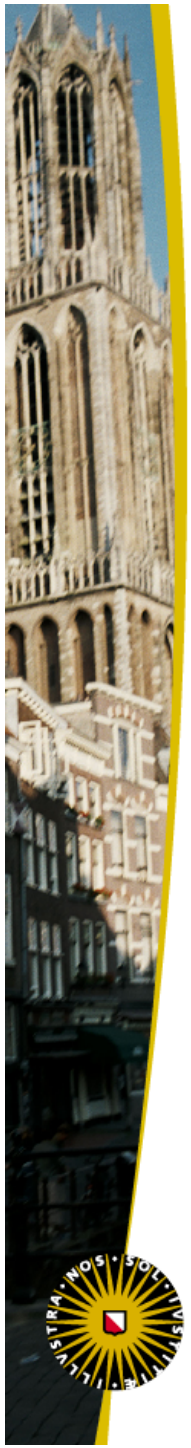


Understanding longitudinal nonresponse in order to correct for it

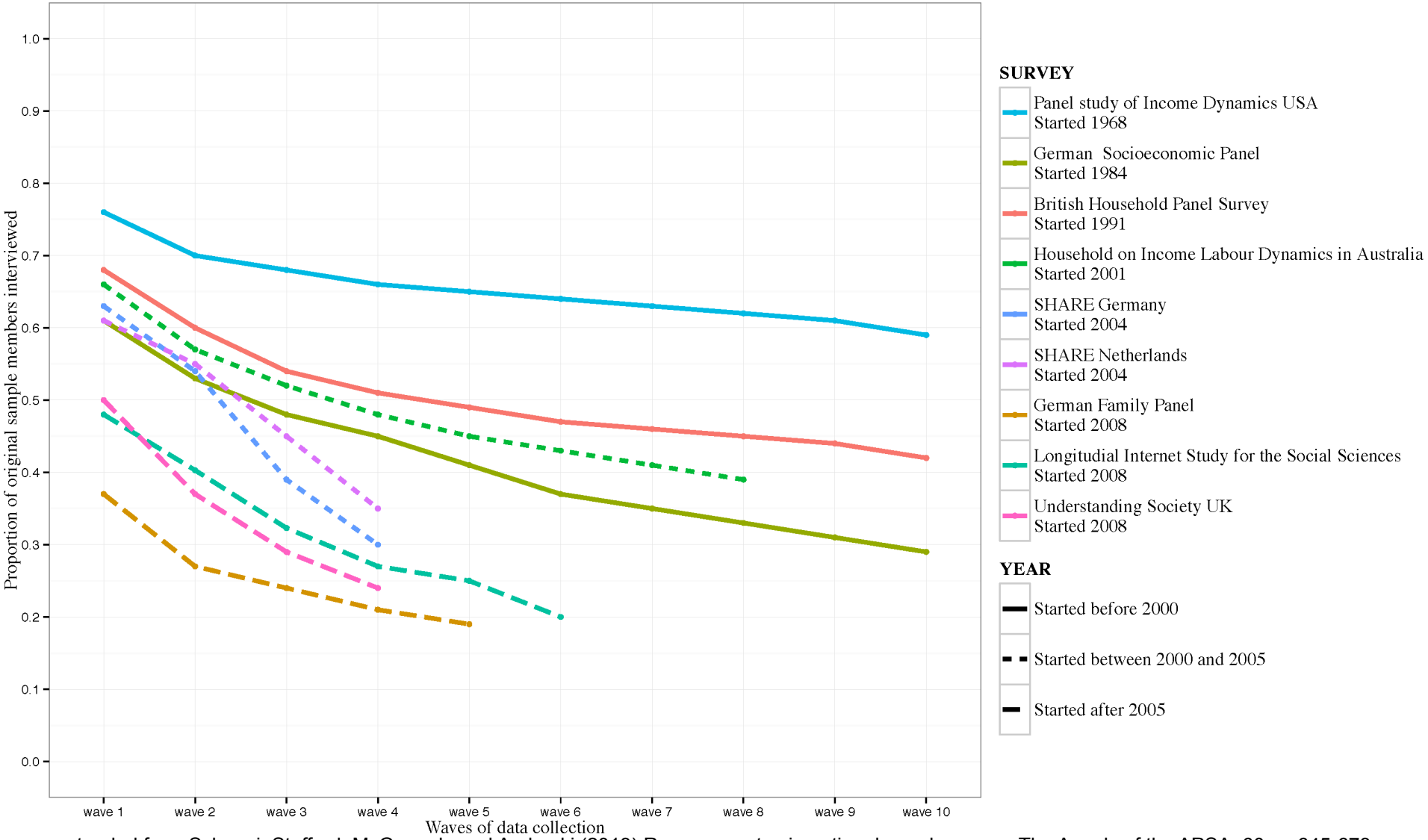
Peter Lugtig
Associate professor in survey methodology
Department of methods and statistics – Utrecht University
www.peterlugtig.com
p.lugtig@uu.nl

Universiteit Utrecht

8 May 2017

Attrition in national panel surveys

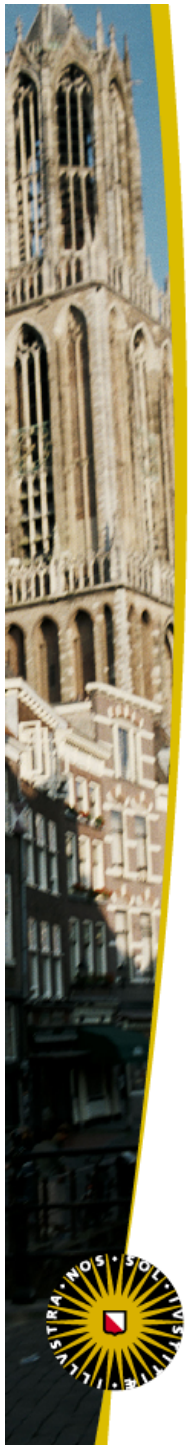
The recent increase in dropout in longitudinal surveys



extended from Schoeni, Stafford, McGonagle and Andreski (2013) Response rates in national panel surveys. The Annals of the APSA, 60, p. 645-673

Outline of talk

1. How can we understand longitudinal nonresponse?
   - BHPS: Face-to-Face household Panel
   - LISS: Probability-based Internet panel
2. How to compensate?
   - Weighting, FIML or imputation?
   - Selecting covariates
   - Blocked imputation approach
3. Future directions



