Modern Strategies to Handle Missing Data: A Showcase of Research on Foster Children

Anouk Goemans, MSc
PhD student
Leiden University
The Netherlands

Email: a.goemans@fsw.leidenuniv.nl

Universiteit Leiden

מכון חרוב (ע"ר)
The Haruv Institute (R.A.)

Modern Strategies to Handle Missing Data:
A Showcase of Research on Foster Children

Issue:

Analysis of Data

How are you going to deal with missing data?

A. I will only have a small number of missing data, so I will not deal with this missing data
B. Pairwise deletion, listwise deletion or mean imputation
C. Multiple imputation or FIML estimation
D. I don’t know yet
E. Not applicable. I don’t have / will not have missing data at all

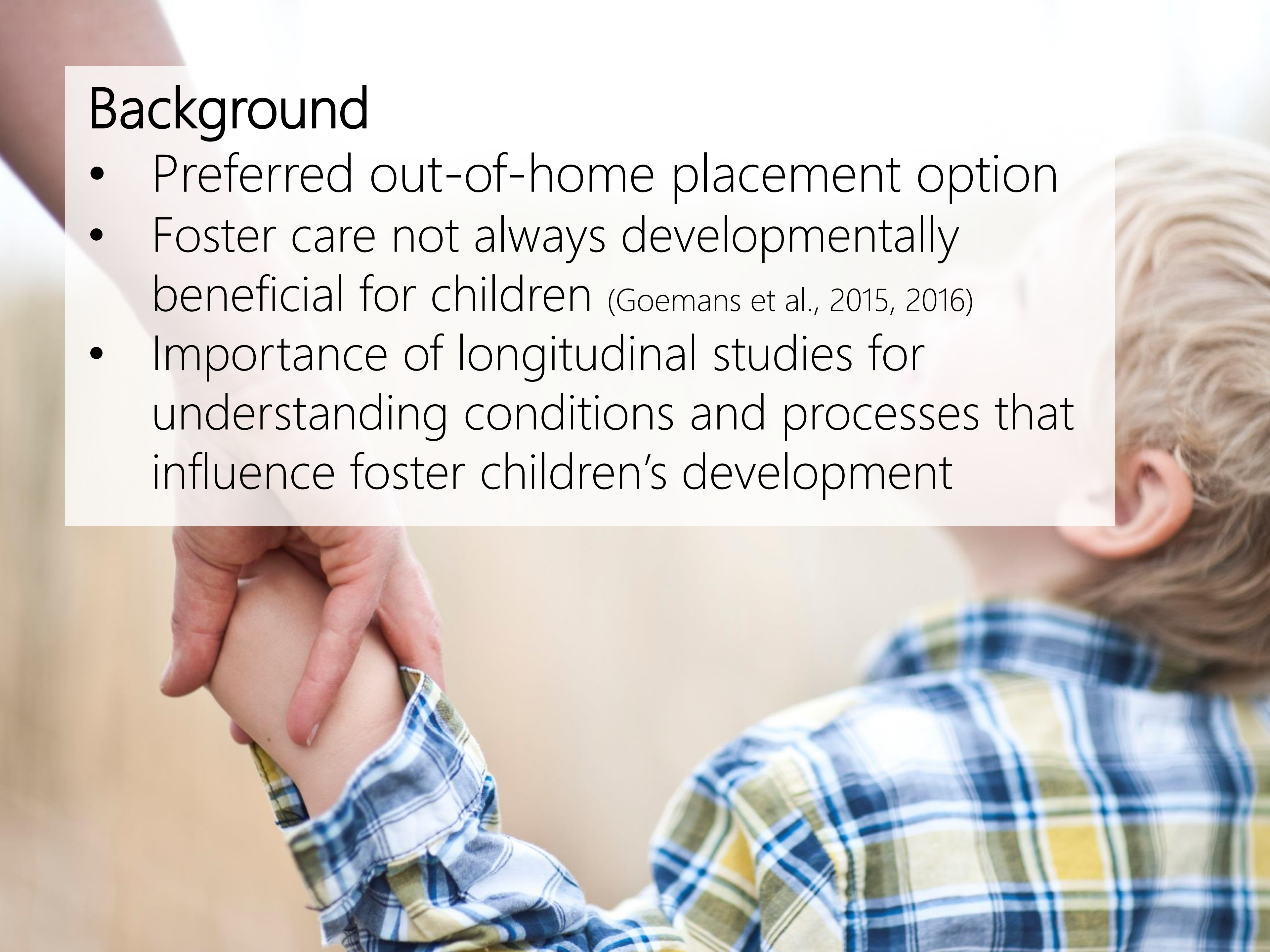
www.menti.com; Code: 94 74 33

Today's Outline:

1. My PhD study
2. Missing data: an introduction
3. Two examples
4. Practical guidelines
5. Summary & Discussion

Background

- Preferred out-of-home placement option
- Foster care not always developmentally beneficial for children (Goemans et al., 2015, 2016)
- Importance of longitudinal studies for understanding conditions and processes that influence foster children's development




Method:
online questionnaires
Design:
3-wave longitudinal study
Goal PhD study:
Examine which factors are related to foster children's development

T1 → T2 → T3

- T1: October 2014 (*N* = 446)

Complete cases: 342/446 = 76.7%

- T2: April 2015

- T3: October 2015 (*N* = 156)

Attrition: 100-(156/431) = 63.9%


Example 2:
FIML estimation
Example 1:
Multiple imputation
Today:
Two examples of modern strategies to handle missing data

Today's Outline:

1. My PhD study
2. Missing data: an introduction
3. Two examples
4. Practical guidelines
5. Summary & Discussion

Missing data: an introduction

- Causes
- Consequences of missing data
- Missing data mechanisms:
  - MCAR
  - MAR
  - MNAR

