FORM PMDD

Priors Unknown

- New Instrument
- Unknown Population

Common X	1	5	9
Set A	2	6	
Set B	3	7	
Set C	4	8	

Stacked

Priors Known

- Data in hand
- Can calculate r

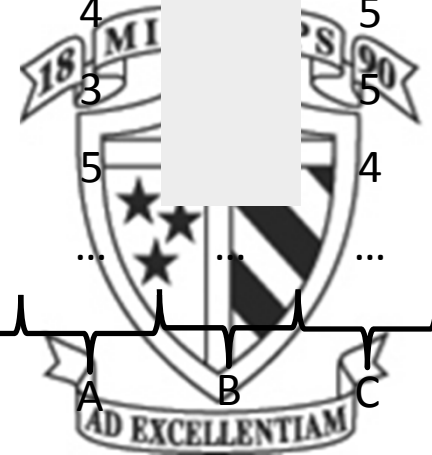
Common X	4		
Set A	7	2	9
Set B	1	3	8
Set C	6	5	

Serpentine



DATASET AFTER DATA COLLECTION

Demographics					Construct 1				Construct 2			
ID	Race	Gender	Age	Major	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8
1	W	M	19	Ele	98	95	99		5	5	4	
2	B	F	20	Early	100	99	100		5	4	5	
3	H	F	21	Sec	95	93	96		4	3	4	
4	H	M	20	ESL	85		83	84	3		3	4
5	B	M	19	Ele	94		92	93	4		3	4
6	B	F	18	SPED	85		82	81	3		3	2
7	A	F	21	ESL	93	91		91	4	4		5
8	A	F	22	Sci	99	97		100	5	3		5
9	W	F	20	Math	100	98		98	5	5		4
...	...	...	...	...	...	...	...	...	...	...	...	...
X					X	A	B	C	X	A	B	C



HOW DO WE REGAIN THE INFORMATION WE LOST?

Imputation

Restricted

- Requires advanced statistical skills – SEM (MPlus, Lisrel)

Unrestricted

- Requires data
- Uses information from each variable
- Most common



PREPARING DATA FOR UNRESTRICTED IMPUTATION

- Data in wide format (one case per row)
- Check for miscoding
- Code text-based variables
- Dummy code unordered categorical variables
- Always, always, always save two copies of the data set
 - Original
 - Working copy without identifiers, strings, or dates



IMPUTING THE DATA

SPSS

- Requires add-on module

R

- Mice (required coding)
- Amelia (GUI)



Thank You!