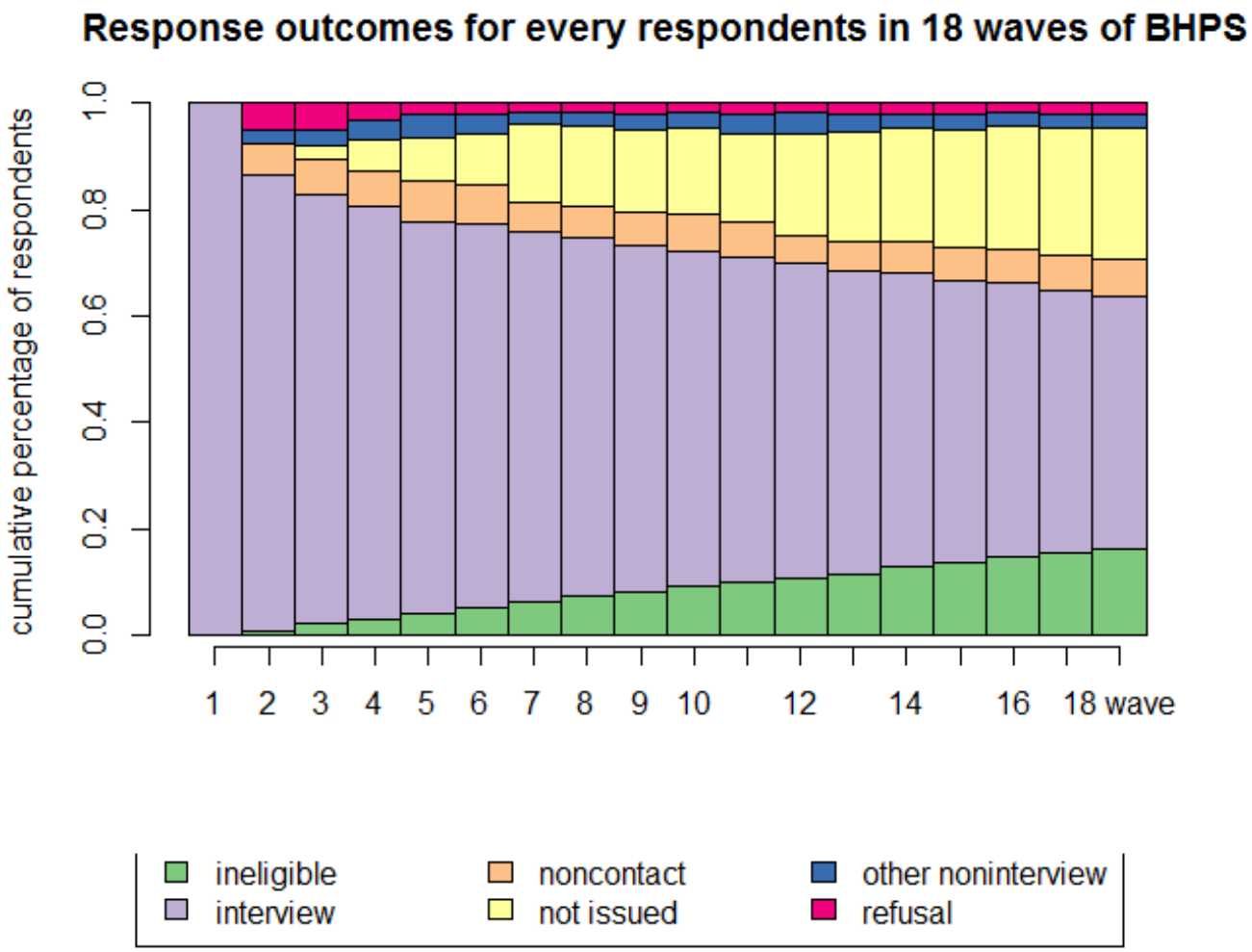
1. Understanding attrition BHPS

Universiteit Utrecht

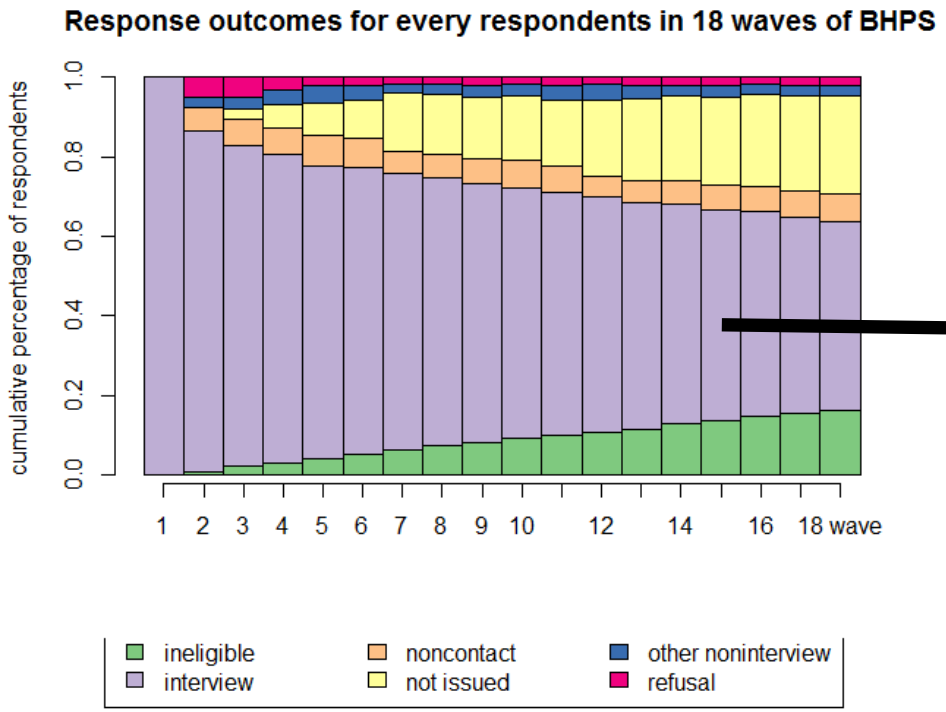
ISER

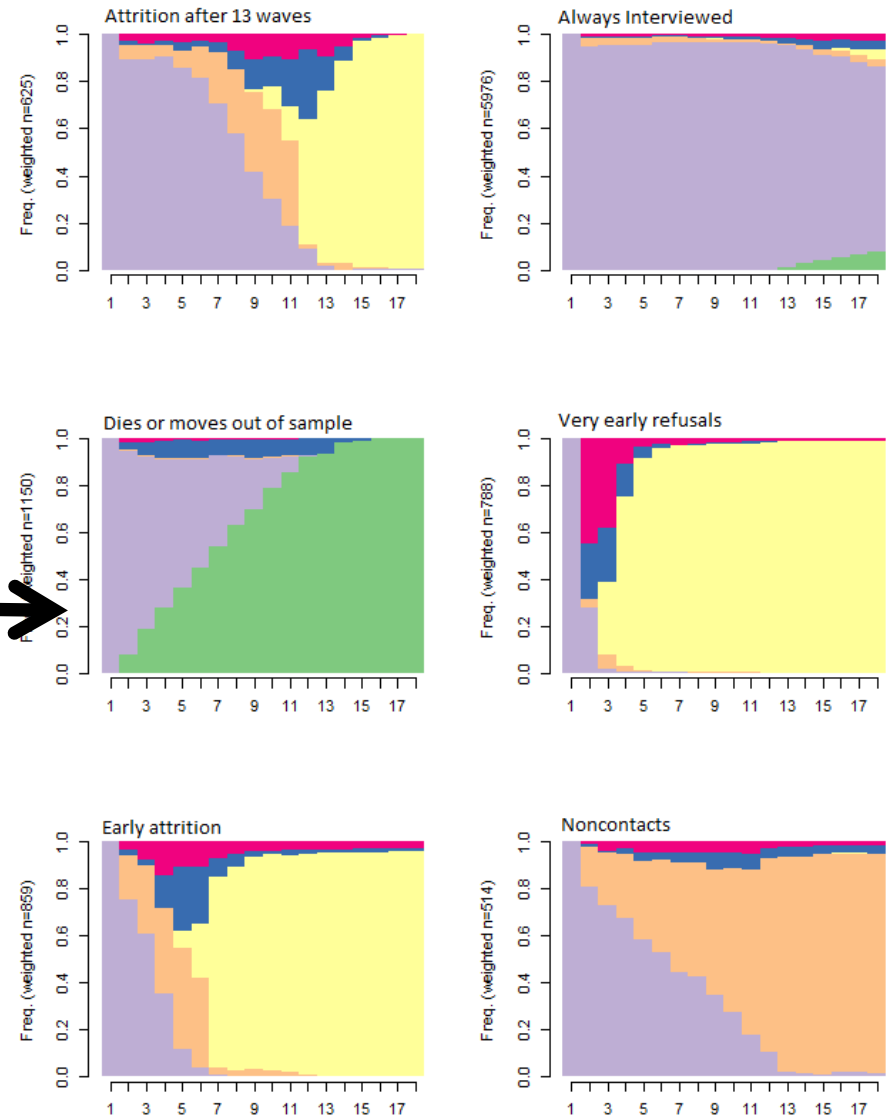
British Household Panel Survey

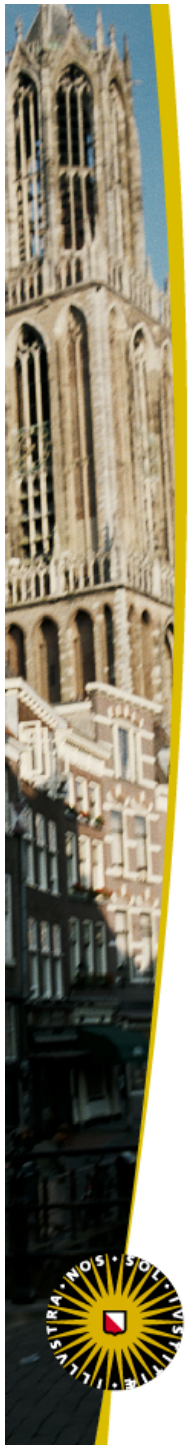
- Face-to-Face panel, started in 1991

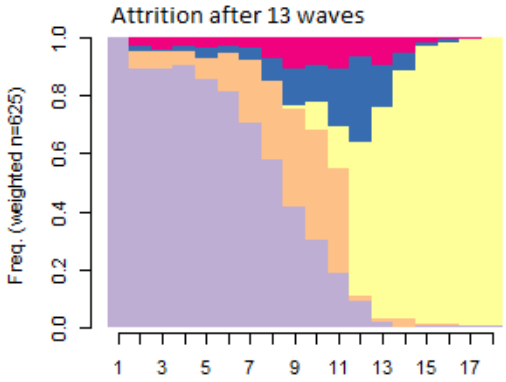


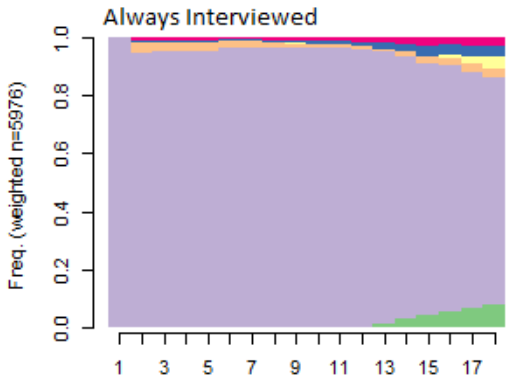
Sequence cluster analysis



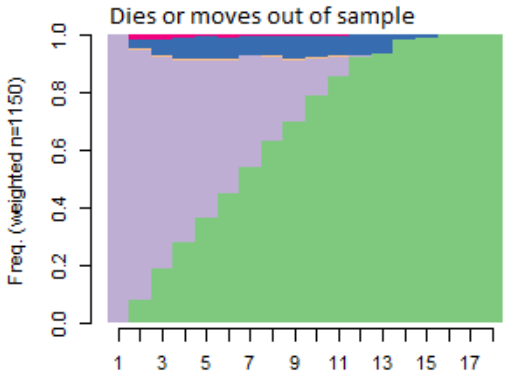


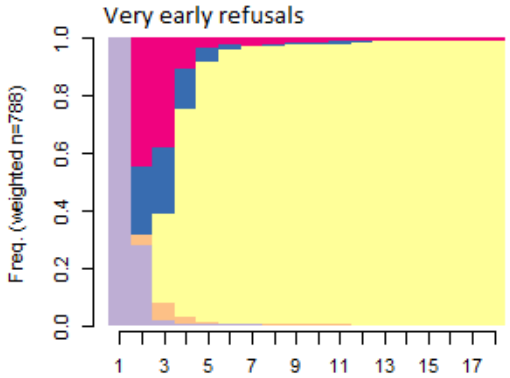




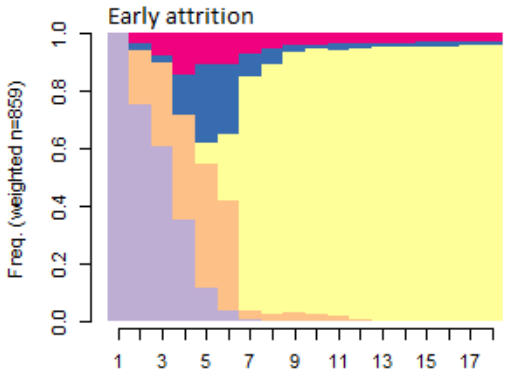


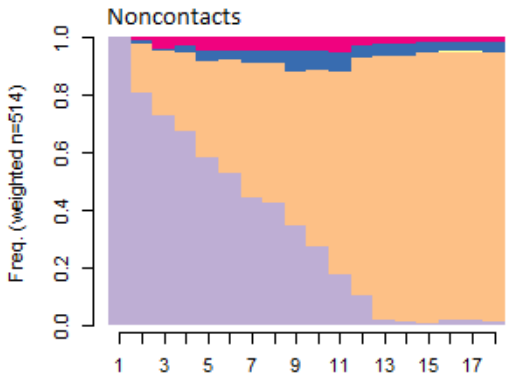
Attrition after 13 waves
7%
Always interviewed
65%





Dies
13%
Very early refusals
9%

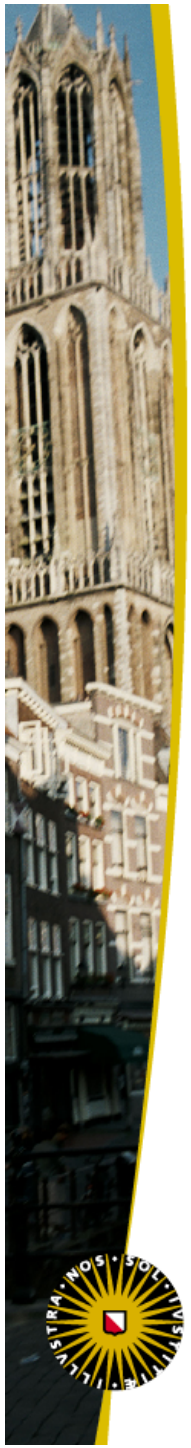




Early attrition
9%
Noncontacts
6%

Classifying attrition types is useful

- 1. Learn about the process of dropout
  - Noncontact often precedes a refusal or 'other non-interview'
  - About 6% of respondents gets 'lost'
- 2. Predict attrition in every group separately
  - People who drop out early are probably different to late dropouts
  - People who die are different from others
- 3. Identify groups that need special attention
  - You cannot prevent people dying
  - But can try to do something about the noncontacts
    - Adaptive designs (Lynn, 2014, 2016)



1. Understanding attrition LISS panel

Universiteit Utrecht

ISER

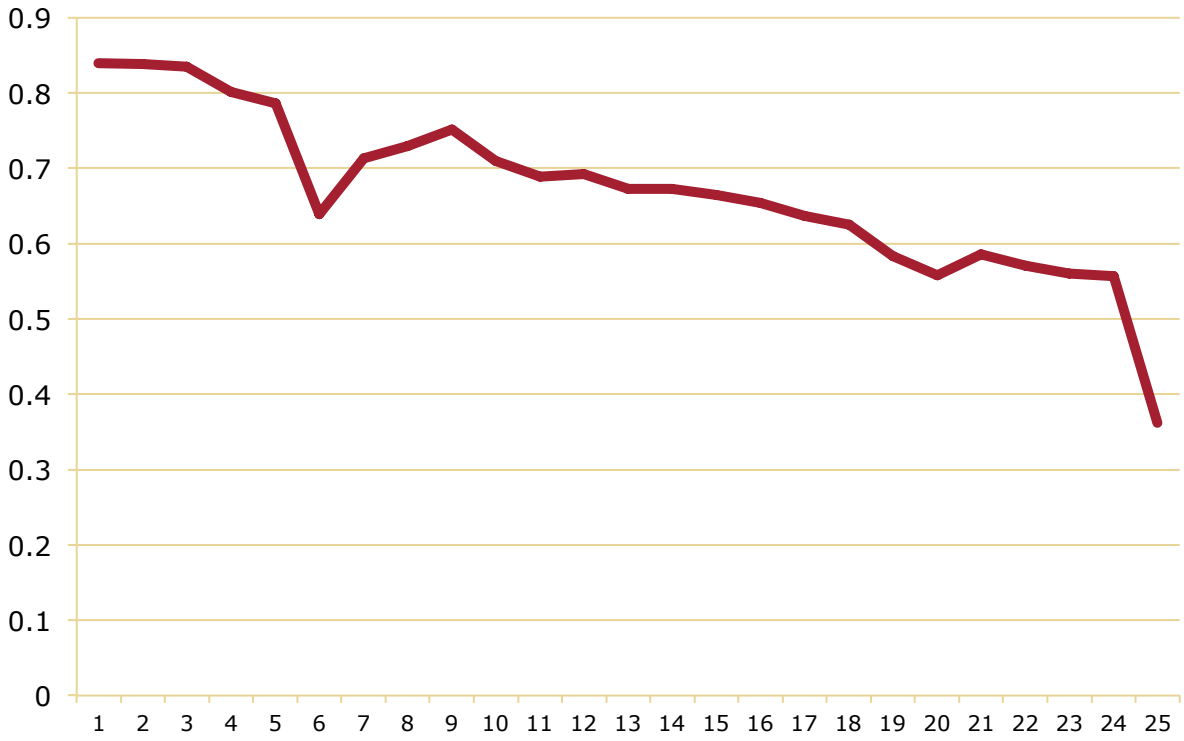
The LISS panel

- LISS Panel
  - Internet household panel
  - Random sample of Dutch population
  - 10.000 respondents in 2007
  - Internet access provided
  - January 2012, about 5.000 still ‘active’

- Monthly surveys (see www.lissdata.nl)
  - Psychological
  - Social
  - Economical
  - Bio-medical measures

LISS overall pattern

- No distinction between types of attrition
- Attrition is actually not monotone

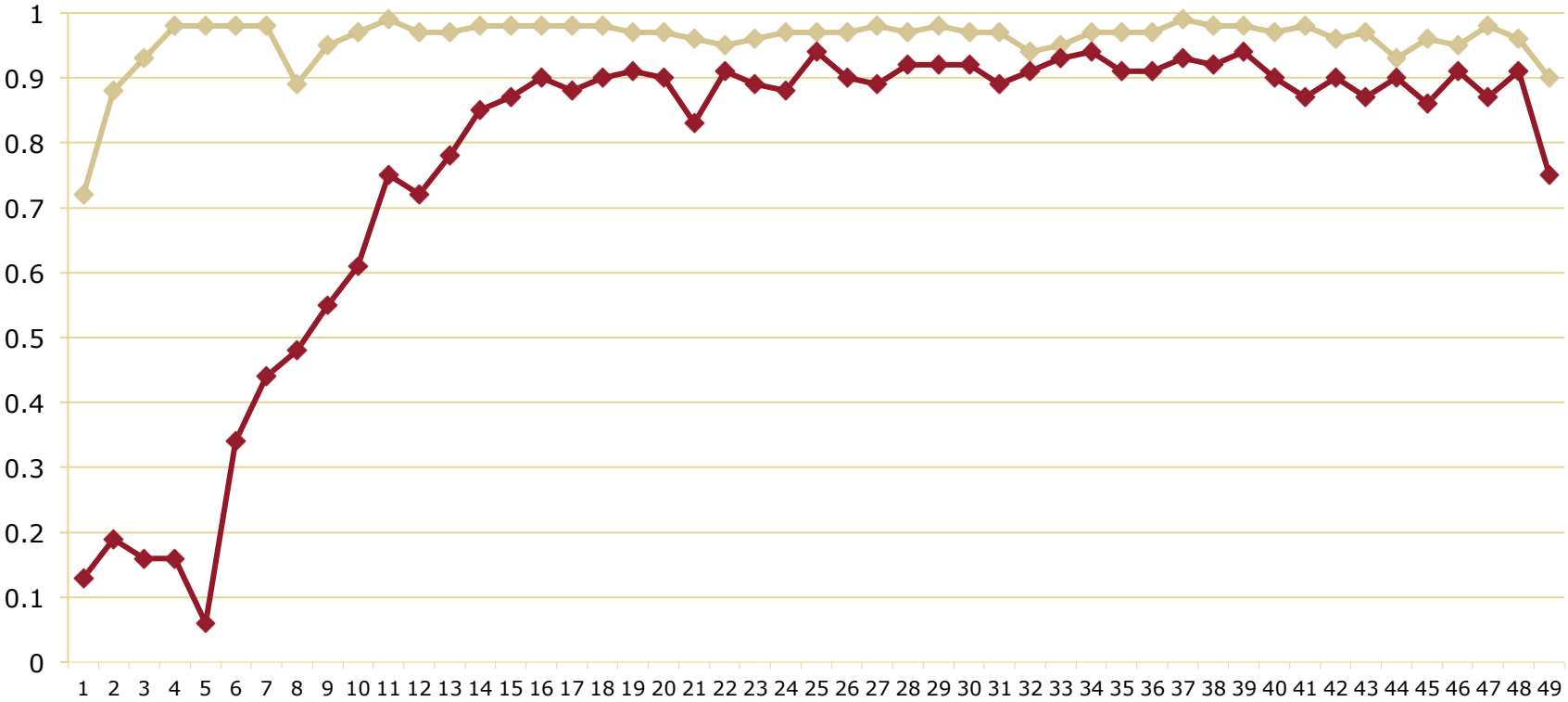


Universiteit Utrecht

Non-monotone attrition

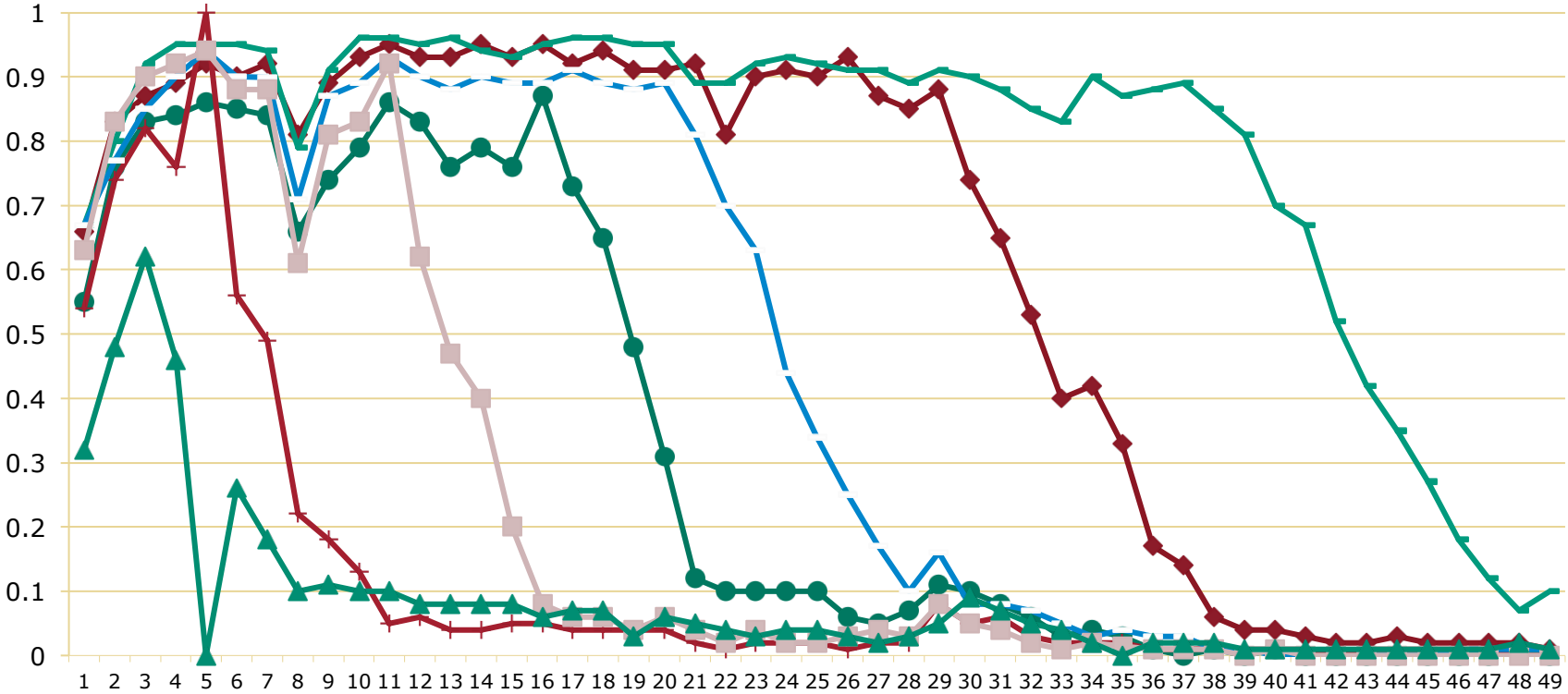
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	.....	48
X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	.....	x
X	x	x	x	x	x	x	x	x	x	x	x	x					.....	
X	x	x	x	x	x	x	x										.....	
X	x	x	x	x	x		x	x	x	x	x	x	x	x		x	.....	
X	x	x	x	x	x	x	x				x	x	x	x	x	x	.....	x
X	x	x	x		x		x	x	x		x	x		x	x	x	.....	
X		x			x				x	x	x	x	x	x	x	x	.....	x