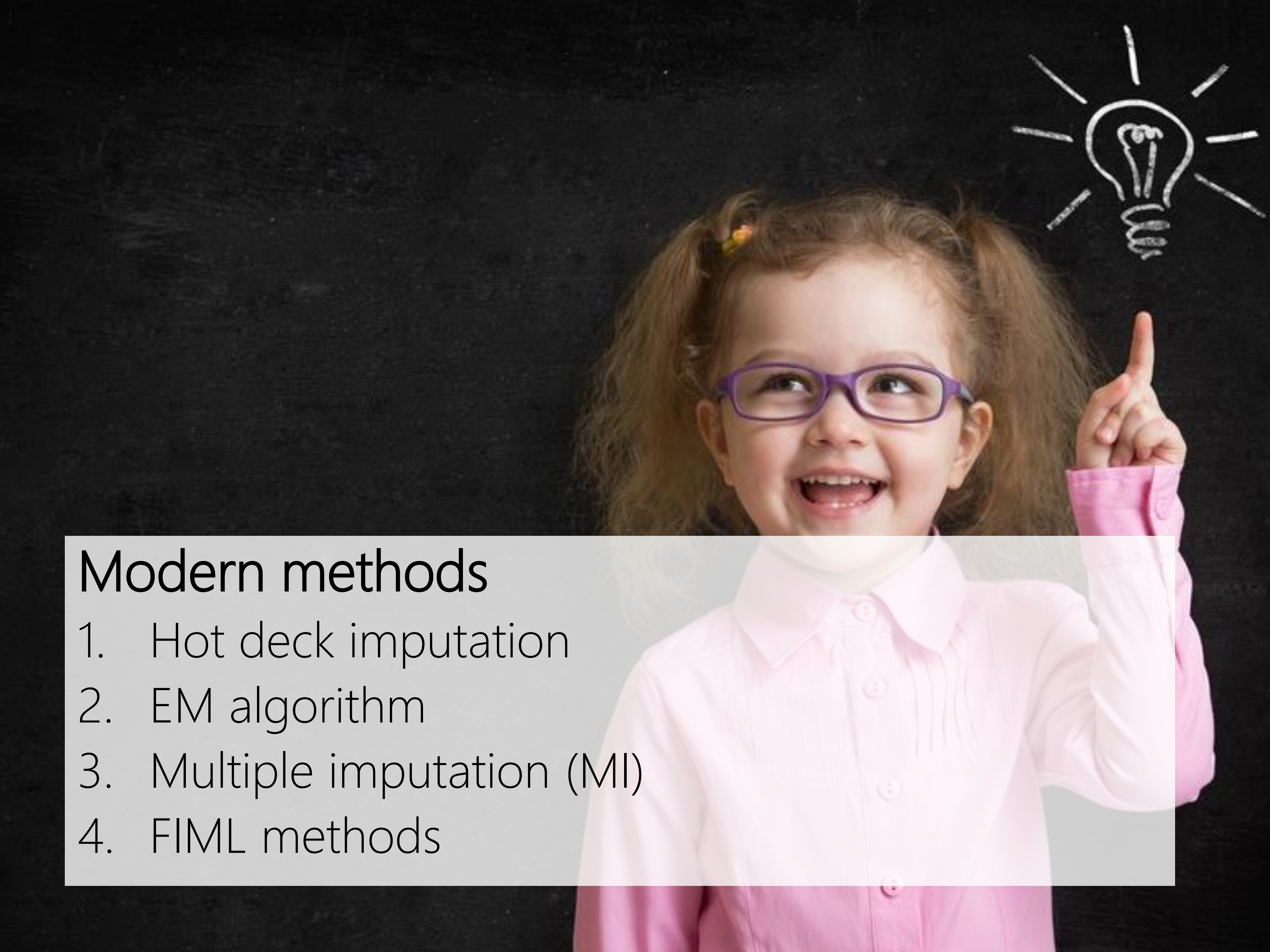
  Check this with Little's MCAR test: SPSS > Analyze > MVA / and other methods
- Ways to handle missing data
  - Traditional/simple methods – assumption: MCAR
  - Modern strategies – assumption: MCAR/MAR

Traditional/Simple methods

1. Listwise deletion – complete case analysis
2. Pairwise deletion – available case analysis
3. Mean substitution

Conclusions:

- All simple methods make strong and often unrealistic assumptions
- Listwise deletion is the least flawed, but very wasteful.
- Avoid and use modern techniques!



Modern methods

1. Hot deck imputation
2. EM algorithm
3. Multiple imputation (MI)
4. FIML methods

Today's Outline:

