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

8 May 2017
Attrition in national panel surveys

The recent increase in dropout in longitudinal surveys

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
British Household Panel Survey

- Face-to-Face panel, started in 1991

Response outcomes for every respondents in 18 waves of BHPS

- ineligible
- noncontact
- not issued
- refusal
- interview
- other noninterview
Sequence cluster analysis

Response outcomes for every respondents in 18 waves of BHPS

- Attrition after 13 waves
- Always Interviewed
- Dies or moves out of sample
- Very early refusals
- Early attrition
- Noncontacts

Legend:
- ineligible
- noncontact
- not issued
- refusal
- interview
- other noninterview
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
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
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  x  x  x  x  x  x  x  x  x  x  x  x  .....  x
LISS panel – attrition patterns

Loyal groups
- Stayers (always) 36.0%
- Late adopters 2.9%

38.9% combined
Results - dropout

Dropout groups
- Dropout around wave 40
- Dropout around wave 32
- Dropout around wave 24
- Dropout around wave 18
- Dropout around wave 12
- Dropout around wave 6
- Never participating group

46.7% combined
- 4.7%
- 4.3%
- 5.1%
- 5.6%
- 7.2%
- 10.9%
- 8.9%
Lurker groups
- Active lurkers 8.2%
- Lurkers activated around wave 30 2.5%
- Low active 3.8%

14.5% combined
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
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)

Y (variable of interest)

Z (auxiliary variables)

R (missingness)
Cross sectional nonresponse adjustments using frame variables

**X** (covariates)

**Y** (variable of interest)

**Z** (auxiliary variables)

**R** (missingness)

Weighting is sensible approach
Longitudinal nonresponse adjustments using earlier measures

X (covariates)

Y (variable of interest)

Z (auxiliary variables)

R (missingness)

imputation is sensible approach
How weak is relation x -> R?
2. Compensating for missing data

Correlates of attrition in BHPS
## Ever attrition – using wave 1 predictors

<table>
<thead>
<tr>
<th>Ever attrition (1=yes)</th>
<th>Strongest predictors</th>
<th>R2 nagelkerke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td></td>
<td>,09</td>
</tr>
<tr>
<td>Gender (F)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being in paid work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likely to move London</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using previous wave predictors</td>
<td>Strongest predictors</td>
<td>R2 nagelkerke</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Wave 2</td>
<td>Age, single non-elderly, noncompletion of tracking, not cooperative, higher degree, London</td>
<td>.15</td>
</tr>
<tr>
<td>Wave 3</td>
<td>Age, gender(M), not cooperative, higher degree, London</td>
<td>.07</td>
</tr>
<tr>
<td>Wave 4</td>
<td>Not having a car, couple with children, urban areas</td>
<td>.05</td>
</tr>
<tr>
<td>Wave 5</td>
<td>Age, completion of tracking, gender (m), couple with children</td>
<td>.07</td>
</tr>
<tr>
<td>Wave 6</td>
<td>Age, poor health, London</td>
<td>.08</td>
</tr>
<tr>
<td>Wave 7</td>
<td>Age, likely to move, self completion returned</td>
<td>.09</td>
</tr>
<tr>
<td>Wave 8</td>
<td>Not cooperative, self completion returned, higher degree</td>
<td>.08</td>
</tr>
<tr>
<td>Wave 9</td>
<td>Poor health, higher degree</td>
<td>.05</td>
</tr>
<tr>
<td>Wave 10</td>
<td>Not having a car</td>
<td>.09</td>
</tr>
<tr>
<td>Wave 11</td>
<td>Likely to move, higher degree</td>
<td>.06</td>
</tr>
<tr>
<td>Wave 12</td>
<td>Completion of tracking</td>
<td>.09</td>
</tr>
<tr>
<td>Wave 13</td>
<td>Gender(m)</td>
<td>.07</td>
</tr>
<tr>
<td>Wave 14</td>
<td>Poor health, Not cooperative, self completion returned</td>
<td>.12</td>
</tr>
<tr>
<td>Wave 15</td>
<td>Age, Likely to move, self completion returned</td>
<td>.12</td>
</tr>
<tr>
<td>Wave 16</td>
<td>Couple no children, no dependent children</td>
<td>.15</td>
</tr>
<tr>
<td>Wave 17</td>
<td>self completion returned</td>
<td>.05</td>
</tr>
<tr>
<td>Wave 18</td>
<td>Age, cooperation with interviewer</td>
<td>.15</td>
</tr>
</tbody>
</table>
## Type of attrition at wave 2 using wave 1 predictors