LISS panel – attrition patterns



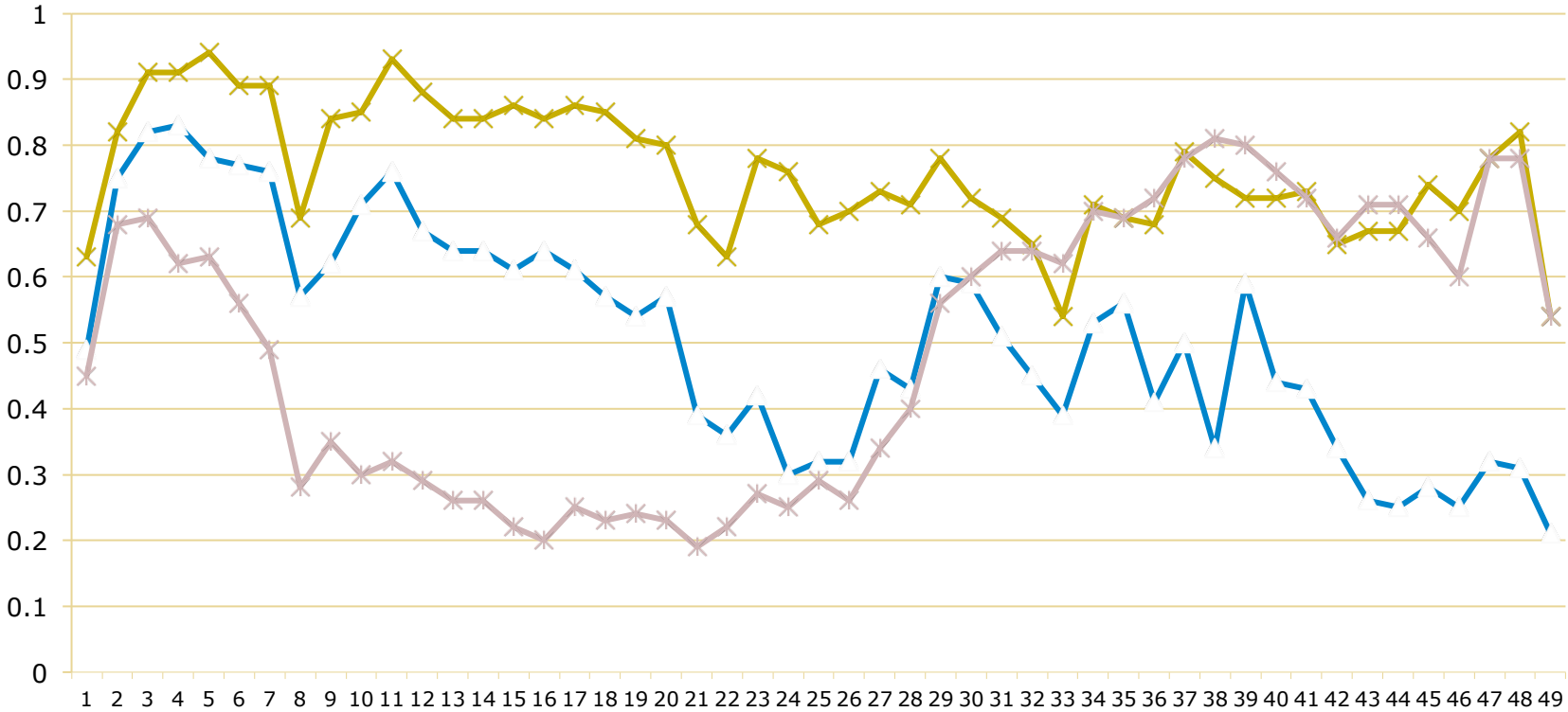
Loyal groups	**38.9% combined**
• Stayers (always)	36.0%
• Late adopters	2.9%

Results - dropout



Dropout groups	**46.7% combined**
• Dropout around wave 40	4.7 %
• Dropout around wave 32	4.3%
• Dropout around wave 24	5.1%
• Dropout around wave 18	5.6%
• Dropout around wave 12	7.2%
• Dropout around wave 6	10.9%
• Never participating group	8.9 %

Results - lurkers

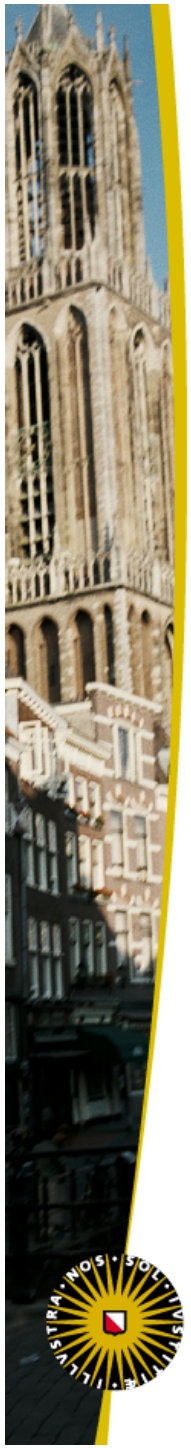


Lurker groups	**14.5% combined**
• Active lurkers	8.2 %
• Lurkers activated around wave 30	2.5%
• Low active	3.8%

LISS panel

- Attrition is both monotone and non-monotone
- Do not know why people drop out
- Implications for imputations
  - Use t1 only to predict t2?
  - Or use t1 and t3?

- BHPS and LISS panel may need different imputation procedures
  - Compensating for nonresponse in longitudinal studies should contain some ad-hoc solutions



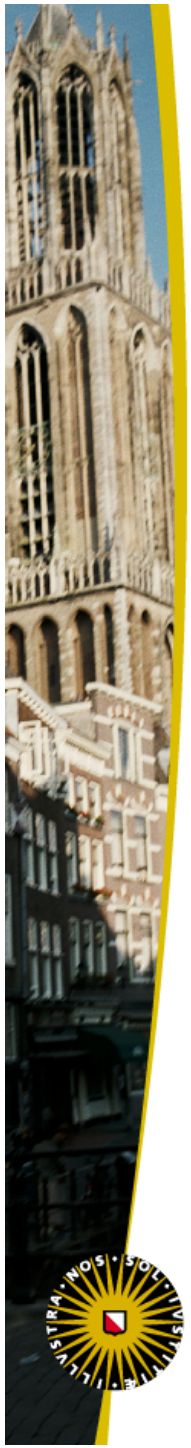
2. Compensating for missing data

Universiteit Utrecht

ISER

Compensating for longitudinal NR

- Do nothing
  - Let users decide
- Weighting
  - Household panels
  - Design-based correction model (1)
- Imputations
  - Household panels (e.g. income imputations)
  - Model-based correction model (multiple)



X (covariates)

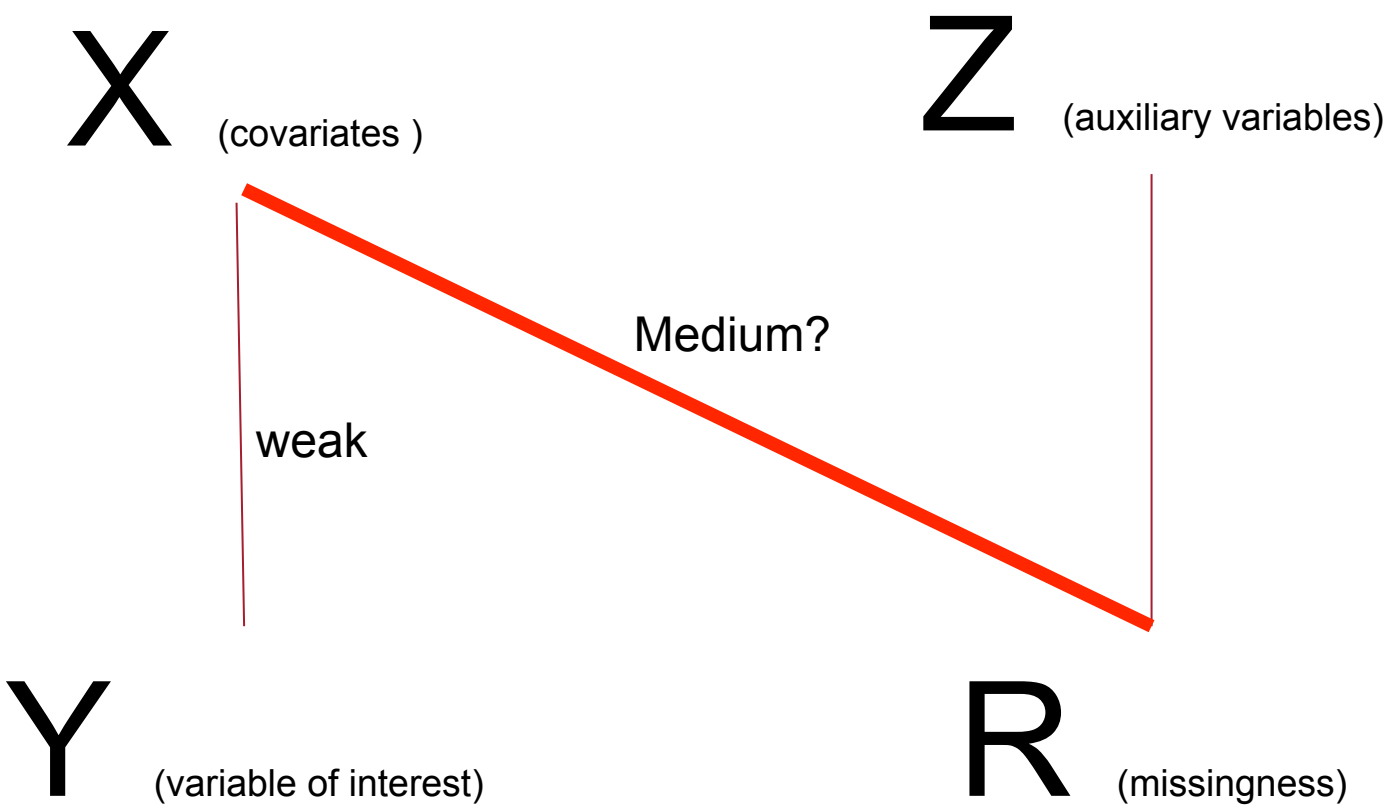
Z (auxiliary variables)

Y (variable of interest)

R (missingness)

Universiteit Utrecht

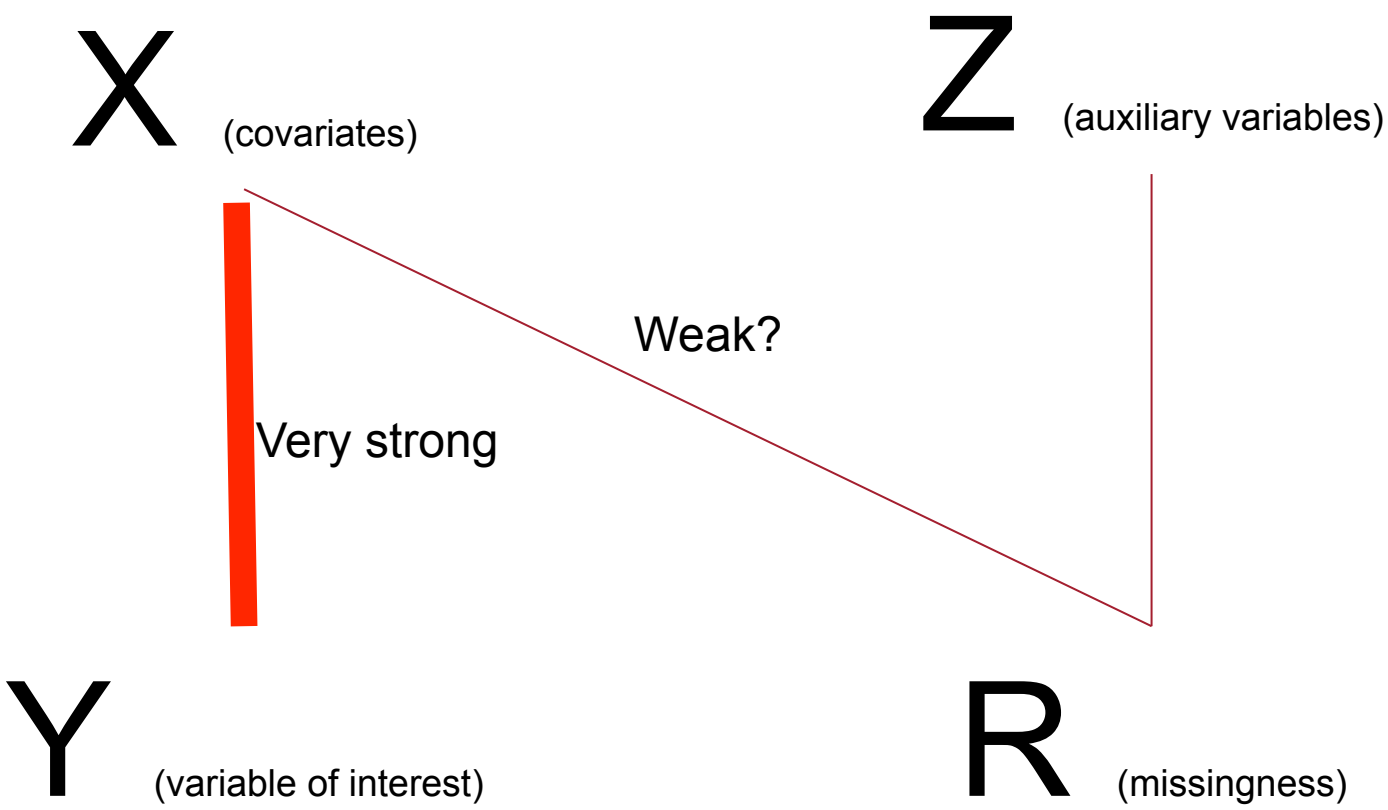
Cross sectional nonresponse adjustments using frame variables



Universiteit Utrecht

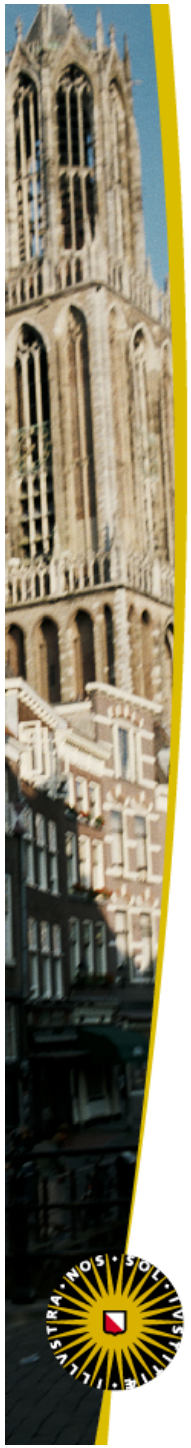
Weighting is sensible approach

Longitudinal nonresponse adjustments using earlier measures



Universiteit Utrecht

imputation is sensible approach
How weak is relation x -> R?