1. My PhD study
2. Missing data: an introduction
3. Two examples
4. Practical guidelines
5. Summary & Discussion

Example 1: Overview

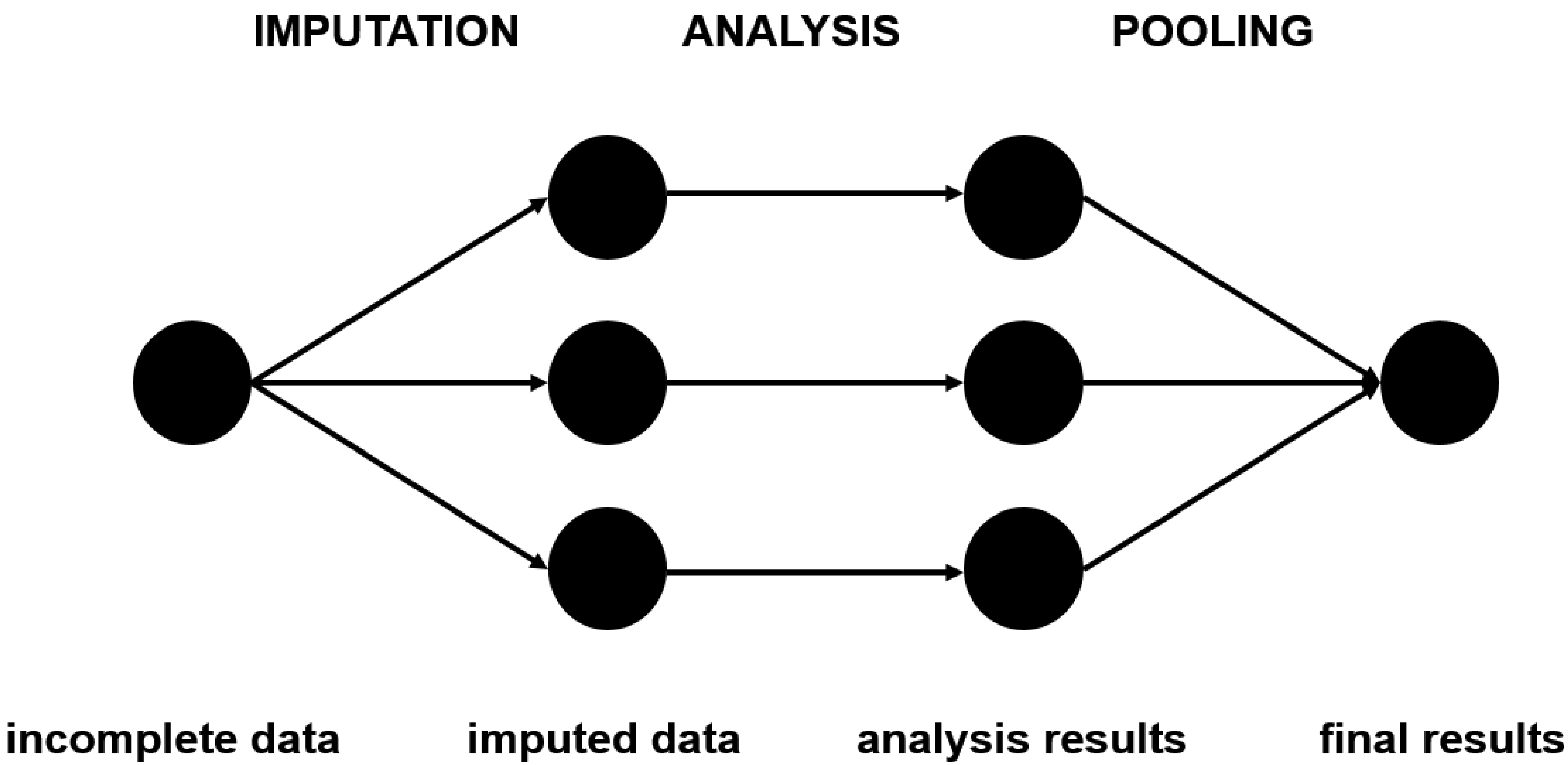
Research question	Which factors are related to foster children's psychosocial functioning? (Goemans et al., 2016)
Data	Wave I
Analysis	Hierarchical regression analysis
Software	SPSS
Type of missing data	Item nonresponse
Strategy	Multiple imputation

Example 1: Missing data

- Sample size: 446
- No more than 10% missing on each variable
- Range missing: 0-7.2%
- Mean missing: 2.0%

- Complete data for 342 (76.7%)

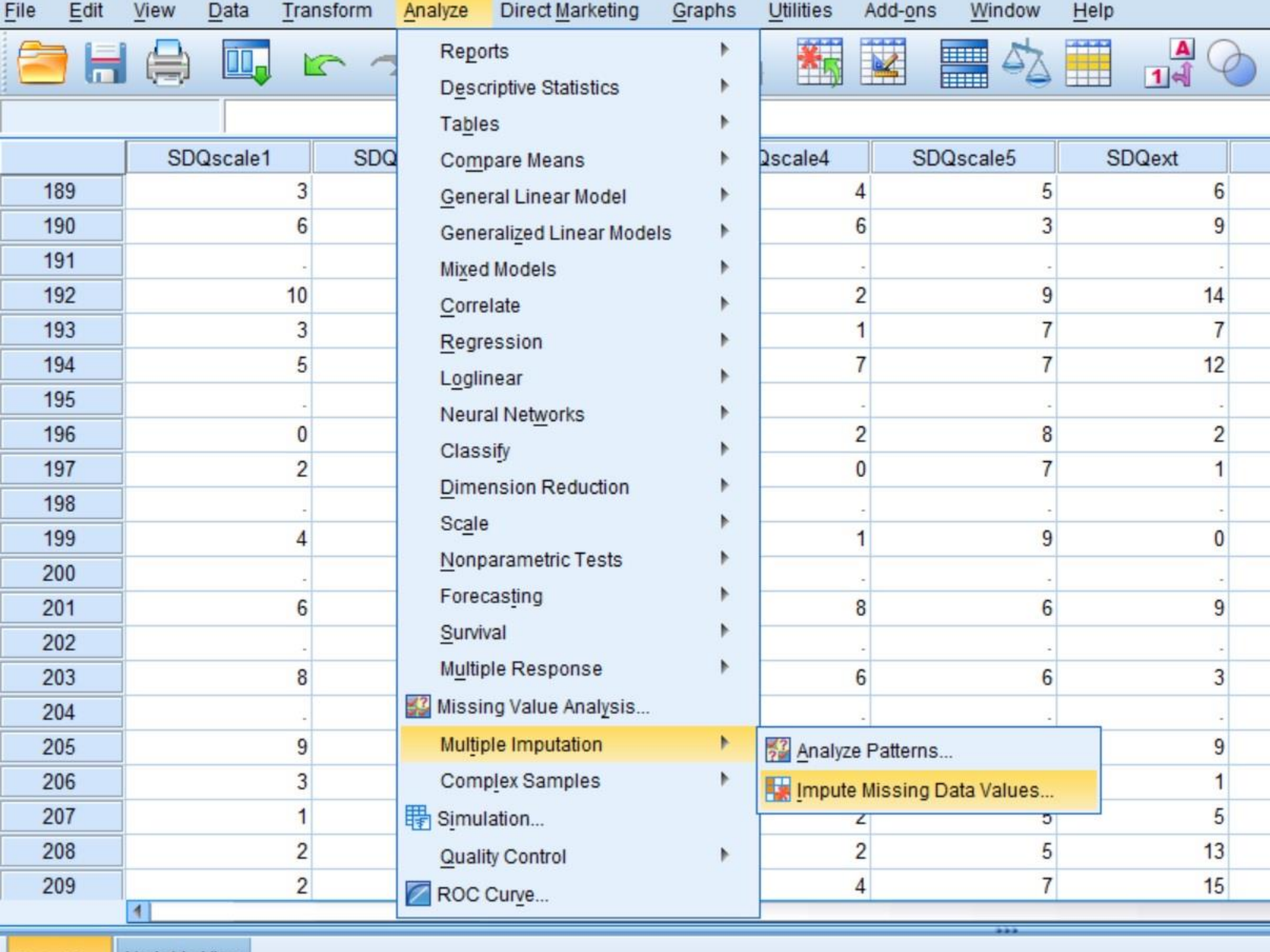
Multiple imputation



Rubin (1987)