<table>
<thead>
<tr>
<th>Using previous wave predictors</th>
<th>Strongest predictors</th>
<th>R² nagelkerke</th>
</tr>
</thead>
</table>
| 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)
## Type of attrition at wave 2 using wave 1 predictors

<table>
<thead>
<tr>
<th>Using previous wave predictors</th>
<th>GLM R2 nagelkerke</th>
<th>RandomForest R2 nagelkerke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncontact</td>
<td>.22</td>
<td>.27</td>
</tr>
<tr>
<td>Refusal</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>Other non-interview</td>
<td>.08</td>
<td>.10</td>
</tr>
<tr>
<td>Ineligible (death)</td>
<td>.37</td>
<td>.39</td>
</tr>
</tbody>
</table>

Reference = Interview at wave 2
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
## Attrition – using wave 1 predictors

<table>
<thead>
<tr>
<th>Category</th>
<th>Strongest predictors</th>
<th>R2 nagelkerke</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 model</td>
<td>Age, single non-elderly, noncompletion of tracking, not cooperative, higher degree, London</td>
<td>0.09</td>
</tr>
<tr>
<td>Separated for reason</td>
<td>Age Planning to move</td>
<td>0.22</td>
</tr>
<tr>
<td>Noncontact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refusal</td>
<td>Bad cooperation with interviewer, Single elderly</td>
<td>0.05</td>
</tr>
<tr>
<td>Other non-interview</td>
<td>Bad cooperation with interviewer</td>
<td>0.08</td>
</tr>
<tr>
<td>Ineligible (death)</td>
<td>Age, Not having a car, All single households</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Reference = Interview at wave 2
Imputation using ‘type’ of attrition
How to include information on missingness?

\[ L(x) = Pr(S = 1|X = x), \]

Example: Propensity to be missing due to
S=1: Death \hspace{1cm} <- X
S=2: Refusal \hspace{1cm} <- X
S=3: Noncontact \hspace{1cm} <- 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
Using different covariates for each block
Does it work?

- Example from BHPS (1991-2008)
  - Dependent variable: moved/non-moving
  - Covariates (from previous wave):
    - educational level, move at previous wave, household size, household type, health status, marriage status, political interest, voted last election
<table>
<thead>
<tr>
<th>Blocked Imputation</th>
<th>Mover</th>
<th>Non mover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refusals</td>
<td>860</td>
<td>8856</td>
</tr>
<tr>
<td>Non-contact</td>
<td>868</td>
<td>8848</td>
</tr>
<tr>
<td>Ineligible</td>
<td>864</td>
<td>8852</td>
</tr>
<tr>
<td>Other non interview</td>
<td>862</td>
<td>8854</td>
</tr>
<tr>
<td>Overall Blocked imputation</td>
<td>866</td>
<td>8850</td>
</tr>
</tbody>
</table>
Simulation study

- Missing data Mechanism:
  - MCAR
  - Weak MAR
  - Strong MAR

- Correlation $Y_1$ and $Y_2$:
  - $\rho = .5$
  - $\rho = .7$

- Percentage missing data:
  - $\text{Missing} = 25\%$
  - $\text{Missing} = 50\%$

- Sample Size:
  - $N = 2,000$
  - $N = 5,000$
Percentage bias in mean

N= 5000, Corr(Y1,Y2)=0.7, missingness = 25%
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
Model-based (FIML)

- Missings on Y are imputed (EM, MI)