2. Compensating for missing data

Correlates of attrition in BHPS

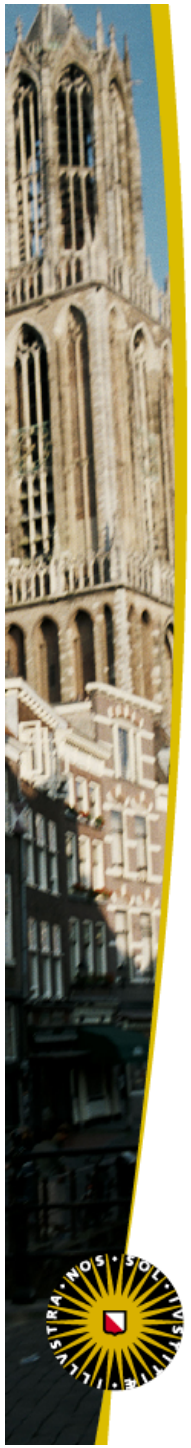
Universiteit Utrecht

ISER

Ever attrition – using wave 1 predictors

	Strongest predictors	R2 nagelkerke
Ever attrition (1=yes)	Higher degree Gender (F) Being in paid work Age Likely to move London	,09

Universiteit Utrecht



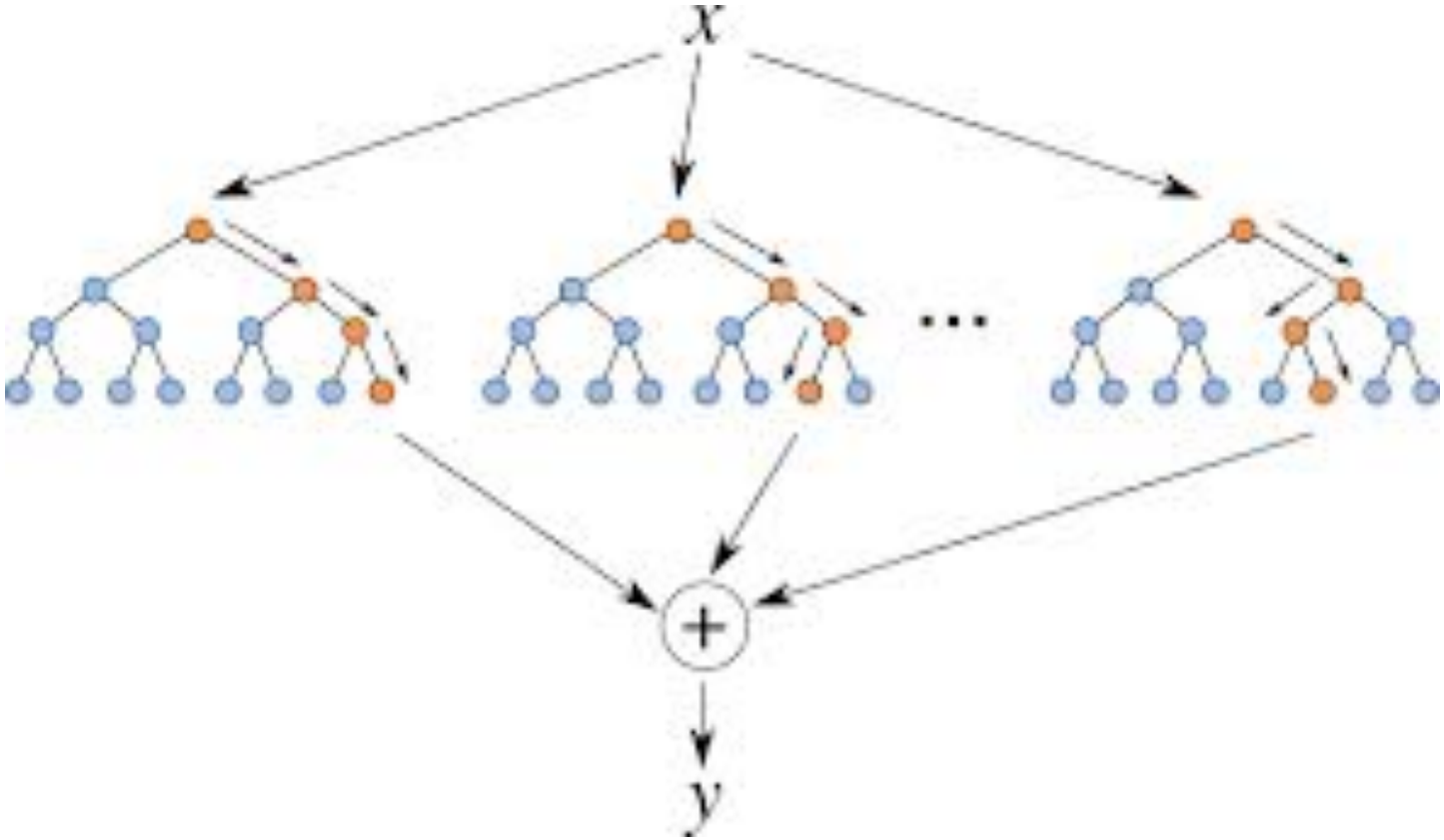
Using previous wave predictors	Strongest predictors	R2 nagelkerke
Wave 2	Age, single non-elderly, noncompletion of tracking, not cooperative, higher degree, London	.15
Wave 3	Age, gender(M), not cooperative, higher degree, London	.07
Wave 4	Not having a car, couple with children, urban areas	.05
Wave 5	Age, completion of tracking, gender (m), couple with children	.07
Wave 6	Age, poor health, London	.08
Wave 7	Age, likely to move, self completion returned	.09
Wave 8	Not cooperative, self completion returned, higher degree	.08
Wave 9	Poor health, higher degree	.05
Wave 10	Not having a car	.09
Wave 11	Likely to move, higher degree	.06
Wave 12	Completion of tracking	.09
Wave 13	Gender(m)	.07
Wave 14	Poor health, Not cooperative, self completion returned	.12
Wave 15	Age, Likely to move, self completion returned	.12
Wave 16	Couple no children, no dependent children	.15
Wave 17	self completion returned	.05
Wave 18	Age, cooperation with interviewer	.15

Type of attrition at wave 2 using wave 1 predictors

Using previous wave predictors	Strongest predictors	R2 nagelkerke
Noncontact	Age Planning to move	.22
Refusal	Not cooperative with Ivwr, Single elderly	.05
Other non-interview	Not cooperative with Ivwr,	.08
Ineligible (death)	Age, Not having a car, All single households	.37
Reference = Interview at wave 2		

RandomForest

- Algorithm from machine learning
- Very complex combinations of variables (X) explaining Y (nonresponse)



Universiteit Utrecht

Type of attrition at wave 2 using wave 1 predictors

Using previous wave predictors	GLM R2 nagelkerke	RandomForest R2 nagelkerke
Noncontact	.22	.27
Refusal	.05	.15
Other non-interview	.08	.10
Ineligible (death)	.37	.39
Reference = Interview at wave 2		

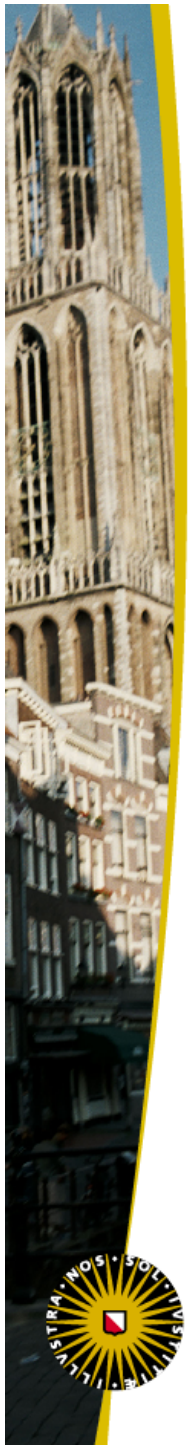
Universiteit Utrecht

Conclusions from BHPS example

- Predictive power of 'traditional models' very low
  - Relation X-> R very weak
- Imputation or FIML models should be preferred over weighting

- Machine learning algorithms can *maybe* increase explanatory power of nonresponse models.
  - But complex patterns!
  - Do not hold across waves very well
- Would imputation per 'type' of attrition help?

Skip to end



2. Compensating – Blocked imputation

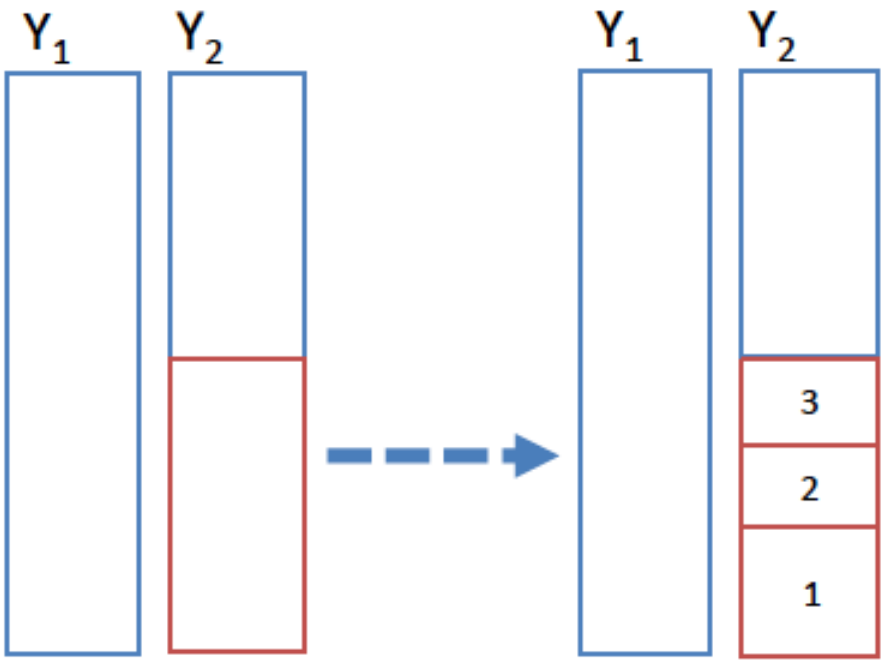
Universiteit Utrecht

ISER

Attrition – using wave 1 predictors

	Strongest predictors	R2 nagelkerke
1 model	Age, single non-elderly, noncompletion of tracking, not cooperative, higher degree, London	,09
Separated for reason Noncontact	Age Planning to move	.22
Refusal	Bad cooperation with interviewer Single elderly	.05
Other non-interview	Bad cooperation with interviewer	.08
Ineligible (death)	Age Not having a car All single households	.37
Reference = Interview at wave 2		

Imputation using 'type' of attrition



Universiteit Utrecht

How to include information on missingness?

$$L(x) = Pr(S = 1|X = x),$$

Example: Propensity to be missing due to

S=1: Death <- X

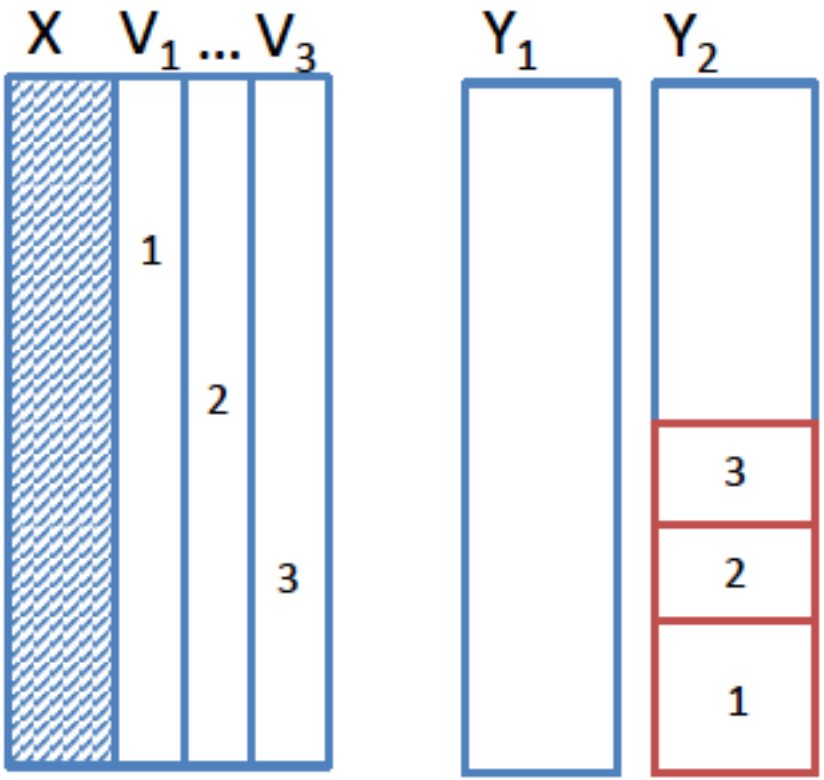
S=2: Refusal <- X

S=3: Noncontact <- X

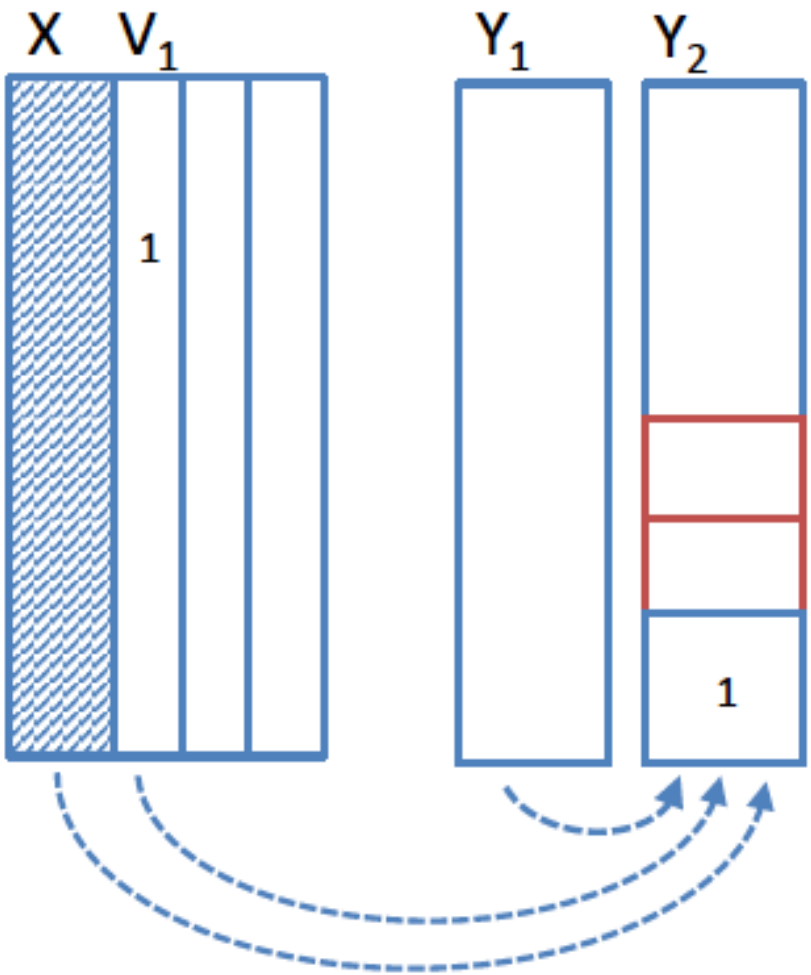
How to do blocked imputation

- 1. Estimate propensities to be missing due to different reasons missingness.
- 2. Multiply impute missing data under three scenarios: using X variables **and** each propensity from step 1.
- 3. Combine imputed missing values according to observed reason for missingness.