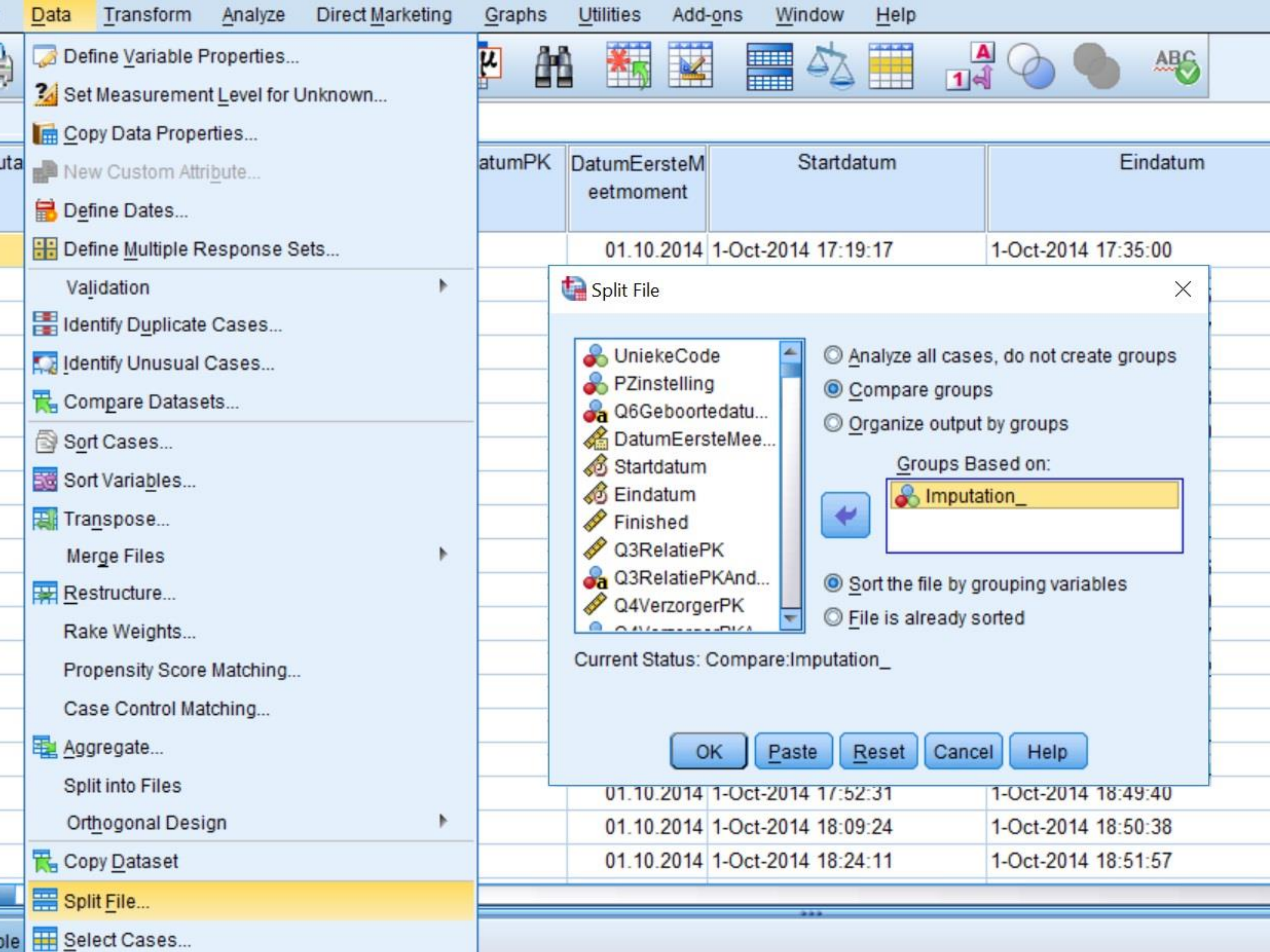
How to actually do it?

