Modelling Long Call Sequences and Final Outcome in the Swedish Labour Force Survey to Reduce the Number of Unproductive Calls

International Workshop on Paradata and Adaptive and Responsive Survey Design

Manchester, 9-10th November 2015

Anton Johansson, Peter Lundquist and Sara Westling (Statistics Sweden)
Gabriele B. Durrant (University of Southampton)

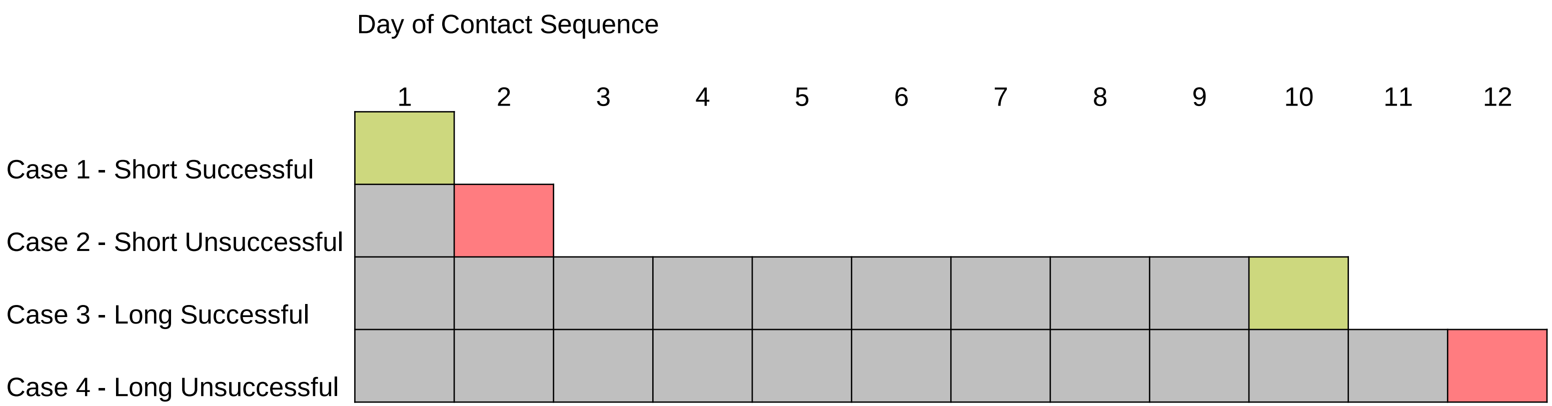
Outline

- Final outcome and contact sequences
- The Swedish Labour Force Survey (LFS)
- Interviewer hours spent on cases
- Applying Logistic regression models
- Examples of classification and trade-offs
- Future work

Introduction and Background

- Large survey resources are being spent on making unproductive calls
- Is it possible to use models to predict unsuccessful call outcomes?
- Ability of ‘classical’ nonresponse models without call data to predict nonresponse is often limited (R-squared values well below 10%)
- Can we improve these models by using paradata?
- This presentation here is motivated by recent work of Durrant et al. (2015) in JSSAM who assess the prediction of nonresponse models using paradata

Final Outcome and Length of Contact Sequence



- Successful if result of sequence is an interview (green)
- Unsuccessful if no interview (red)
- Long sequence if 8 or more contact days

Problem

- Each additional contact attempt requires that resources are allocated to the case.

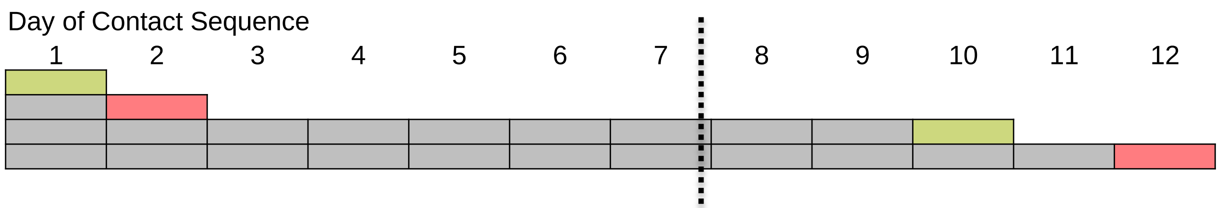
- Long unsuccessful sequences are the most ”unproductive” since they require a lot of resources but do not end in an interview.