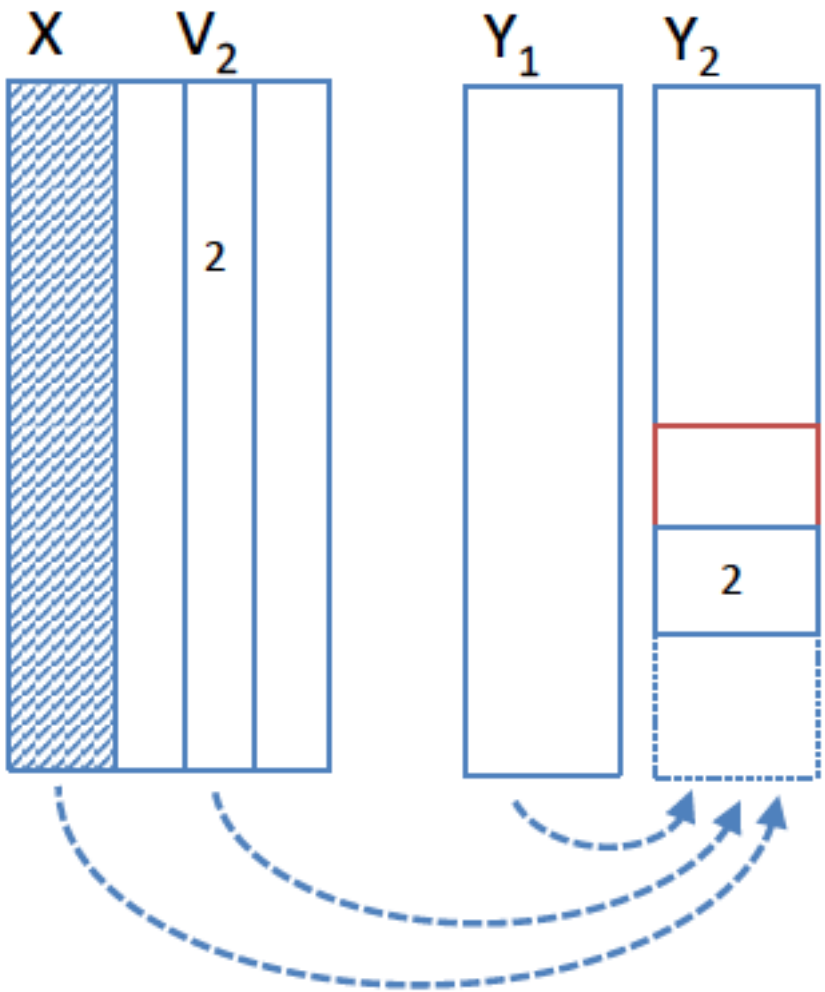
Entering covariates



Using different covariates for each block

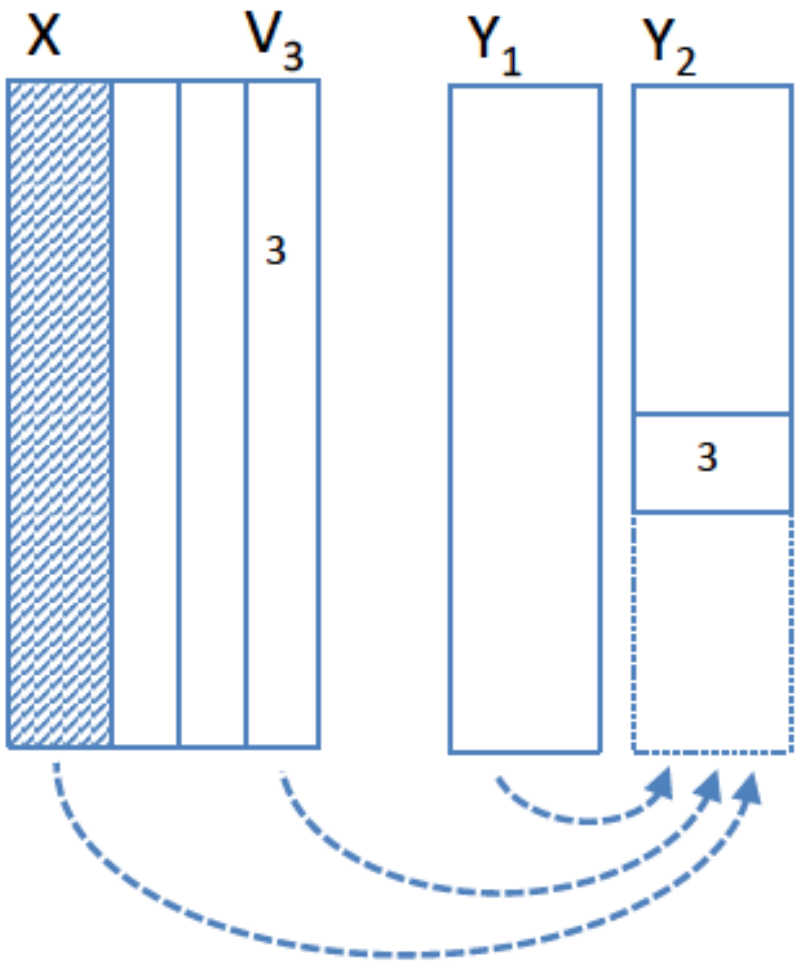


Using different covariates for each block

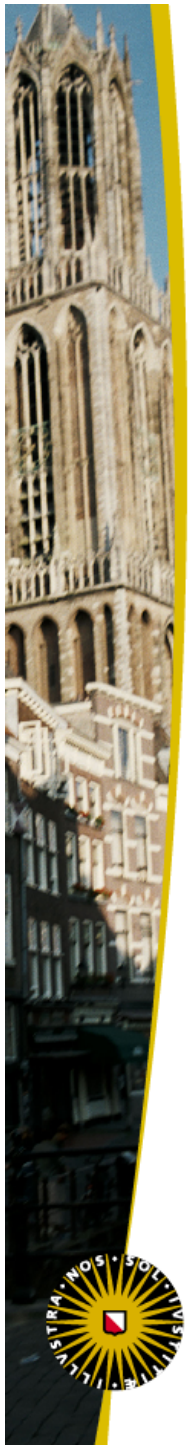


Universit

Using different covariates for each block



Universiteit U



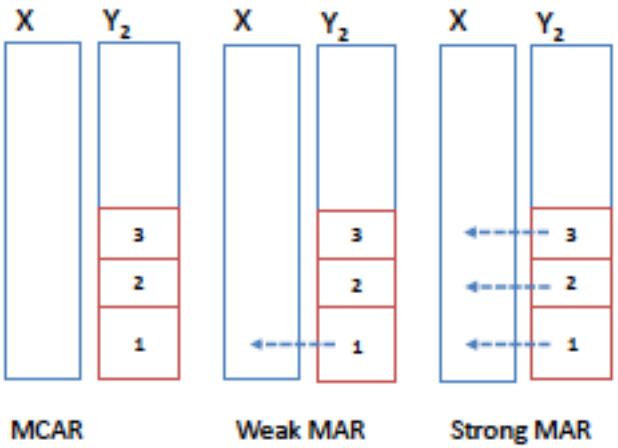
Does it work?

- Example from BHPS (1991-2008)
  - Dependent variable: moved/non-moved
  - Covariates (from previous wave):
    - educational level, move at previous wave, household size, household type, health status, marriage status, political interest, voted last election

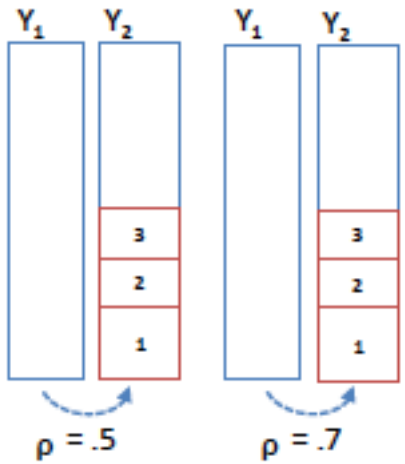
BHPS - results

Blocked Imputation	Mover	Non mover
Refusals	860	8856
Non-contact	868	8848
Ineligible	864	8852
Other non interview	862	8854
Overall Blocked imputation	**866**	**8850**

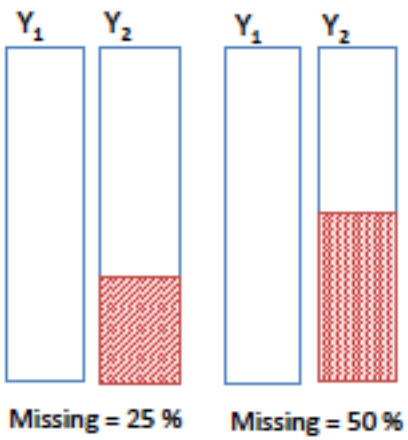
Simulation study



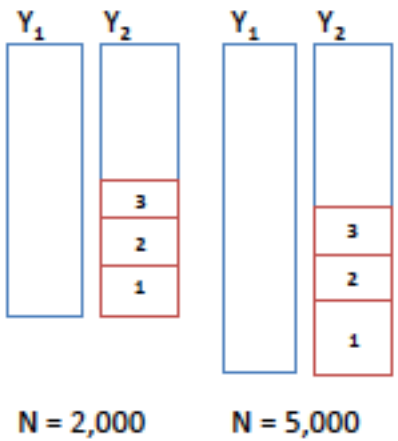
Missing data Mechanism



Correlation Y_1 and Y_2

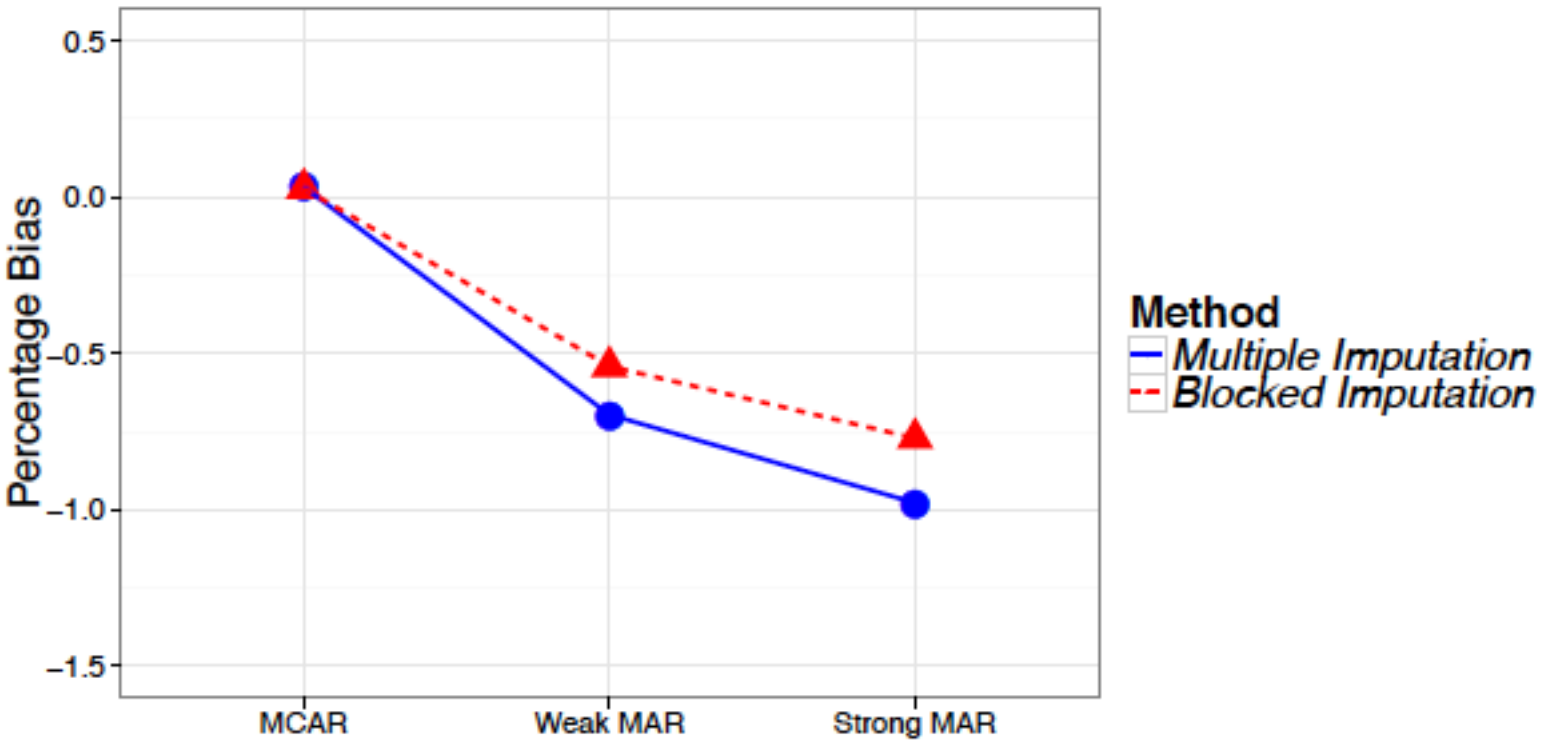


Percentage missing data



Sample Size

Percentage bias in mean



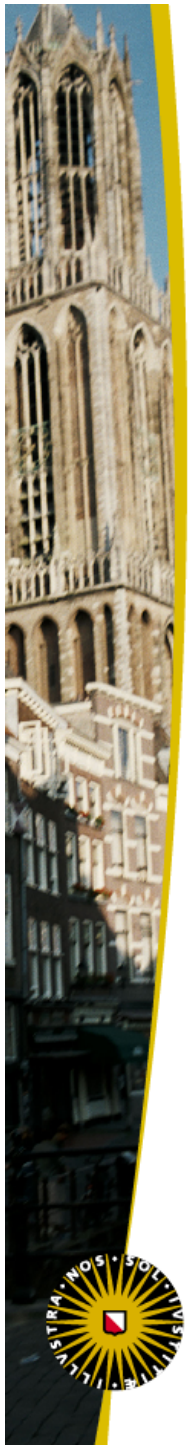
N= 5000, Corr(Y1,Y2)=0.7, missingness = 25%

Universiteit Utrecht

Conclusion Blocked imputation

- Does not perform worse if there are no strong covariates for explaining R (reasons missingness) and Y
- Performs better than Multiple Imputation when there are strong predictors for missingness
- Needs about 200 cases with a particular type of dropout in order to make modeling possible