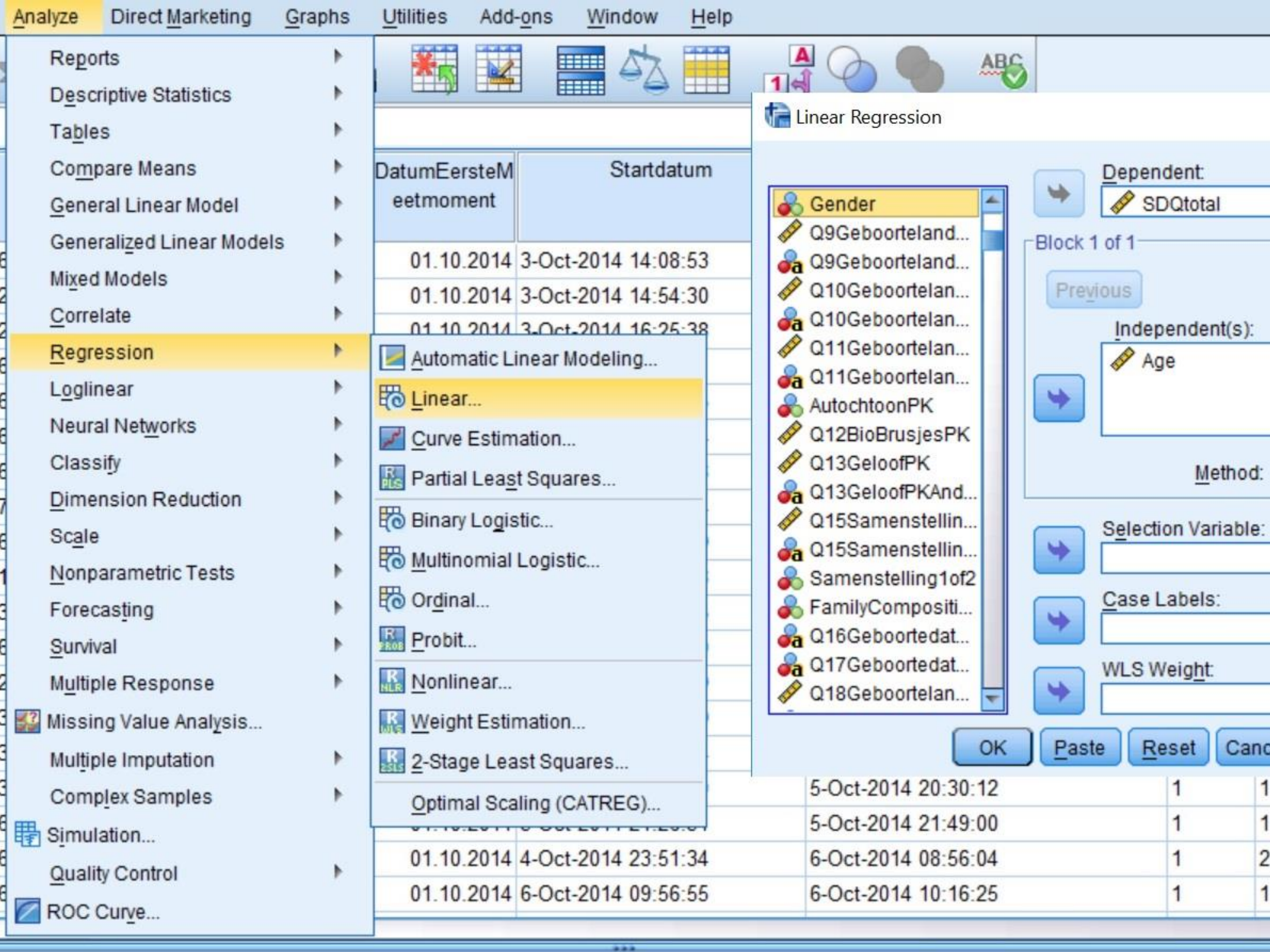
- Missing data mechanism? (MCAR or MAR?)
- SPSS: Analyze > Multiple imputation > Impute missing data values
  - Some suggestions:
    - Variables tab: use all variables, also DV (more reliable), 20 imputations (Graham et al., 2007)
    - Method tab: custom (MCMC), max iterations: 100, model type: dependent on multivariate normality


File
Edit
View
Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Analyze Patterns...
Impute Missing Data Values...
SDQscale1
SDQscale5
SDQext

	Imputation_	UniekeCode	PZinstelling	Q6GeboortedatumPK	DatumEersteMeetmoment	Startdatum
1	0	1	3	05.11.2009	01.10.2014	1-Oct-2014 17:19:17
2	0	2	7	06.08.1999	01.10.2014	1-Oct-2014 17:10:42
3	0	3	3	03.06.2004	01.10.2014	1-Oct-2014 17:18:43
4	0	4	7	27.09.2002	01.10.2014	1-Oct-2014 17:16:04
5	0	5	6	25.08.2002	01.10.2014	1-Oct-2014 17:10:25
6	0	6	6	23.09.2009	01.10.2014	1-Oct-2014 17:15:40
7	0	7	6	08.10.2008	01.10.2014	1-Oct-2014 17:17:00
8	0	9	6	29.07.2004	01.10.2014	1-Oct-2014 17:31:06
9	0	10	2	29.09.2014	01.10.2014	1-Oct-2014 17:11:59
10	0	11	6	18 08 2004	01.10.2014	1-Oct-2014 17:25:51
11	0	13	3	11.02.2000	01.10.2014	1-Oct-2014 17:29:29
12	0	14	2	10.08.2009	01.10.2014	1-Oct-2014 17:07:35
13	0	15	7	10.07.2000	01.10.2014	1-Oct-2014 17:45:58
14	0	17	3	27.03.2000	01.10.2014	1-Oct-2014 18:21:38
15	0	18	7	30.06.1997	01.10.2014	1-Oct-2014 18:02:31
16	0	19	3	15.10.2005	01.10.2014	1-Oct-2014 18:06:43
17	0	20	7	05.05.1998	01.10.2014	1-Oct-2014 17:52:31
18	0	21	2	02.12.2002	01.10.2014	1-Oct-2014 18:09:24
19	0	22	7	01.05.1999	01.10.2014	1-Oct-2014 18:24:11

Data View Variable View


Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Merge Files
Restructure...
Rake Weights...
Propensity Score Matching...
Case Control Matching...
Aggregate...
Split into Files
Orthogonal Design
Copy Dataset
Split File...
Select Cases...
atumPK
DatumEersteMeetmoment
Startdatum
Eindatum
01.10.2014
1-Oct-2014 17:19:17
1-Oct-2014 17:35:00
Split File
UniekeCode
PZinstelling
Q6Geboortedatu...
DatumEersteMee...
Startdatum
Eindatum
Finished
Q3RelatiePK
Q3RelatiePKAnd...
Q4VerzorgerPK
Analyze all cases, do not create groups
Compare groups
Organize output by groups
Groups Based on:
Imputation_
Sort the file by grouping variables
File is already sorted
Current Status: Compare:Imputation_
OK
Paste
Reset
Cancel
Help
01.10.2014
1-Oct-2014 17:52:31
1-Oct-2014 18:49:40
01.10.2014
1-Oct-2014 18:09:24
1-Oct-2014 18:50:38
01.10.2014
1-Oct-2014 18:24:11
1-Oct-2014 18:51:57