Aims and Objectives

- We would like to allocate these resources elsewhere (e.g. to reduce nonresponse bias).

- Could we do this already during fieldwork?

- After “Phase 1”, when a case has had 7 days of unsuccessful contact attempts, would it be possible to predict which cases will end up as unsuccessful even after they are given additional contact attempts?

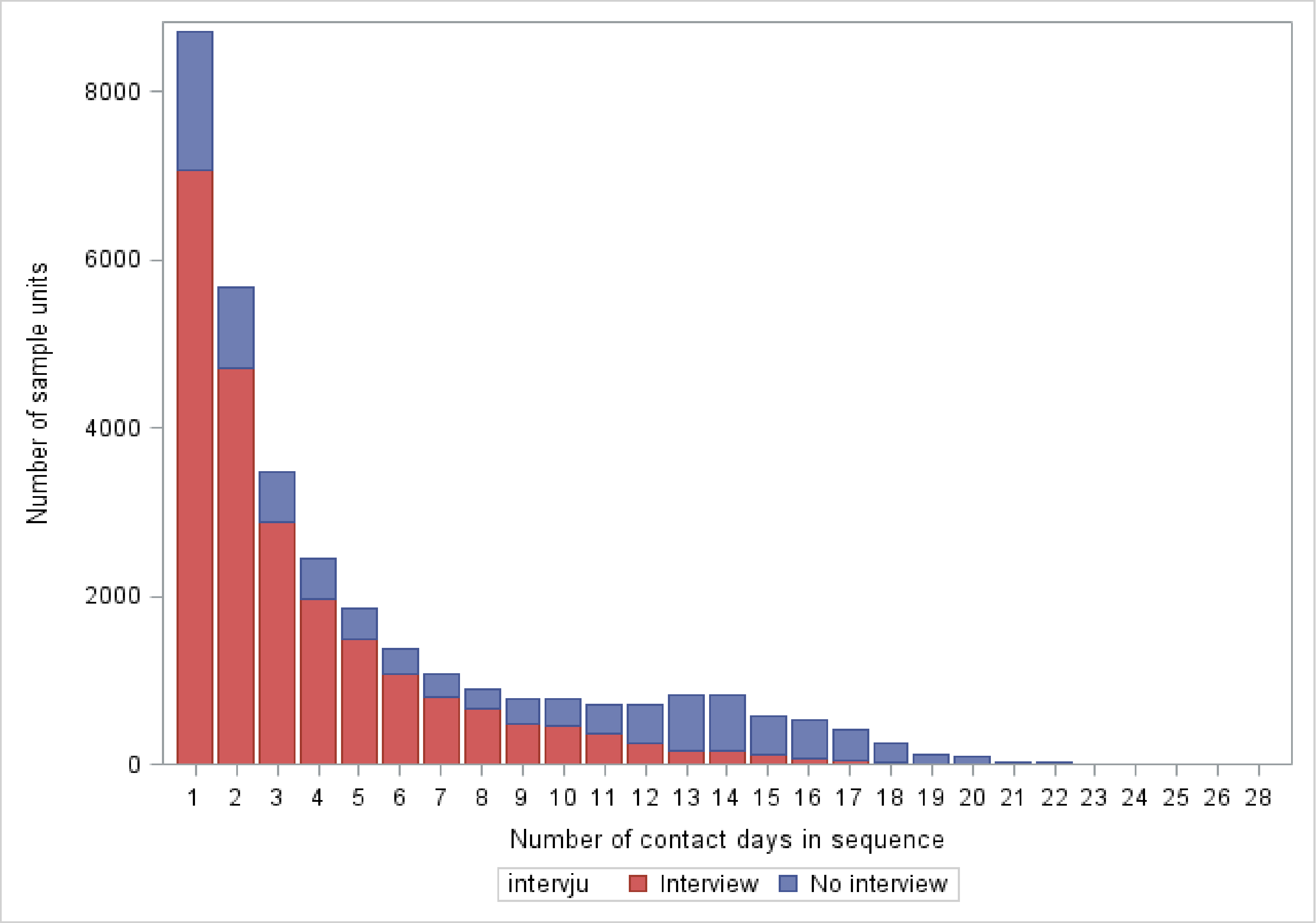


The Swedish Labour Force Survey

- Mode of data collection: CATI
- Monthly panel survey
- 21 500 individuals in ordinary sample each month
- Rotating Panel
  - *Sampled persons are in the sample 8 times (every third month for two years)*
- New panel each month:
  - *Approx. 2650 individuals*