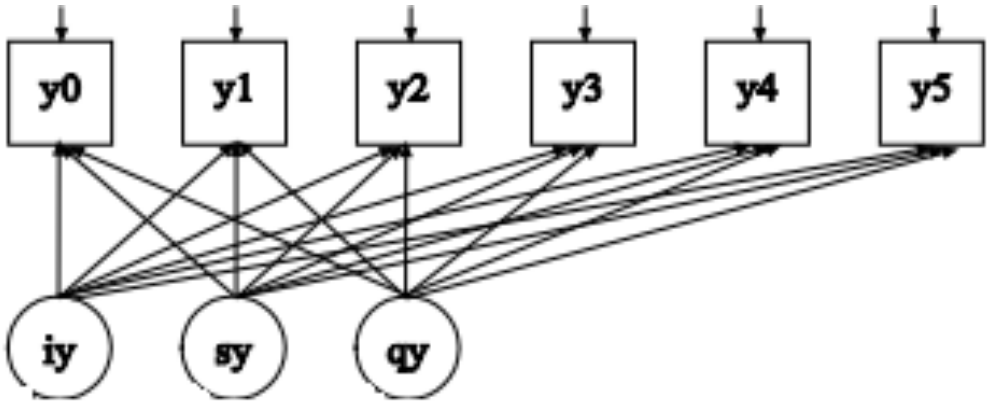
- Skip to end



2. Compensating – joint modeling

Universiteit Utrecht

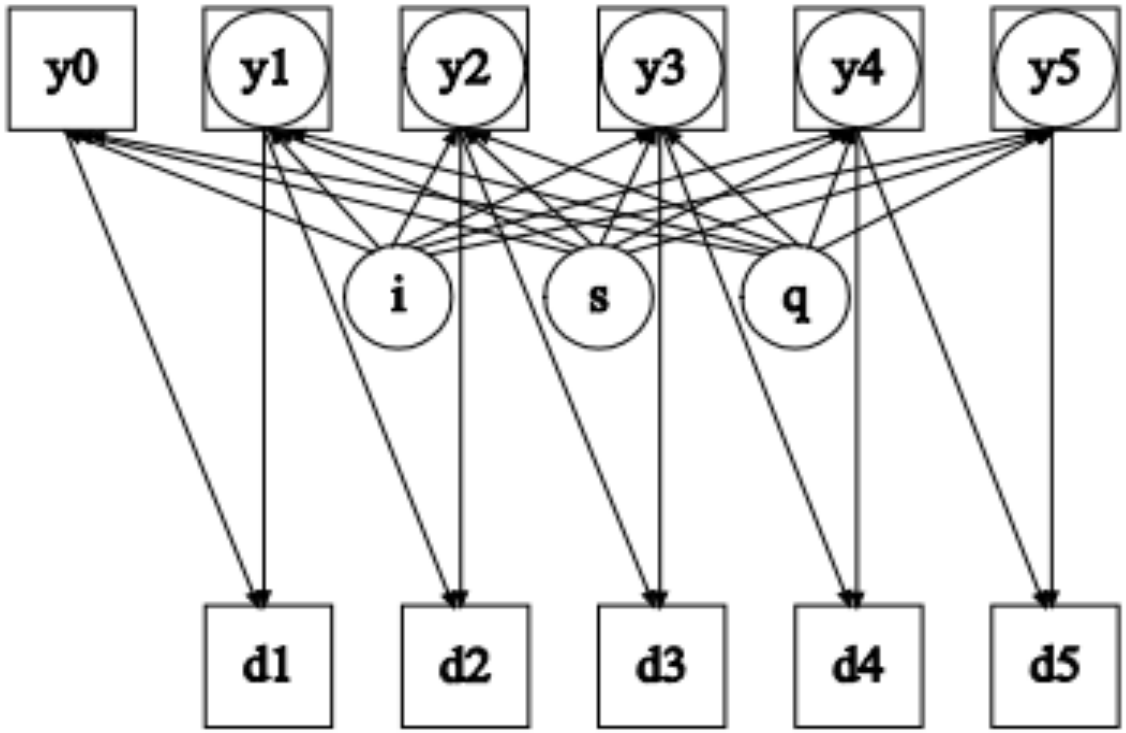
Model-based (FIML)



- Missings on Y are imputed (EM, MI)

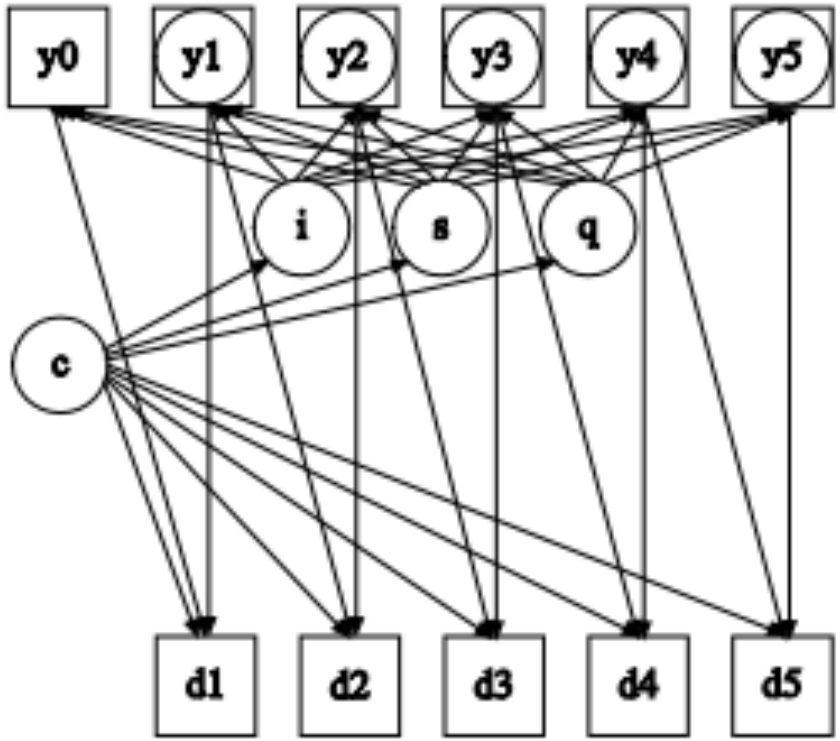
- See Muthen et al. (2010) http://www.statmodel.com/download/FebruaryVersionWide.pdf for all figures presented in the next 10 slides

2. Modeling missing data
Diggle-Kenward election model



- Growth model
- D's indicate who 'survives' from one wave to next (0-1)

2. Diggle-Kenward latent class



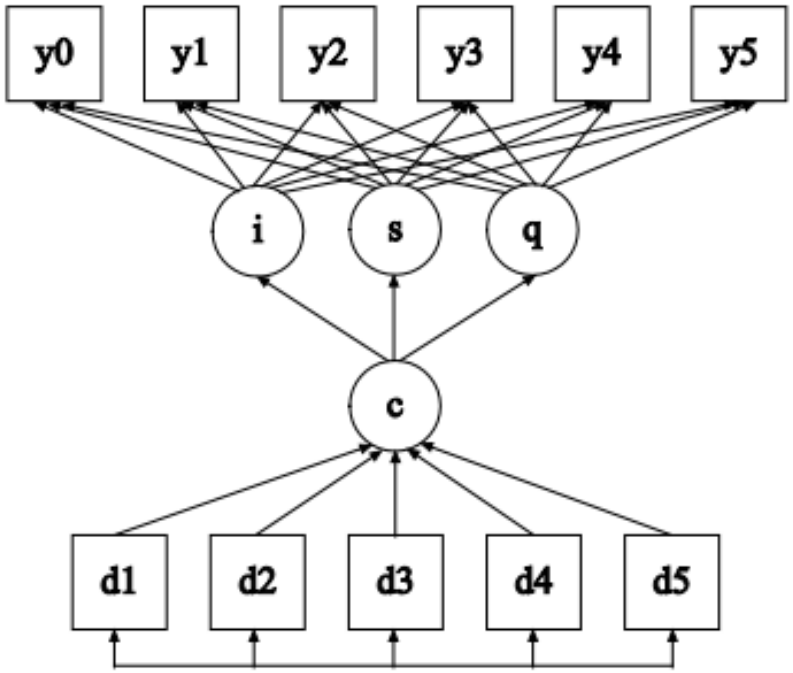
- Similar to previous model
- C indicates that there are unobserved groups (classes) with a different growth pattern **and** missing data process

Latent Class Analysis (LCA)

- Latent variables may be:
  - Continuous
  - Categorical
    - Nominal (AMOS, MPLUS, LISREL, LatentGold)
    - Ordinal (only LatentGold)

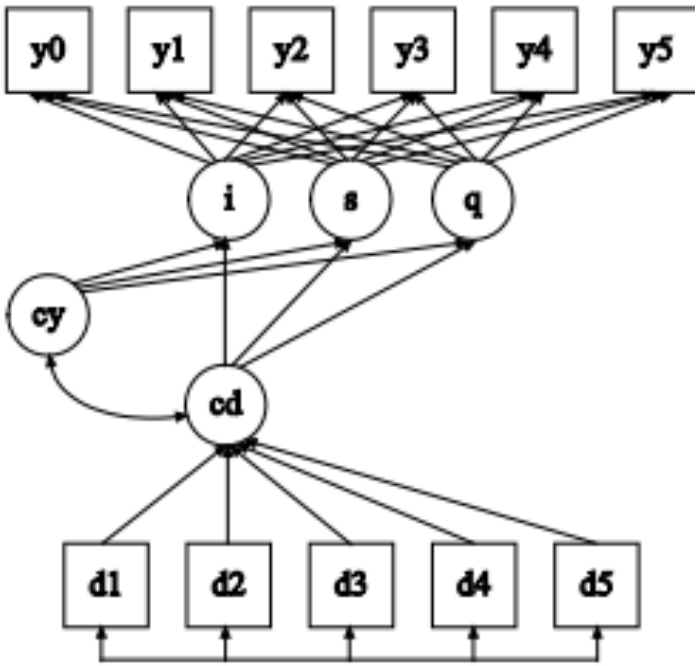
- Instead of Latent factors -> Latent classes
- Observed variables may be continuous or categorical

Roy latent class



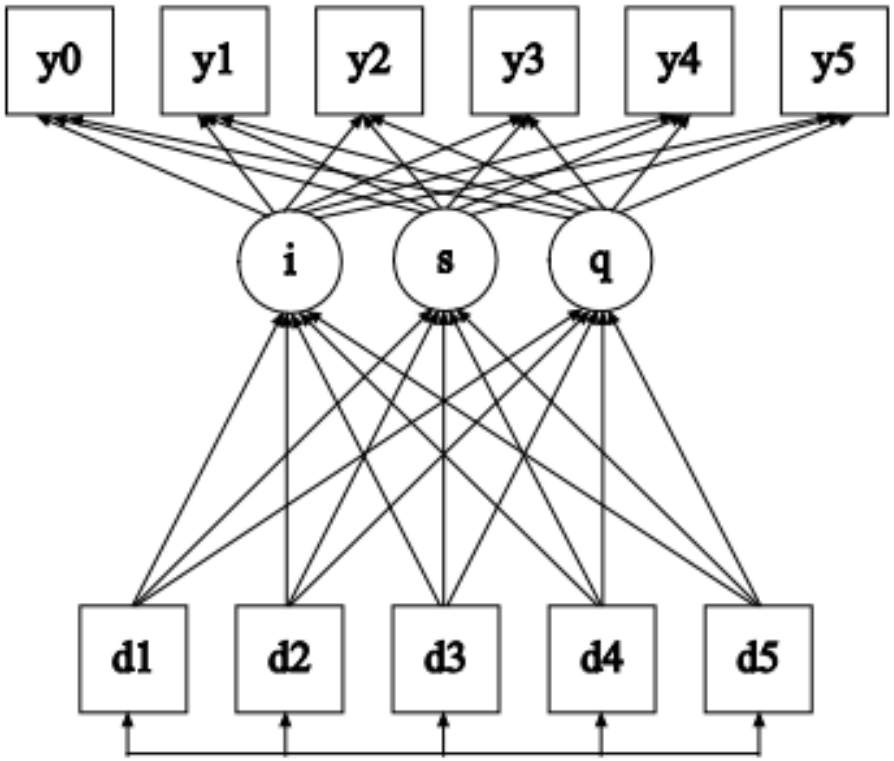
- D's (0-1) here only affect # latent classes (c)
  - D's are correlated…
- Classes of growth patterns formed on by substantive growth (y) and missingness (d)

2. Muthen Roy Latent Class model



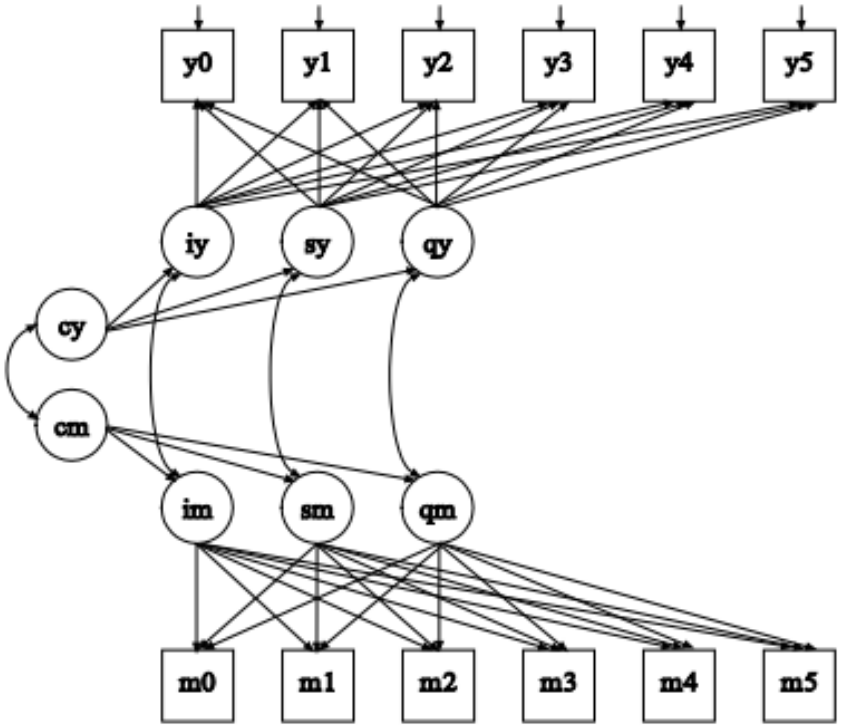
- Extension of Roy model
  - Latent classes for dropout
  - Latent classes for substantive process

2. Modeling missing data
Pattern mixture model



- D's (0-1) now directly affect growth parameters
  - No Latent classes!
- Regression parameters indicate selectiveness of missings

2. Two-part Joint modeling



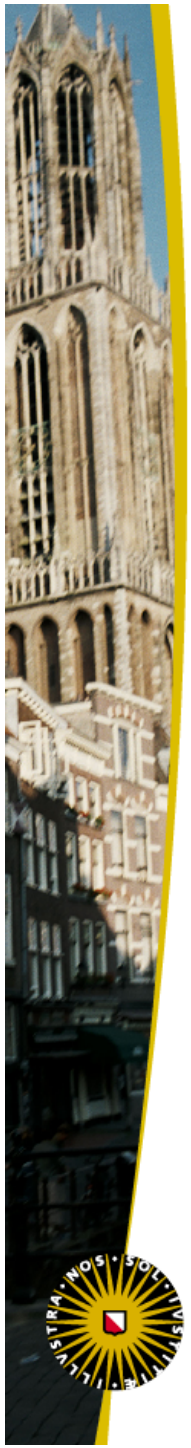
- Missingness here also 'growth' process
  - Two separate growth mixture models for 1. substantive process and 2. missing data

Extra slides

Universiteit Utrecht

Joint modeling of attrition in LISS

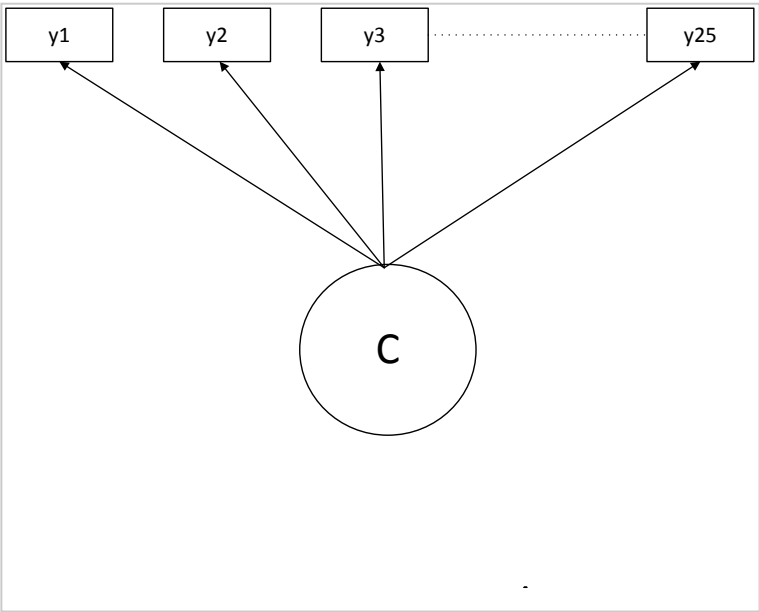
- Characteristics of attrition classes by covariates
  - Personal demographics
  - Household situation
  - Psychosocial variables
    - Big 5
    - Need for cognition, need to evaluate
  - Survey characteristics
    - Internet access and SimPC given
    - Survey enjoyment
    - Survey value
    - Survey burden



Latent class models

- 1. Latent Class
  - Response patterns (0-1) of 24 waves are summarized in a number of classes (1-12)
- 2. Latent Class Growth Analysis (LCGA)
  - Attrition is here considered a 'growth' process
  - Groups of people slowly drop out (or do so fast, or stay)
  - Advantage to Latent Class: parametric model, fewer parameters to estimate
- 3. Growth Mixture models (GMM)
  - Same as LCGA
  - Only difference: within classes room for 'unobserved heterogeneity
  - Growth parameters (i, s, q) may differ within a class.