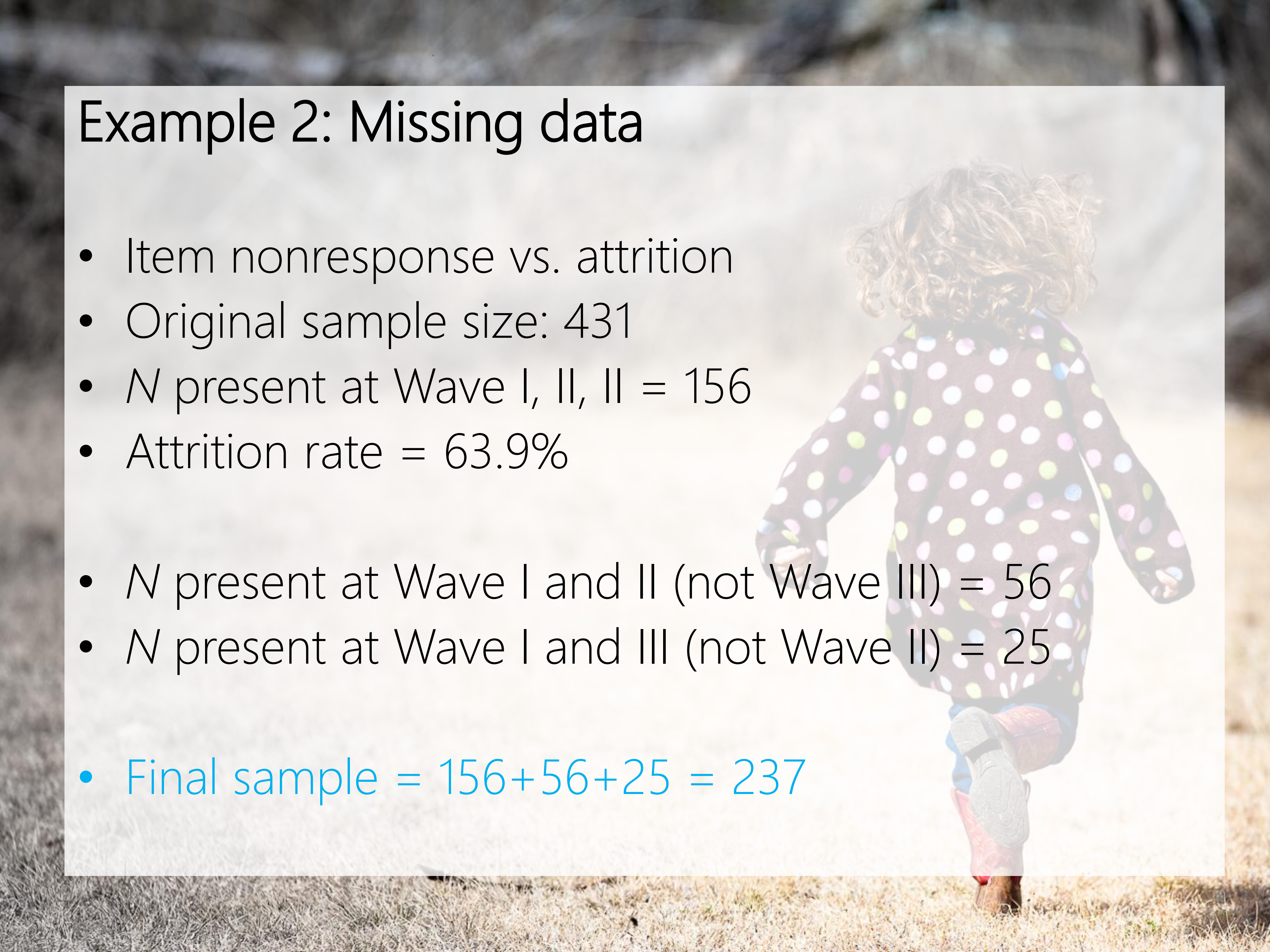

Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...
DatumEersteMeetmoment
Startdatum
01.10.2014 3-Oct-2014 14:08:53
01.10.2014 3-Oct-2014 14:54:30
01.10.2014 4-Oct-2014 23:51:34 6-Oct-2014 08:56:04 1 2
01.10.2014 6-Oct-2014 09:56:55 6-Oct-2014 10:16:25 1 1
5-Oct-2014 20:30:12 1 1
5-Oct-2014 21:49:00 1 1
Linear Regression
Gender
Q9Geboorteland...
Q9Geboorteland...
Q10Geboortelan...
Q10Geboortelan...
Q11Geboortelan...
Q11Geboortelan...
AutochtoonPK
Q12BioBrusjesPK
Q13GeloofPK
Q13GeloofPKAnd...
Q15Samenstellin...
Q15Samenstellin...
Samenstelling1of2
FamilyCompositi...
Q16Geboortedat...
Q17Geboortedat...
Q18Geboortelan...
Dependent:
SDQtotal
Block 1 of 1
Previous
Independent(s):
Age
Method:
Selection Variable:
Case Labels:
WLS Weight:
OK
Paste
Reset
Canc

Coefficients[a]

Imputation Number	Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	M
Original data	1	(Constant)	12,861	,862		14,912	,000	
		Leeftijd numeriek	-,046	,076	-,027	-,606	,545	
1	1	(Constant)	12,751	,859		14,849	,000	
		Leeftijd numeriek	-,038	,076	-,023	-,501	,617	
2	1	(Constant)	12,769	,859		14,863	,000	
		Leeftijd numeriek	-,040	,076	-,024	-,523	,601	
3	1	(Constant)	12,799	,860		14,888	,000	
		Leeftijd numeriek	-,042	,076	-,025	-,560	,576	
18	1	(Constant)	12,849	,862		14,902	,000	
		Leeftijd numeriek	-,047	,076	-,028	-,621	,535	
19	1	(Constant)	12,766	,859		14,856	,000	
		Leeftijd numeriek	-,039	,076	-,023	-,520	,604	
20	1	(Constant)	12,774	,858		14,881	,000	
		Leeftijd numeriek	-,040	,076	-,024	-,530	,596	
Pooled	1	(Constant)	12,806	,861		14,870	,000	
		Leeftijd numeriek	-,043	,076		-,568	,570	