We are studying these cases in our preliminary analysis. We use data from twelve months of data collection in LFS 2012.

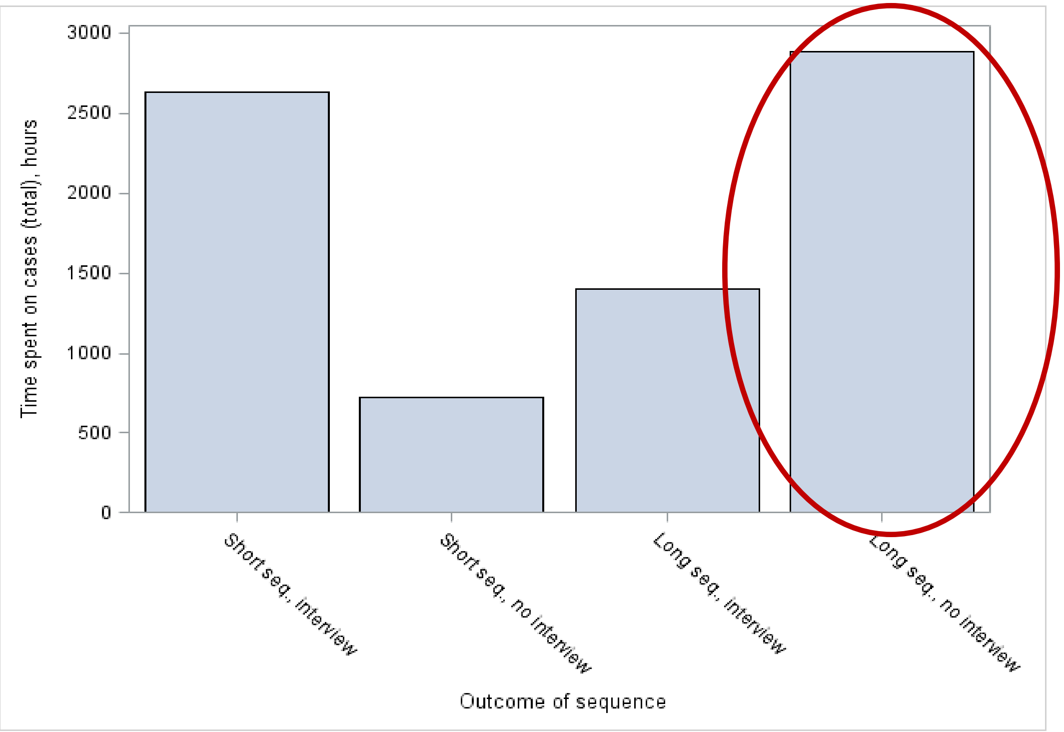
Sequences for LFS 2012



Sequences for LFS 2012

	Final Outcome Interview?		
Length of call sequence	Yes	No	Total
1-7 contact days	**62%**	**14%**	**76%**
	19987	*4666*	24653
8+ contact days	**9%**	**15%**	**24%**
	2805	*4807*	*7612*
Total	**71%**	**29%**	**100%**
	22792	9473	32265