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
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.
1b. Latent Class Model with covariates

Covariates can be used to predict class membership

Gender (1=f)
Age
Income (13 cat)
Education (7 cat)
Urbanicity
Partner (1=yes)
SimPC (1=yes)
Psychological
Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism
Need to evaluate
Need for cognition
Survey attitude
-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
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); ! Specify 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
One or more variances estimates freely for every respondent? -> GMM
**MPLUS input for Latent Class Growth analysis**

**Variable:**
categorical = vragenlijstenjan08 vragenlijstenfeb08 vragenlijstenmar08 vragenlijstenapr08 (...)

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:**

```plaintext
! %overall%
```
```plaintext
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@0.8 vragenlijstensep08@0.9 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;
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 classes
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 opcat sted partner simpc neuroticism
  extraversion agreeable openness conscientious needdevalue
  needcognition workethic health08
  sathouse satsurrou leftright satscience;
GMM model + covariates (graphically)

- gender
- age
- income(cat)
- education
- urbanicity
- partner(0-1)
- simpc
- neuroticism(factor)
- extraversion(factor)
- agreeableness(factor)
- openness (factor)
- conscientiousness (factor)
- need evaluate (factor)
- need for cognition (factor)
- work ethic (factor)
- survey attitude
  * enjoy Internet
  * enjoy interviewed
  * interesting
  * important
  * a lot can be learned
  * waste of time
  * too many requests
  * privacy concerns
  * exhaustive
### Effect of covariates – Stayers vs. all

<table>
<thead>
<tr>
<th>Logit coefficients/Covariates</th>
<th>Class 2</th>
<th>Class 8</th>
<th>Class 9</th>
<th>Class 6</th>
<th>Class 11</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 3</th>
<th>Class 7</th>
<th>Class 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-demographic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (1=f)</td>
<td>-.26 (.16)</td>
<td>-.11 (.19)</td>
<td>.02 (.14)</td>
<td>.23 (.15)</td>
<td>.08 (.15)</td>
<td>.23 (.12)</td>
<td>.11 (.13)</td>
<td>-.14 (.18)</td>
<td>.39 (.16)</td>
<td>.003 (.09)</td>
</tr>
<tr>
<td>Age</td>
<td>-.05 (.06)*</td>
<td>-.02 (.01)*</td>
<td>-.01 (.05)</td>
<td>-.03 (.006)**</td>
<td>-.03 (.006)*</td>
<td>-.02 (.004)*</td>
<td>-.04 (.005)*</td>
<td>-.04 (.007)*</td>
<td>-.06 (.008)*</td>
<td>-.02 (.003)*</td>
</tr>
<tr>
<td>Income (13 cat)</td>
<td>-.04 (.05)</td>
<td>-.07 (.06)</td>
<td>-.08 (.04)*</td>
<td>-.08 (.04)</td>
<td>.04 (.04)</td>
<td>.08 (.03)*</td>
<td>.06 (.04)</td>
<td>.09 (.04)*</td>
<td>.14 (.04)*</td>
<td>-.064 (.03)</td>
</tr>
<tr>
<td>Education (7 cat)</td>
<td>-.32 (.08)*</td>
<td>-.07 (.06)</td>
<td>-.04 (.05)</td>
<td>-.09 (.05)</td>
<td>-.07 (.05)</td>
<td>-.07 (.04)</td>
<td>.1 (.05)*</td>
<td>-.07 (.06)</td>