Example 2: Overview

Research question	Are there transactional relations between foster children's internalizing and externalizing behaviors and foster parents' stress?
Data	Wave I, II, III
Analysis	Structural Equation Modeling (SEM)
Software	EQS
Type of missing data	Attrition (wave nonresponse)
Strategy	FIML estimation



Example 2: Missing data

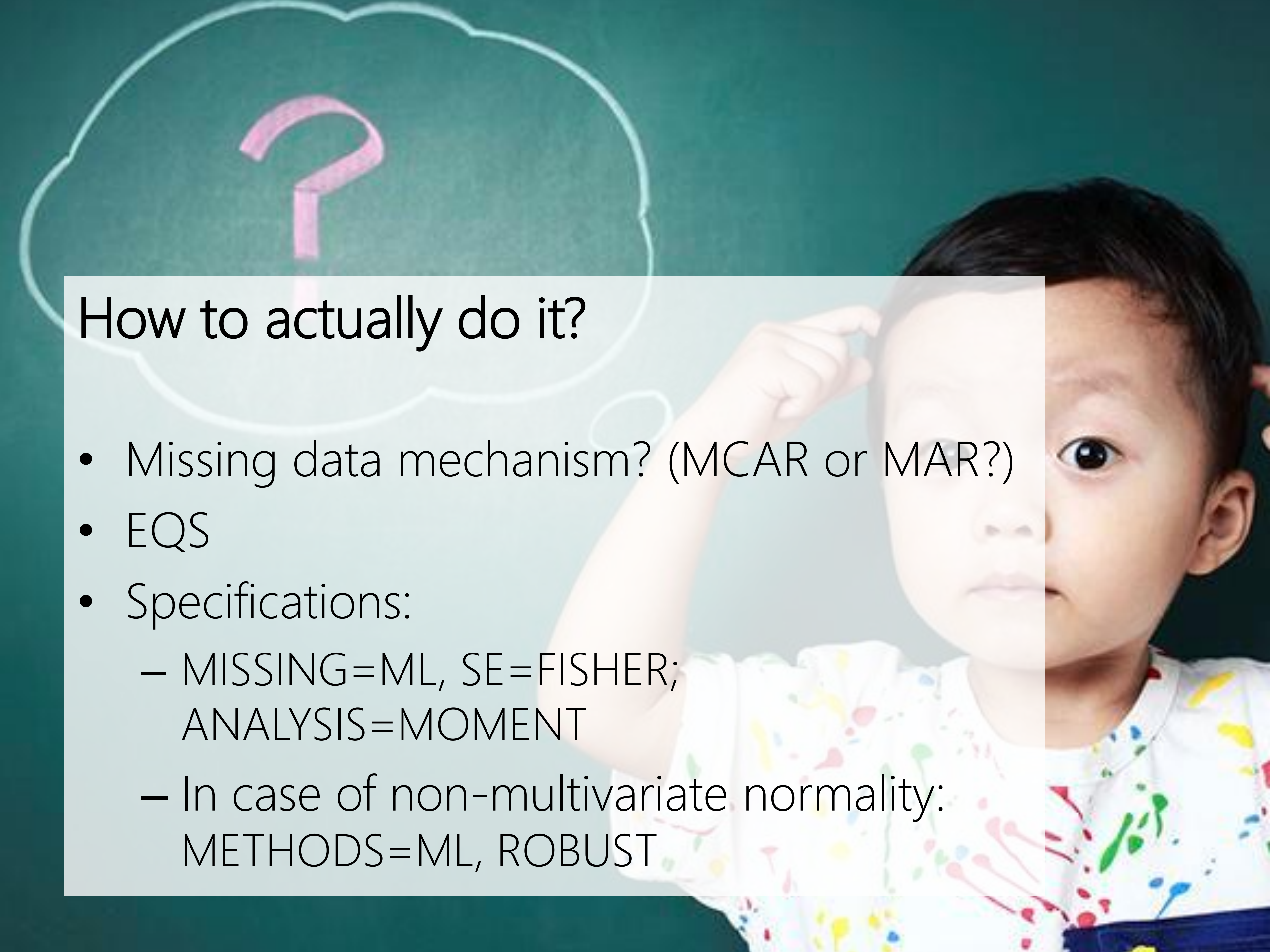
- Item nonresponse vs. attrition
- Original sample size: 431
- *N* present at Wave I, II, II = 156
- Attrition rate = 63.9%

- *N* present at Wave I and II (not Wave III) = 56
- *N* present at Wave I and III (not Wave II) = 25

- Final sample = 156+56+25 = 237

FIML estimation:

- Does not impute data
- Use all available information to estimate parameter values and standard errors
- Available in the SEM software packages



How to actually do it?

- Missing data mechanism? (MCAR or MAR?)
- EQS
- Specifications:
  - MISSING=ML, SE=FISHER; ANALYSIS=MOMENT
  - In case of non-multivariate normality: METHODS=ML, ROBUST

Today's Outline:

1. My PhD study
2. Missing data: an introduction
3. Two examples
4. Practical guidelines
5. Summary & Discussion

Suggestions for making it work:

- Books and articles
- YouTube tutorials
- Read other articles with the same study design / in which the method is described

- Take your time!
- Trial and error
- Ask for help

Further reading:

- Baraldi & Enders (2010). An introduction to modern missing data analyses.
- Graham (2009). Missing data analysis: making it work in the real world.
- Jackson et al. (2012). Strategies for longitudinal research with youth in foster care: A demonstration of methods, barriers, and innovations.
- Jelicic et al. (2009) – Use of missing data methods in longitudinal studies: the persistence of bad practices in developmentla psychology.
- Peeters et al. (2015). How to handle missing data: A comparison of different approaches.