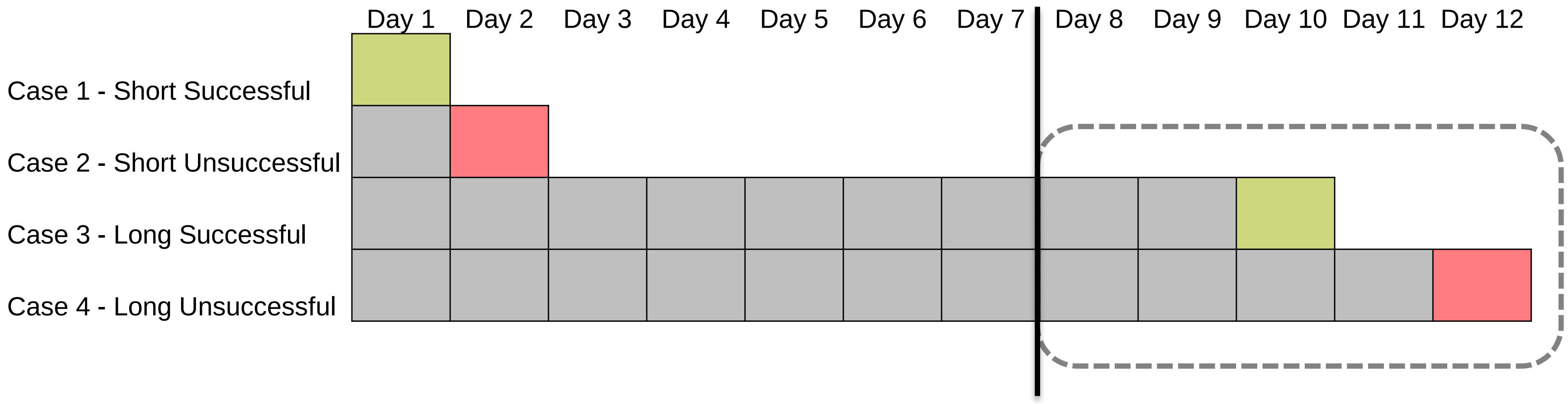
Interviewer Hours Worked



- Time measures of contact attempts used as a proxy for resources (hours) allocated to each sequence
- A large part of the hours are used for the long unsuccessful sequences

Applying Logistic Regression Models to Predict Long Unsuccessful cases

- Only study the cases with long sequences (8+ contact days)
- Use logistic regression model on already collected data in LFS 2012
- Can we use register data and paradata to predict the cases with ”long unsuccessful” final outcome?



Initial Models

- Model 1 – Register variables
  - Age (grouped),
  - High education (yes/no),
  - House Ownership (yes/no),
  - Benefits (yes/no),
  - High income (yes/no)
- Model 2 – Register variables + paradata
  - Variables above
  - Appointment during contact sequence (yes/no)

Assessment of models (1)

Table 2. Model summary	Nagelkerke R^2	Classification table	AUC (ROC-curve)
Model 1 – Register variables	0,0373	62,8%	0,5960
Model 2 – Register variables + paradata	0,0582	64,1%	0,6234

- **Pseudo-R^2 statistic**
  - Proportion of variation in the dependent variable that is explained by the model

Assessment of models (2)

Table 2. Model summary	Nagelkerke R^2	Classification table	AUC (ROC-curve)
Model 1 – Register variables	0,0373	62,8%	0,5960
Model 2 – Register variables + paradata	0,0582	64,1%	0,6234

- **Classification table**
  - Number of correctly classified cases
  - $P(y_i = 1 \; and \; \hat{y}_i = 1) + P(y_i = 0 \; and \; \hat{y}_i = 0)$
  - Cut-off value $\pi_0 = 0.5$
  - $\hat{y}_i = 1$ if $\pi_i > 0.5$
  - The number of correctly classified cases depends on the chosen cut-off value

Assessment of models (3)

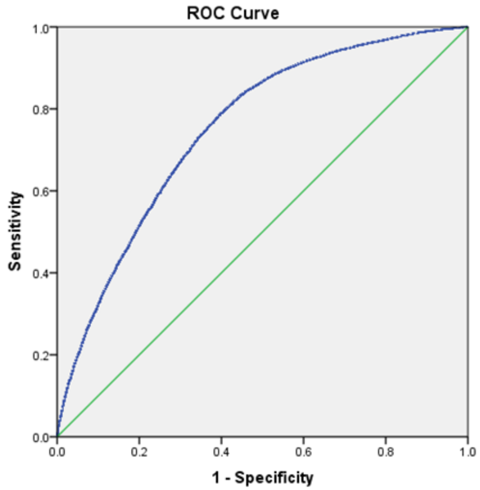
Table 2. Model summary	Nagelkerke R^2	Classification table	AUC (ROC-curve)
Model 1 – Register variables	0,0373	62,8%	0,5960
Model 2 – Register variables + paradata	0,0582	64,1%	0,6234