<td>-.11 (.05)*</td>
<td>-.09 (.04)*</td>
</tr>
<tr>
<td>Urbanicity</td>
<td>-.05 (.05)</td>
<td>.04 (.06)</td>
<td>.01 (.05)</td>
<td>-.02 (.05)</td>
<td>-.03 (.05)</td>
<td>-.07 (.04)</td>
<td>.05 (.29)</td>
<td>-.07 (.07)</td>
<td>-.13 (.08)</td>
<td>-.09 (.04)*</td>
</tr>
<tr>
<td>Partner(1=yes)</td>
<td>.05 (.2)</td>
<td>.07 (.17)</td>
<td>-.14 (.017)</td>
<td>-.07 (.15)</td>
<td>-.04 (.17)</td>
<td>.14 (.14)</td>
<td>.2 (.13)</td>
<td>.18 (.22)</td>
<td>.23 (.23)</td>
<td>.39 (.12)*</td>
</tr>
<tr>
<td>SimPC (1=yes)</td>
<td>.3 (.3)</td>
<td>.05 (.32)</td>
<td>-.1 (.42)*</td>
<td>-.5 (.38)</td>
<td>-.21 (.81)*</td>
<td>-.81 (.37)*</td>
<td>.05 (.29)</td>
<td>-.09 (.78)</td>
<td>-.27 (.44)</td>
<td>-.43 (.27)</td>
</tr>
<tr>
<td>Psychological</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>-.16 (.1)</td>
<td>-.05 (.09)</td>
<td>.01 (.08)</td>
<td>.04 (.08)</td>
<td>.04 (.08)</td>
<td>.21 (.08)*</td>
<td>.03 (.08)</td>
<td>-.1 (.1)</td>
<td>.02 (.09)</td>
<td>-.04 (.06)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.08 (.08)</td>
<td>-.05 (.08)</td>
<td>-.08 (.08)</td>
<td>-.06 (.08)</td>
<td>-.12 (.07)</td>
<td>-.06 (.06)</td>
<td>-.21 (.06)*</td>
<td>-.31 (.08)*</td>
<td>-.15 (.08)</td>
<td>-.08 (.04)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.13 (.09)</td>
<td>.26 (.09)*</td>
<td>.11 (.08)</td>
<td>.1 (.06)</td>
<td>.13 (.07)</td>
<td>.16 (.06)</td>
<td>.1 (.06)</td>
<td>.19 (.09)*</td>
<td>.28 (.09)*</td>
<td>.27 (.04)*</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.03 (.07)</td>
<td>-.02 (.09)</td>
<td>-.14 (.08)</td>
<td>.07 (.09)</td>
<td>.11 (.08)</td>
<td>.01 (.07)</td>
<td>.24 (.07)*</td>
<td>.19 (.09)*</td>
<td>-.05 (.09)</td>
<td>.13 (.05)*</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.16 (.1)</td>
<td>-.02 (.08)</td>
<td>-.06 (.08)</td>
<td>-.02 (.07)</td>
<td>-.06 (.07)</td>
<td>-.06 (.06)</td>
<td>.08 (.06)</td>
<td>-.04 (.08)</td>
<td>.004 (.09)</td>
<td>.05 (.04)</td>
</tr>
<tr>
<td>Need to evaluate</td>
<td>.04 (.08)</td>
<td>-.08 (.08)</td>
<td>-.06 (.08)</td>
<td>-.1 (.06)</td>
<td>-.12 (.07)</td>
<td>-.14 (.06)</td>
<td>-.12 (.06)*</td>
<td>-.16 (.1)</td>
<td>-.19 (.09)*</td>
<td>-.2 (.05)*</td>
</tr>
<tr>
<td>Need for cognition</td>
<td>.09 (.1)</td>
<td>.15 (.1)</td>
<td>-.02 (.09)</td>
<td>.006 (.09)</td>
<td>.05 (.1)</td>
<td>-.19 (.07)*</td>
<td>.08 (.08)</td>
<td>.27 (.11)*</td>
<td>.17 (.11)</td>
<td>.02 (.05)</td>
</tr>
<tr>
<td>Survey attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1 Enjoy Internet</td>
<td>-.15 (.08)</td>
<td>-.2 (.05)*</td>
<td>-.18 (.05)*</td>
<td>-.25 (.05)*</td>
<td>-.21 (.05)*</td>
<td>-.29 (.05)*</td>
<td>-.07 (.04)</td>
<td>-.11 (.08)</td>
<td>-.04 (.06)</td>
<td>-.25 (.04)</td>
</tr>
<tr>
<td>-2 Enjoy interviewed</td>
<td>.11 (.06)</td>
<td>.02 (.05)</td>
<td>.06 (.05)</td>
<td>-.06 (.04)</td>
<td>-.03 (.04)</td>
<td>-.08 (.04)*</td>
<td>.05 (.04)</td>
<td>.007 (.08)</td>
<td>.03 (.06)</td>
<td>-.04 (.03)</td>
</tr>
<tr>
<td>-3. Interesting</td>
<td>-.08 (.09)</td>
<td>-.1 (.06)</td>
<td>-.13 (.07)*</td>
<td>-.05 (.06)</td>
<td>-.18 (.05)*</td>