Guidelines for reporting:

- What did you do to prevent missing data?
- How much missing data do you have?
- What is the missing data mechanism?
- How did you handle the missing data?

Burton & Altman (2004); Jelicic et al. (2009); Peeters et al. (2014), Peugh & Enders (2004); Schlomer et al. (2010)

Today's Outline:

1. My PhD study
2. Missing data: an introduction
3. Two examples
4. Practical guidelines
5. Summary & Discussion

PROS	CONS
• MI and FIML best methods currently available • General methods	• More work • More complex?

Developmental Psychology
2009, Vol. 45, No. 4, 1195–1199

Use of Missing Data Methods in Longitudinal Studies: The Persistence of Bad Practices in Developmental Psychology

“Modern” Missing Data Analysis Methods

… wish is that 10 years … applications.”

However, the ML and MI methods yield more valid results than listwise and pairwise deletion approaches and, therefore, should become part of the developmental scientist’s analytical tool kit.

Developmental Psychology
2009, Vol. 45, No. 4, 1195–1199

… applications of maximum likelihood and multiple imputation are appearing with greater frequency in published research articles, a substantial gap still exists between the procedures that the methodological literature recommends and those that are actually used in the applied research studies

these method …

data so that a …

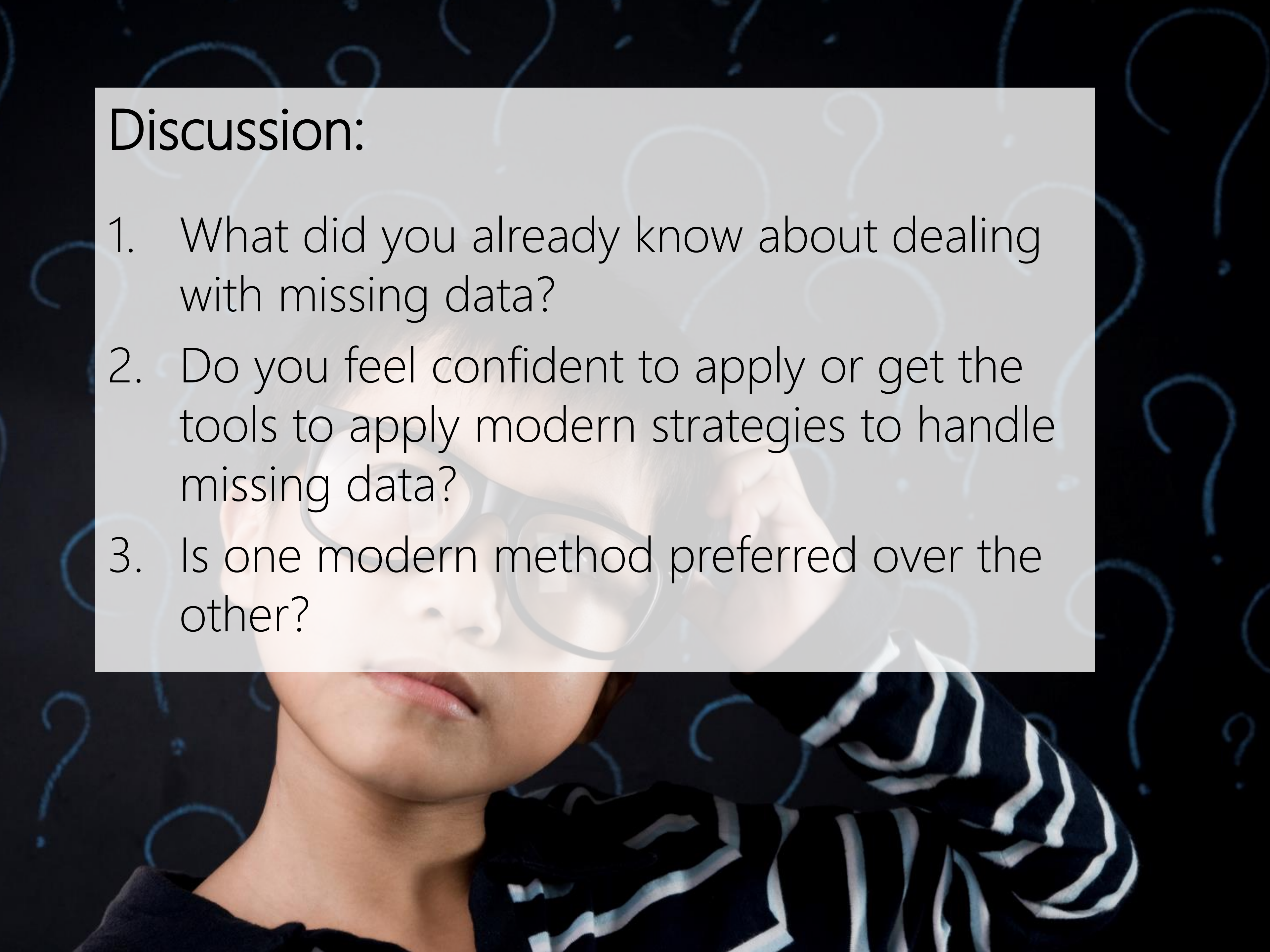
John W. Gra…

J.W. Graham, *Missing Da…*

Vote again!

A. I will only have a small number of missing data, so I will not deal with this missing data
B. Pairwise deletion, listwise deletion or mean imputation
C. Multiple imputation or FIML estimation
D. I don't know yet
E. Not applicable. I don't have / will not have missing data at all

www.menti.com; Code: 94 74 33



Discussion:

1. What did you already know about dealing with missing data?
2. Do you feel confident to apply or get the tools to apply modern strategies to handle missing data?
3. Is one modern method preferred over the other?

Modern Strategies to Handle Missing Data:

A Showcase of Research on Foster Children

Thank you for your attention!

Anouk Goemans

Email: a.goemans@fsw.leidenuniv.nl

Universiteit Leiden

מכון חרוב (ע"ר)

The Haruv Institute (R.A.)