Area under ROC-curve

- Receiver Operating Curve (ROC)
- Summarizes predictive power for all possible cut-offs for π, by plotting sensitivity as a function of (1-specificity)

- Sensitivity: $P(\hat{y}_i = 1 | y_i = 1)$
- Specificity: $P(\hat{y}_i = 0 | y_i = 0)$

- The greater area under the curve (AUC) the greater the predictive power.
- AUC values range from 1 (perfect discrimination) to 0.5 (no discrimination).



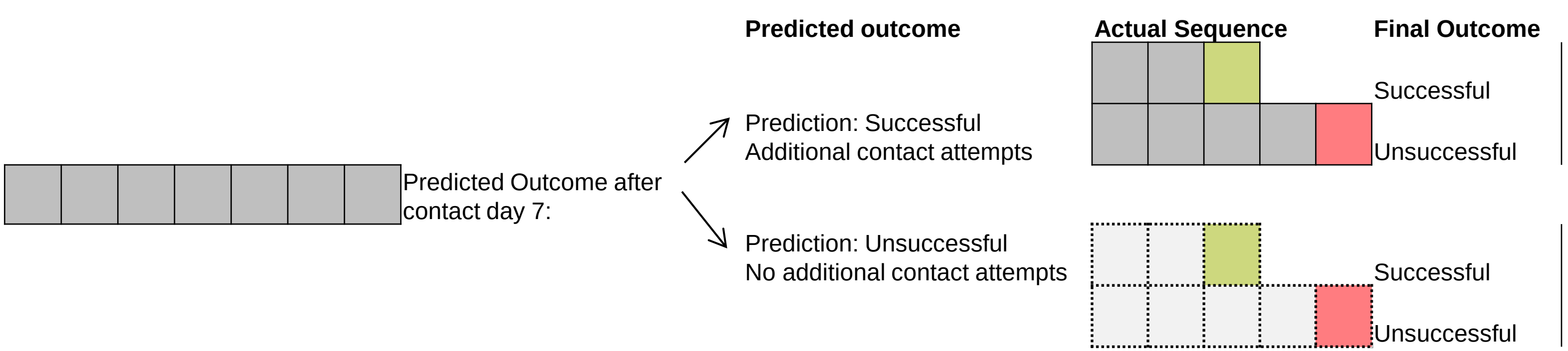
Diagonal segments are produced by ties.

Models need to be developed further

- Initial models not very useful for prediction
- Further analyses need to include additional variables in the models.

Stopping rule strategies and case prioritization

- We could consider (at least) two stopping rules after day 7
  - Stop calling all cases
  - Use model to predict ”final nonrespondents” and exclude them from further follow-up
- As an illustration of the second approach, we use the model on already collected data.



Trade-off

- Trade-off between saving resources and the effect the ”lost” data might have on key estimates.
- For cases predicted to be unsuccessful there are no additional contact attempts (i.e. no additional ”cost”).
- But, if the model do not discriminate well between successful and unsuccesful cases, we also lose data.
- The next step in our work is to look at how key estimates would have been affected by the stopping rule.

Examples of Classification and Trade-offs

Table 3. Stopping rule with chosen cutoff for predicting a case as interview ($\hat{\pi}_i \geq 0.5$)

Predicted outcome	**Actual outcome**	**Number of Cases**	**Time spent on contact attempts for the cases(hours)**
Nonrespondent	Nonrespondent	2197	680
	Respondent	1176	247
Respondent	Nonrespondent	178	55
	Respondent	207	38
	Total	3758	1020
Correctly classified cases:		64%	(2197+207)/3758
Correctly classified as nonrespondents:		93%	2197/(2197+178)
Correctly classified as respondents:		15%	207/(1176+207)
Hours "saved" (Hours spent on cases predicted to be nonrespondents):		927	(680+247)
Interviews "lost" (Respondents classified as nonrespondents):		1176	