<td>-.14 (.05)*</td>
<td>-.06 (.05)</td>
<td>.17 (.09)</td>
<td>-.06 (.08)</td>
<td>-.07 (.05)</td>
</tr>
<tr>
<td>-4. Important</td>
<td>-.03 (.1)</td>
<td>-.05 (.06)</td>
<td>.08 (.06)</td>
<td>.05 (.06)</td>
<td>-.02 (.06)</td>
<td>-.03 (.05)</td>
<td>-.02 (.05)</td>
<td>-.19 (.1)</td>
<td>-.05 (.08)</td>
<td>.01 (.06)</td>
</tr>
<tr>
<td>-5. A lot can be learned</td>
<td>.09 (.09)</td>
<td>.02 (.06)</td>
<td>-.06 (.06)</td>
<td>-.07 (.06)</td>
<td>.09 (.07)</td>
<td>.02 (.05)</td>
<td>.06 (.05)</td>
<td>-.04 (.11)</td>
<td>.03 (.12)</td>
<td>-.05 (.05)</td>
</tr>
<tr>
<td>-6. Waste of time</td>
<td>.06 (.07)</td>
<td>-.06 (.05)</td>
<td>.06 (.05)</td>
<td>.02 (.05)</td>
<td>.01 (.05)</td>
<td>.03 (.04)</td>
<td>.07 (.04)</td>
<td>.07 (.06)</td>
<td>.002 (.06)</td>
<td>.06 (.04)</td>
</tr>
<tr>
<td>-7. Too many requests</td>
<td>.03 (.06)</td>
<td>.01 (.04)</td>
<td>-.04 (.04)</td>
<td>.05 (.04)</td>
<td>-.02 (.04)</td>
<td>-.01 (.03)</td>
<td>-.02 (.03)</td>
<td>.03 (.06)</td>
<td>.003 (.05)</td>
<td>.004 (.03)</td>
</tr>
<tr>
<td>-8. Privacy concerns</td>
<td>.05 (.05)</td>
<td>-.03 (.04)</td>
<td>-.03 (.04)</td>
<td>.02 (.04)</td>
<td>.04 (.04)</td>
<td>.11 (.03)*</td>
<td>-.03 (.04)</td>
<td>-.05 (.08)</td>
<td>.006 (.06)</td>
<td>-.01 (.03)</td>
</tr>
<tr>
<td>-9. exhaustive</td>
<td>.1 (.06)</td>
<td>.09 (.04)*</td>
<td>.02 (.04)</td>
<td>.05 (.04)</td>
<td>.16 (.04)*</td>
<td>.13 (.03)*</td>
<td>.09 (.04)*</td>
<td>.05 (.06)</td>
<td>.11 (.05)*</td>
<td>.09 (.03)*</td>
</tr>
</tbody>
</table>
# Effect of covariates – Stayers vs. all (significant)

<table>
<thead>
<tr>
<th>Logit coefficients/ Covariates</th>
<th>Class 2</th>
<th>Class 8</th>
<th>Class 9</th>
<th>Class 6</th>
<th>Class 11</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 3</th>
<th>Class 7</th>
<th>Class 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-demographic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (1=f)</td>
<td>-05 (.06)</td>
<td>-02 (.01)</td>
<td>-03 (.06)</td>
<td>-03 (.06)</td>
<td>-02 (.004)</td>
<td>-04 (.005)</td>
<td>-04 (.007)</td>
<td>-06 (.008)</td>
<td>-02 (.003)</td>
<td>-04 (.004)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income (13 cat)</td>
<td>-.08 (.04)</td>
<td>.08 (.03)</td>
<td>.09 (.04)*</td>
<td>.14 (.04)</td>
<td>.1 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (7 cat)</td>
<td>-.32 (.08)</td>
<td>.1 (.05)</td>
<td>-.11 (.05)</td>
<td>-.09 (.04)</td>
<td>-.11 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urbanicity</td>
<td>-.09 (.04)</td>
<td>.39 (.12)</td>
<td>.28 (.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimPC (1=yes)</td>
<td>-1.0 (.42)</td>
<td>-2.11 (.81)</td>
<td>-.81 (.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.21 (.08)</td>
<td>-.21 (.06)</td>
<td>-.31 (.08)</td>
<td>-.13 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>.26 (.09)</td>
<td></td>
<td>.19 (.09)</td>
<td>.28 (.09)*</td>
<td>.27 (.04)*</td>
<td>.28 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.24 (.07)</td>
<td>.19 (.09)</td>
<td>.13 (.05)*</td>
<td>.12 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need to evaluate</td>