1. Latent Class Model

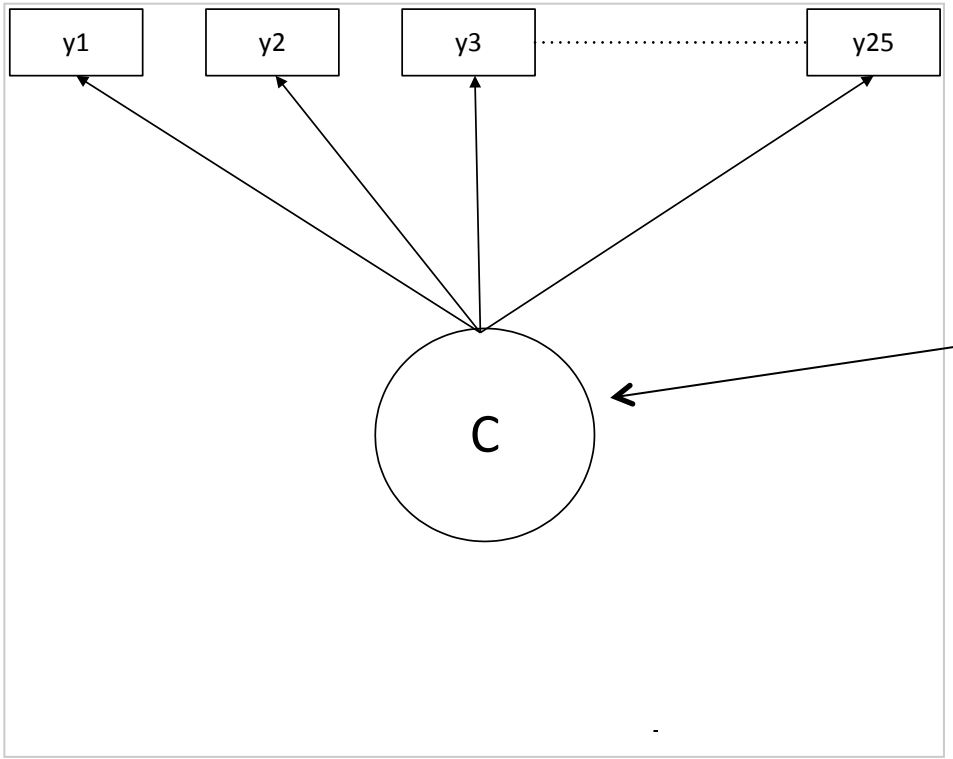


Response patterns (0-1) of 24 waves are summarized in a number of classes (1-12)

All 12 models run -> best model selected

Universiteit Utrecht

1b. Latent Class Model with covariates



Gender (1=f)
Age
Income (13 cat)
Education (7 cat)
Urbanicity
Partner(1=yes)
SimPC (1=yes)
<u>Psychological</u>
Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism
Need to evaluate
Need for cognition
<u>Survey attitude</u>
-1 Enjoy Internet
-2 Enjoy -interviewed
-3. Interesting
-4. Important
-5. A lot can be learned
-6. Waste of time
-7. Too many requests
-8. Privacy concerns
-9. exhaustive

Covariates can be used to predict class membership

MPLUS input for Latent Class Analysis

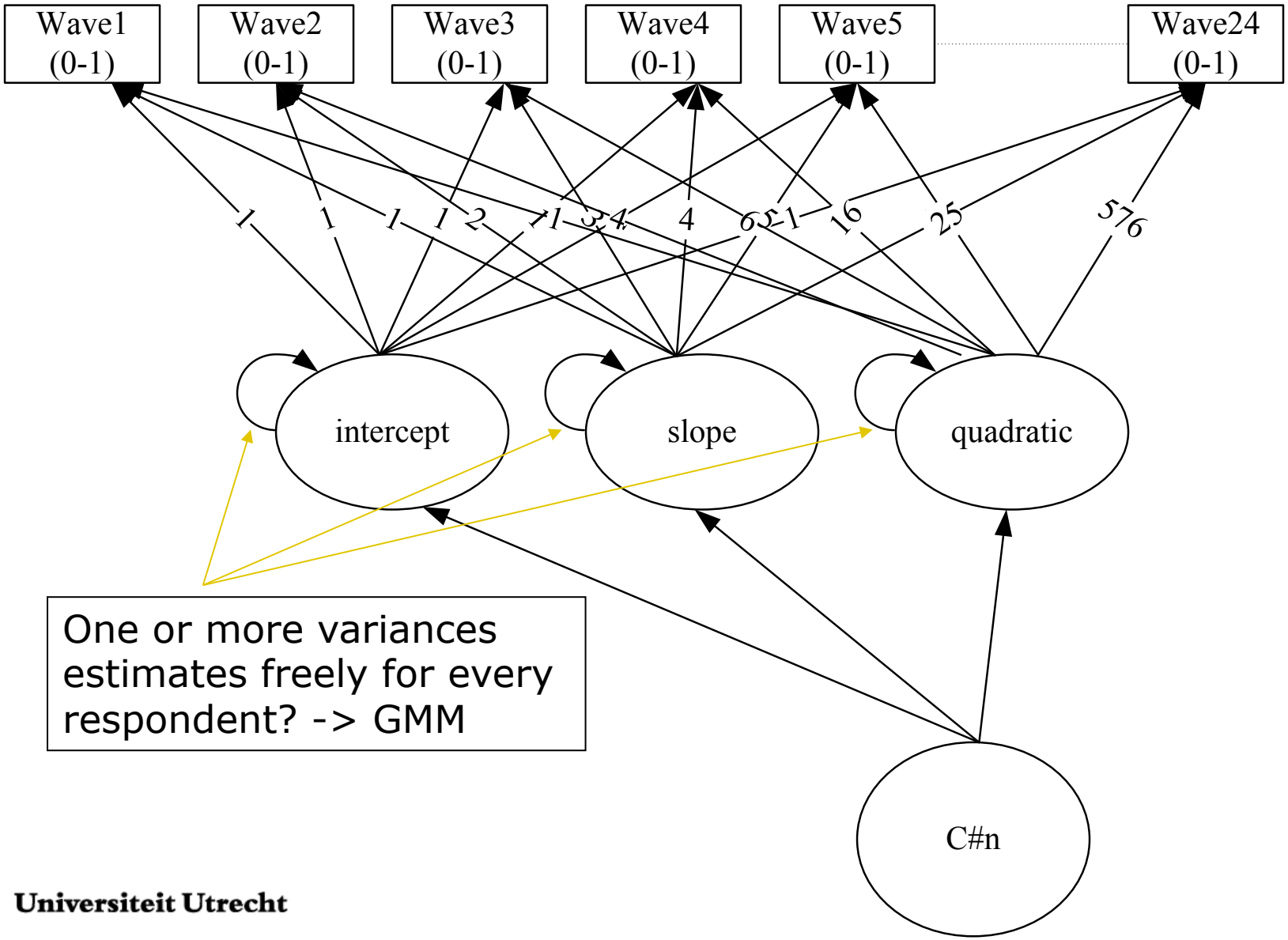
```
Variable:
categorical = vragenlijstenjan08
    vragenlijstenfeb08 vragenlijstenmar08 vragenlijstenapr08
    vragenlijstenmei08 vragenlijstenjun08 vragenlijstenjul08
    vragenlijstenaug08 vragenlijstensep08 vragenlijstenokt08
    vragenlijstennov08 vragenlijstendec08 vragenlijstenjan09
    vragenlijstenfeb09 vragenlijstenmar09
    vragenlijstenapr09 vragenlijstenmei09 vragenlijstenjun09
    vragenlijstenjul09 vragenlijstenaug09 vragenlijstensep09
    vragenlijstenokt09 vragenlijstennov09 vragenlijstendec09 != indicators for wave response
idvariable = nomem_encr;
    classes = c(2); ! Specifiy number of classes here
    cluster = nohouse_encr;

  analysis:
    Type = complex mixture;
    algorithm=integration;
  ! integration=10;        !reducing number of integration points may speed up at cost of precision
    starts = 100 10;
    processors = 2;

  model:
%overall% !nothing more is needed
```

Universiteit Utrecht

GMM/LCGA model + covariates (graphically)



One or more variances estimates freely for every respondent? -> GMM

Universiteit Utrecht

MPLUS input for Latent Class Growth analysis

```
Variable:
categorical = vragenlijstenjan08
   vragenlijstenfeb08 vragenlijstenmar08 vragenlijstenapr08 (…)!= indicators for wave response
idvariable = nomem_encr;
   classes = c(2); ! Specify the number of clases
   cluster = nohouse_encr;

 analysis:
   Type = complex mixture;
   algorithm=integration;
 ! integration=10;       !reducing number of integration points may speed up at cost of precision
   starts = 100 10;
   processors = 2;

 model:   !
%overall%
i s q| vragenlijstenjan08@0.1 vragenlijstenfeb08@0.2
     vragenlijstenmar08@0.3 vragenlijstenapr08@0.4
     vragenlijstenmei08@0.5 vragenlijstenjun08@0.6 vragenlijstenjul08@0.7
     vragenlijstenaug08@.8 vragenlijstensep08@.9 vragenlijstenokt08@1
     vragenlijstennov08@1.1 vragenlijstendec08@1.2 vragenlijstenjan09@1.3
     vragenlijstenfeb09@1.4 vragenlijstenmar09@1.5
     vragenlijstenapr09@1.6 vragenlijstenmei09@1.7 vragenlijstenjun09@1.8
     vragenlijstenjul09@1.9 vragenlijstenaug09@2.0 vragenlijstensep09@2.1
     vragenlijstenokt09@2.2 vragenlijstennov09@2.3 vragenlijstendec09@2.4;
```

MPLUS input for Growth Mixture model

```
Variable:
categorical = vragenlijstenjan08
   vragenlijstenfeb08 vragenlijstenmar08 vragenlijstenapr08 (…)!= indicators for wave response
idvariable = nomem_encr;
   classes = c(2); ! Specify the number of clases
   cluster = nohouse_encr;

  analysis:
   Type = complex mixture;
   algorithm=integration;
  ! integration=10;        !reducing number of integration points may speed up at cost of precision
   starts = 100 10;
   processors = 2;

  model:
%overall%
i s q| vragenlijstenjan08@0.1 vragenlijstenfeb08@0.2
     vragenlijstenmar08@0.3 vragenlijstenapr08@0.4
     vragenlijstenmei08@0.5 vragenlijstenjun08@0.6 vragenlijstenjul08@0.7
     vragenlijstenaug08@.8 vragenlijstensep08@.9 vragenlijstenokt08@1
     vragenlijstennov08@1.1 vragenlijstendec08@1.2 vragenlijstenjan09@1.3
     vragenlijstenfeb09@1.4 vragenlijstenmar09@1.5 (…)
       i@0;
       s*; ! Estimates a random slope variance
      q@0;
```

Universiteit Utrecht

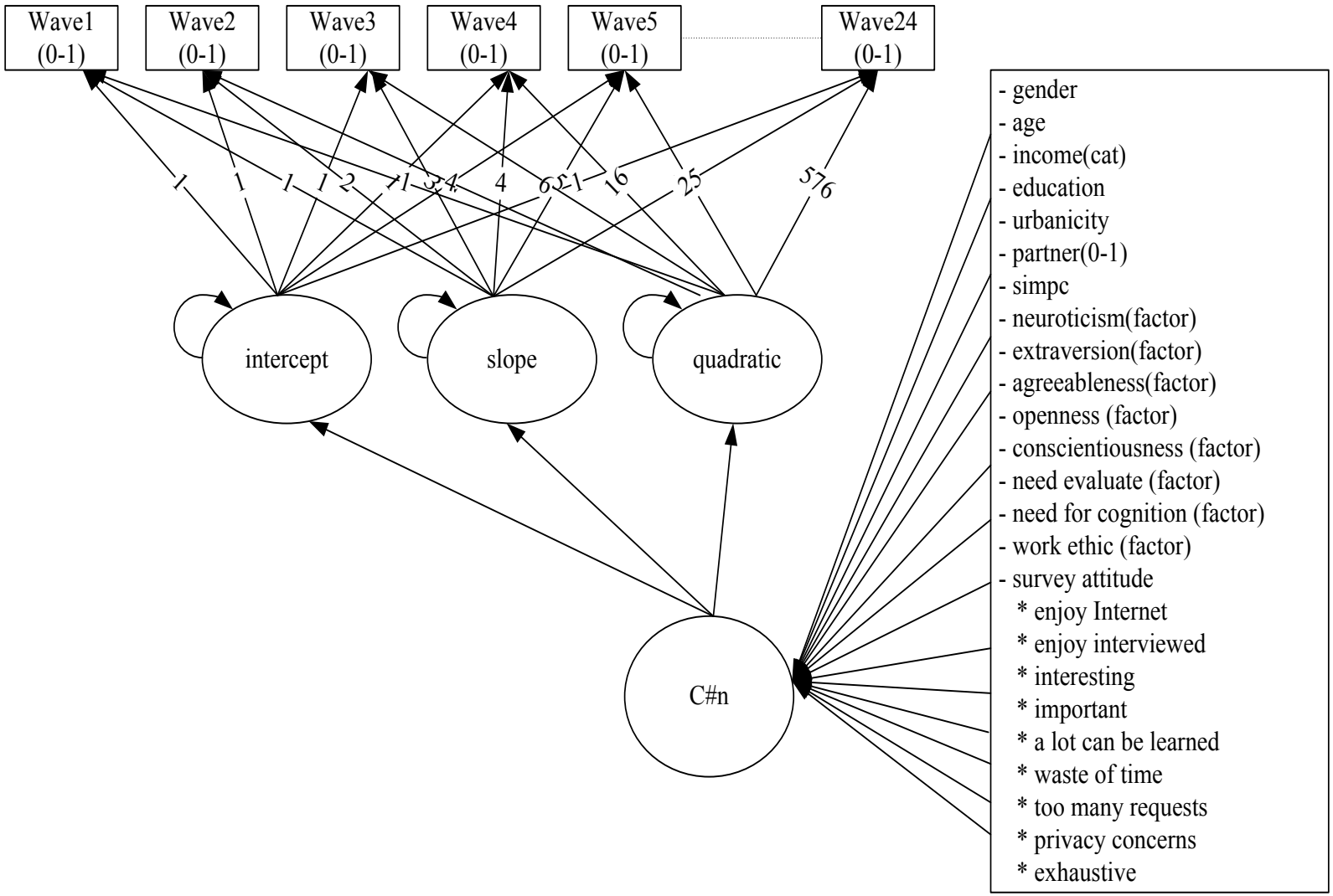
MPLUS input for GMM+ covariates

```
Variable:
categorical = vragenlijstenjan08
    vragenlijstenfeb08 vragenlijstenmar08 vragenlijstenapr08 (…)!= indicators for wave response
idvariable = nomem_encr;
    classes = c(2); ! Specify the number of clases
    cluster = nohouse_encr;

  analysis:
    Type = complex mixture;
    algorithm=integration;
  ! integration=10;        !reducing number of integration points may speed up at cost of precision
    starts = 100 10;
    processors = 2;

  model:   !
%overall%
i s q| vragenlijstenjan08@0.1 vragenlijstenfeb08@0.2 (…)
        i@0;
        s*; ! Estimates a random slope variance
       q@0;
 c on   geslacht nettocat oplcat sted partner simpc neuroticism
     extraversion agreeable openness conscientious needevaluate
     needcognition workethic health08
     sathouse satsurrou leftright satscience;
```

GMM model + covariates (graphically)



Effect of covariates – Stayers vs. all

Logit coefficients/ Covariates	Class 2 Slow begin, stayers	Class 8 Dropout after 40 months	Class1 Dropout after 30 months	Class 9 Dropout after 24 months	Class 6 Dropout after 18 months	Class 11 Dropout after 12 months	Class 4 High lurkers	Class 5 Lurkers, activated at month 30	Class 3 Low lurkers	Class 7 Dropout after 6 months	Class 12 Never participate
Socio-demographic											
Gender (1=f)	-.26 (.16)	-.11 (.19)	.02 (.14)	.23 (.15)	.08 (.15)	.23 (.12)	.11 (.13)	-.14 (.18)	.39 (.16)*	.003 (.09)	-.09 (.1)
Age	-.05 (.006)*	-.02 (.01)*	-.01 (.05)	-.03 (.006)*	-.03 (.006)*	-.02 (.004)*	-.04 (.005)*	-.04 (.007)*	-.06 (.008)*	-.02 (.003)*	-.04 (.004)*
Income (13 cat)	-.04 (.05)	-.07 (.06)	-.08 (.04)*	.08 (.04)	.04 (.04)	.08 (.03)*	.06 (.04)	.09 (.04)*	.14 (.04)*	-.064 (.03)	.1 (.03)*
Education (7 cat)	-.32 (.08)*	-.07 (.06)	-.04 (.05)	-.09 (.05)	-.07 (.05)	-.07 (.04)	.1 (.05)*	-.07 (.06)	-.11 (.05)*	-.09 (.04)*	-.11 (.04)*
Urbanicity	-.05 (.05)	.04 (.06)	.01 (.05)	-.02 (.05)	-.003 (.05)	-.07 (.04)	.05 (.29)	-.07 (.07)	-.13 (.08)	-.09 (.04)*	-.01 (.04)
Partner(1=yes)	.05 (.2)	.07 (.17)	-.14 (.017)	-.07 (.15)	-.004 (.17)	.14 (.14)	.2 (.13)	.18 (.22)	.23 (.23)	.39 (.12)*	.28 (.14)*
SimPC (1=yes)	.3 (.3)	.05 (.32)	-1.0 (.42)*	-.5 (.38)	-2.11 (.81)*	-.81 (.37)*	.05 (.29)	-.09 (.78)	-.27 (.44)	-.43 (.27)	-.95 (.52)
Psychological											
Openness	-.16 (.1)	-.05 (.09)	.01(.08)	.04 (.08)	.04 (.08)	.21 (.08)*	.03 (.08)	-.1 (.1)	.02 (.09)	-.04 (.06)	.003 (.05)
Conscientiousness	.08 (.08)	-.05 (.08)	-.08 (.08)	-.06 (.08)	-.12 (.07)	-.06 (.06)	-.21 (.06)*	-.31 (.08)*	-.15 (.08)	-.08 (.04)	-.13 (.04)*
Extraversion	.13 (.09)	.26 (.09)*	.11 (08)	.1 (.06)	.13 (.07)	.16 (.06)	.1 (.06)	.19 (.09)*	.28 (.09)*	.27 (.04)*	.28 (.04)*
Agreeableness	.03 (.07)	-.02 (.09)	-.14 (.08)	.07 (.09)	.11 (.08)	.01 (.07)	.24 (.07)*	.19 (.09)*	-.05 (.09)	.13 (.05)*	.12 (.04)*
Neuroticism	-.16 (.1)	-.02 (.08)	-.06 (.08)	-.02 (.07)	-.06 (.07)	-.07 (.06)	.08 (.06)	-.04 (.08)	.004 (.09)	.05 (.04)	.03 (.04)
Need to evaluate	.04 (.08)	-.08 (.08)	-.06 (.08)	-.1 (.08)	-.12 (.07)	-.14 (.06)	-.12 (.06)*	-.16 (.1)	-.19 (.09)*	-.2 (.05)*	-.22 (.04)*
Need for cognition	.09 (.1)	.15 (.1)	.02 (.09)	.006 (.09)	.05 (.1)	-.19 (.07)*	.08 (.08)	.27 (.11)*	.17 (.11)	.02 (.05)	.12 (.05)*
Survey attitude											
-1 Enjoy Internet	-.15 (.08)	-.2 (.05)*	-.18 (.05)*	-.25 (.05)*	-.21 (.05)*	-.29 (.05)*	-.07 (.04)	-.11 (.08)	-.04 (.06)	-.25 (.04)	-.18 (.06)*
-2 Enjoy - interviewed	.11 (.06)	.02 (.05)	.06 (.05)	-.06 (.04)	-.03 (.04)	-.08 (.04)*	-.05 (.04)	.007 (.08)	.03 (.06)	-.04 (.03)	-.005 (.04)
-3. Interesting	-.08 (.09)	-.1 (.06)	-.13 (.07)*	-.05 (.06)	-.18 (.05)*	-.14 (.05)*	-.06 (.05)	.17 (.09)	-.06 (.08)	-.07 (.05)	-.12 (.06)*
-4. Important	-.03 (.1)	-.05 (.06)	.08 (.06)	.05 (.06)	-.02 (.06)	-.03 (.05)	-.02 (.05)	-.19 (.1)	-.05 (.08)	.01 (.06)	.14 (.07)*
-5. A lot can be learned	.09 (.09)	.02 (.06)	-.06 (.06)	-.07 (.06)	.09 (.07)	.02 (.05)	.06 (.05)	-.04 (.11)	.03 (.12)	-.05 (.05)	.03 (.07)
-6. Waste of time	.06 (.07)	-.06 (.05)	.06 (.05)	.02 (.05)	.01 (.05)	.03 (.04)	.07 (.04)	.07 (.06)	.002 (.06)	.06 (.04)	.11 (.04)*
-7. Too many requests	.03 (.06)	.01 (.04)	-.04 (.04)	.05 (.04)	-.02 (.04)	-.01 (.03)	-.02 (.03)	.03 (.06)	.003 (.05)	.004 (.03)	-.01 (.03)
-8. Privacy concerns	.05 (.05)	-.03 (.04)	-.03 (.04)	.02 (.04)	.04 (.04)	.11 (.03)*	-.03 (.04)	-.05 (.08)	.006 (.06)	-.01 (.03)	.02 (.04)
-9. exhaustive	.1 (.06)	.09 (.04)*	.02 (.04)	.05 (.04)	.16 (.04)*	.13 (.03)*	.09 (.04)*	.05 (.06)	.11 (.05)*	.09 (.03)*	.09 (.03)*

Effect of covariates – Stayers vs. all (significant)

	Class 2	Class 8	Class1	Class 9	Class 6	Class 11	Class 4	Class 5	Class 3	Class 7	Class 12
Logit coefficients/ Covariates	Slow begin, stayers	Dropout after 40 months	Dropout after 30 months	Dropout after 24 months	Dropout after 18 months	Dropout after 12 months	High lurkers	Lurkers, activated at month 30	Low lurkers	Dropout after 6 months	Never participate
Socio-demographic											
Gender (1=f)									.39 (.16)		
Age	-.05 (.006)	-.02 (.01)		-.03 (.006)	-.03 (.006)	-.02 (.004)	-.04 (.005)	-.04 (.007)	-.06 (.008)	-.02 (.003)	-.04 (.004)
Income (13 cat)			-.08 (.04)			.08 (.03)		.09 (.04)*	.14 (.04)		.1 (.03)
Education (7 cat)	-.32 (.08)						.1 (.05)		-.11 (.05)	-.09 (.04)	-.11 (.04)
Urbanicity										-.09 (.04)	
Partner(1=yes)										.39 (.12)	.28 (.14)
SimPC (1=yes)			-1.0 (.42)		-2.11 (.81)	-.81 (.37)					
Psychological											
						.21 (.08)					
Conscientiousness							-.21 (.06)	-.31 (.08)			-.13 (.04)
Extraversion		.26 (.09)						.19 (.09)	.28 (.09)*	.27 (.04)*	.28 (.04)
Agreeableness							.24 (.07)	.19 (.09)		.13 (.05)*	.12 (.04)
Need to evaluate							-.12 (.06)		-.19 (.09)*	-.2 (.05)*	-.22 (.04)
Need for cognition						-.19 (.07)*		.27 (.11)			.12 (.05)
Survey attitude											
-1 Enjoy Internet		-.2 (.05)	-.18 (.05)	-.25 (.05)	-.21 (.05)	-.29 (.05)					-.18 (.06)
-2 Enjoy -interviewed						-.08 (.04)					
-3. Interesting			-.13 (.07)		-.18 (.05)	-.14 (.05)					-.12 (.06)
-4. Important											.14 (.07)
-6. Waste of time											.11 (.04)
-8. Privacy concerns						.11 (.03)					
-9. exhaustive		.09 (.04)			.16 (.04)	.13 (.03)	.09 (.04)		.11 (.05)	.09 (.03)	.09 (.03

Universiteit Utrecht

Effect of covariates – Stayers vs. slow starters

	Class 2
Logit coefficients/ Covariates	Slow begin, stayers
Socio-demographic	
Gender (1=f)	
Age	-.05 (.006)
Income (13 cat)	
Education (7 cat)	-.32 (.08)
Urbanicity	
Partner(1=yes)	
SimPC (1=yes)	
Psychological	
Conscientiousness	
Extraversion	
Agreeableness	
Need to evaluate	
Need for cognition	
Survey attitude	
-1 Enjoy Internet	
-2 Enjoy -interviewed	
-3. Interesting	
-4. Important	
-6. Waste of time	
-8. Privacy concerns	
-9. exhaustive	

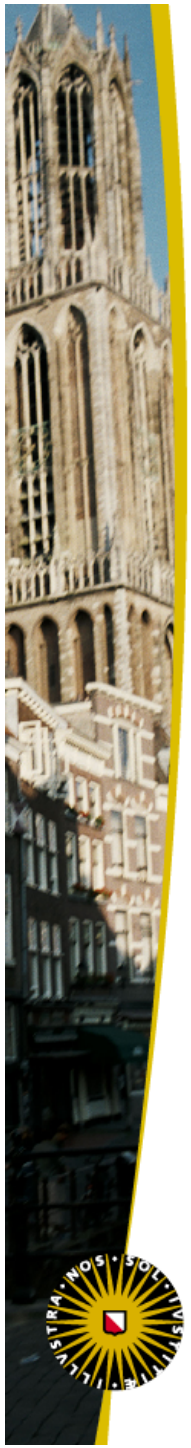
Universiteit Utrecht

Effect of covariates – stayers vs. dropout

	Class 8	Class1	Class 9	Class 6	Class 11	Class 7	Class 12
Logit coefficients/ Covariates	Dropout after 40 months	Dropout after 30 months	Dropout after 24 months	Dropout after 18 months	Dropout after 12 months	Dropout after 6 months	Never participate
Socio-demographic							
Gender (1=f)							
Age	-.02 (.01)		-.03 (.006)	-.03 (.006)	-.02 (.004)	-.02 (.003)	-.04 (.004)
Income (13 cat)		-.08 (.04)			.08 (.03)		.1 (.03)
Education (7 cat)						-.09 (.04)	-.11 (.04)
Urbanicity						-.09 (.04)	
Partner(1=yes)						.39 (.12)	.28 (.14)
SimPC (1=yes)		-1.0 (.42)		-2.11 (.81)	-.81 (.37)		
Psychological							
					.21 (.08)		
Conscientiousness							-.13 (.04)
Extraversion	.26 (.09)					.27 (.04)*	.28 (.04)
Agreeableness						.13 (.05)*	.12 (.04)
Need to evaluate						-.2 (.05)*	-.22 (.04)
Need for cognition					-.19 (.07)*		.12 (.05)
Survey attitude							
-1 Enjoy Internet	-.2 (.05)	-.18 (.05)	-.25 (.05)	-.21 (.05)	-.29 (.05)		-.18 (.06)
-2 Enjoy -interviewed					-.08 (.04)		
-3. Interesting		-.13 (.07)		-.18 (.05)	-.14 (.05)		-.12 (.06)
-4. Important							.14 (.07)
-6. Waste of time							.11 (.04)
-8. Privacy concerns					.11 (.03)		
-9. exhaustive	.09 (.04)			.16 (.04)	.13 (.03)	.09 (.03)	.09 (.03

Effect of covariates

	Class 4	Class 5	Class 3
Logit coefficients/ Covariates	Highly active lurkers	Lurkers, activated at month 30	Low active lurkers
Socio-demographic			
Gender (1=f)			.39 (.16)
Age	-.04 (.005)	-.04 (.007)	-.06 (.008)
Income (13 cat)		.09 (.04)*	.14 (.04)
Education (7 cat)	.1 (.05)		-.11 (.05)
Urbanicity			
Partner(1=yes)			
SimPC (1=yes)			
Psychological			
Conscientiousness	-.21 (.06)	-.31 (.08)	
Extraversion		.19 (.09)	.28 (.09)*
Agreeableness	.24 (.07)	.19 (.09)	
Need to evaluate	-.12 (.06)		-.19 (.09)*
Need for cognition		.27 (.11)	
Survey attitude			
-1 Enjoy Internet			
-2 Enjoy -interviewed			
-3. Interesting			
-4. Important			
-6. Waste of time			
-8. Privacy concerns			
-9. exhaustive	.09 (.04)		.11 (.05)



Results

- Differences with stayers
  - The earlier the dropout, the larger the differences
- Dropouts are:
  - Younger
  - More extravert
  - Enjoy being interviewed on the Internet less
  - Find it less interesting and more exhaustive
- Lurkers are:
  - Younger
  - Less conscientious, more extravert, more agreeable
  - Have a lower need to evaluate
  - Have a very similar survey attitude

Jump to 'what to do'

Universiteit Utrecht

Lugtig, P., Das, M. and Scherpenzeel, A.C. (2014) Nonresponse and attrition in a probability-based Online panel for the general population. Pp 135-154 in Callegaro, M., Lavrakas, P. J., Krosnick, J., Baker, Reginald P., Bethlehem, J. and Göritz, A. S. (Eds). Online panel research: A data quality perspective. Wiley series in survey methodology, New York: Wiley

References

- www.peterlugtig.com

Longitudinal survey errors

- Lynn, P. and Lugtig, P. (2017) Total Survey Error for Longitudinal Surveys. Chapter 13 in Biemer, P.P., De Leeuw, E.D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, C. and West, B.T. (eds.) *Total Survey Error in Practice.* New York: Wiley

Modeling of attrition

- Lugtig, P. (2014) Panel attrition: separating stayers, fast attriters, gradual attriters and lurkers. *Sociological Methods and Research* 43: 699-723. DOI: 10.1177/0049124113520305
- Lugtig, P., Das, M. and Scherpenzeel, A.C. (2014) Nonresponse and attrition in a probability-based Online panel for the general population. Pp 135-154 in Callegaro, M., Lavrakas, P. J., Krosnick, J., Baker, Reginald P., Bethlehem, J. and Göritz, A. S. (Eds). Online panel research: A data quality perspective. Wiley series in survey methodology, New York: Wiley

Imputation:

- Kamphuis, R.S., Jolani, S. and Lugtig, P. (submitted) The blocked imputation method. *Journal of the Royal Statistical Society: series A.*