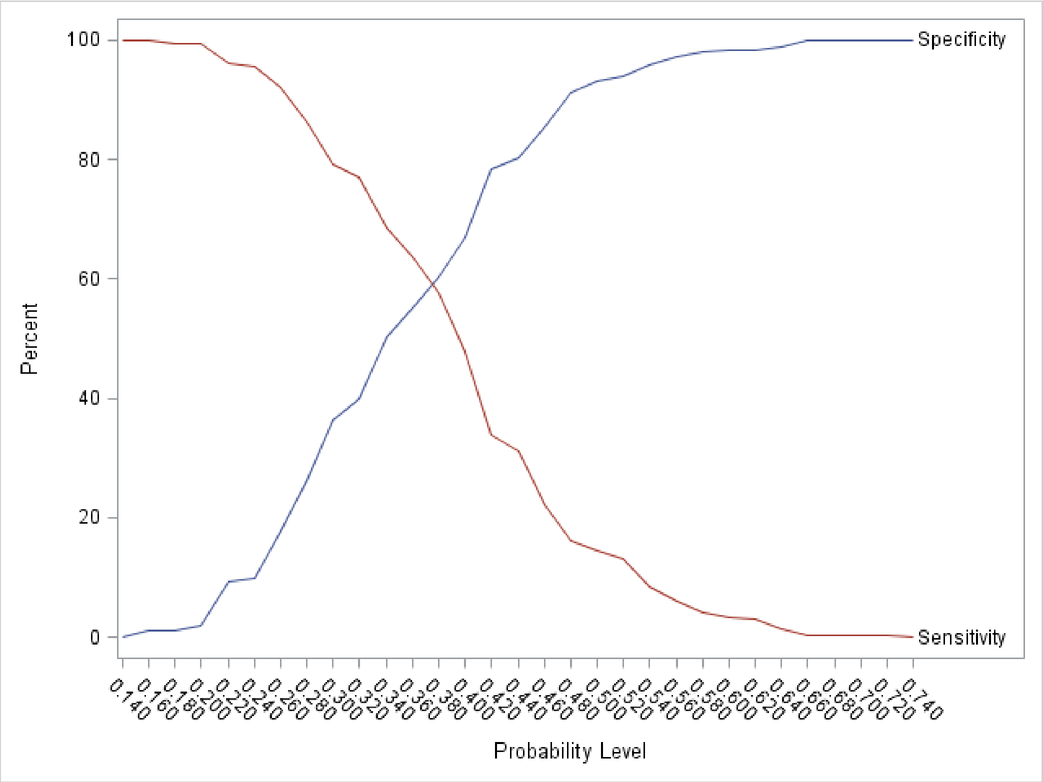
Impact of cut-off point

Table 4. Stopping rule with chosen cutoff for predicting a case as interview ($\hat{\pi}_i \geq 0.38$)

Predicted outcome	**Actual outcome**	**Number of Cases**	**Time spent on contact attempts for the cases(hours)**
Nonrespondent	Nonrespondent	1437	432
	Respondent	591	128
Respondent	Nonrespondent	938	304
	Respondent	792	158
	Total	3758	1022
Correctly classified cases:		59%	(1437+792)/3758
Correctly classified as nonrespondents:		61%	1437/(1437+938)
Correctly classified as respondents:		57%	792/(591+792)
Hours "saved" (Hours spent on cases predicted to be nonrespondents):		560	(432+128)
Interviews "lost" (Respondents classified as nonrespondents):		591	

Cut-off in example 2

- Sensitivity: $P(\hat{y}_i = 1 | y_i = 1)$
- Specificity: $P(\hat{y}_i = 0 | y_i = 0)$



Future Work

- Developing the logistic models predicting nonresponse using further paradata
  - *Is it possible to predict nonresponse? With what sort of variable?*

- Assessment on how key estimates in the LFS are affected by the stopping rules, excluding already collected observations according to the prediction of the models.

- Sensitivity analysis of the cut-off value of 8+ calls.

THANK YOU!