<td>-.12 (.06)</td>
<td>-.19 (.09)*</td>
<td>-.2 (.05)*</td>
<td>-.22 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for cognition</td>
<td>-.19 (.07)*</td>
<td>.27 (.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1 Enjoy Internet</td>
<td>-.2 (.05)</td>
<td>-.18 (.05)</td>
<td>-.25 (.05)</td>
<td>-.21 (.05)</td>
<td>-.29 (.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 Enjoy -interviewed</td>
<td>-.08 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3. Interesting</td>
<td>-.13 (.07)</td>
<td>-.18 (.05)</td>
<td>-.14 (.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4. Important</td>
<td>.14 (.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6. Waste of time</td>
<td>.11 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-8. Privacy concerns</td>
<td>.11 (.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-9. exhaustive</td>
<td>.09 (.04)</td>
<td>.16 (.04)</td>
<td>.13 (.03)</td>
<td>.09 (.04)</td>
<td>.11 (.05)</td>
<td>.09 (.03)</td>
<td>.09 (.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Effect of covariates – Stayers vs. slow starters

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

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

<table>
<thead>
<tr>
<th>Logit coefficients/ Covariates</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-demographic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (1=f)</td>
<td></td>
<td></td>
<td>.39 (.16)</td>
</tr>
<tr>
<td>Age</td>
<td>-.04 (.005)</td>
<td>-.04 (.007)</td>
<td>-.06 (.008)</td>
</tr>
<tr>
<td>Income (13 cat)</td>
<td>.09 (.04)*</td>
<td>.14 (.04)</td>
<td></td>
</tr>
<tr>
<td>Education (7 cat)</td>
<td>.1 (.05)</td>
<td>-.11 (.05)</td>
<td></td>
</tr>
<tr>
<td>Urbanicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner (1=yes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimPC (1=yes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Psychological</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.21 (.06)</td>
<td>-.31 (.08)</td>
<td>.28 (.09)*</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.19 (.09)</td>
<td>.19 (.09)</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.24 (.07)</td>
<td>.19 (.09)</td>
<td></td>
</tr>
<tr>
<td>Need to evaluate</td>
<td>-.12 (.06)</td>
<td>.27 (.11)</td>
<td>-.19 (.09)*</td>
</tr>
<tr>
<td>Need for cognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey attitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1 Enjoy Internet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 Enjoy -interviewed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3. Interesting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4. Important</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6. Waste of time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-8. Privacy concerns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-9. exhaustive</td>
<td>.09 (.04)</td>
<td>.11 (.05)</td>
<td></td>
</tr>
</tbody>
</table>
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

References

• www.peterlugtig.com

Longitudinal survey errors

Modeling of attrition

Imputation: