State-of-the-Art
University of Manchester

Kick-off Meeting
26-27 February 2015

Natalie Shlomo (Natalie.Shlomo@Manchester.ac.uk)

Social Statistics, School of Social Sciences
University of Manchester

The BADEN network is funded by the Leverhulme Trust International Network Grant, reference number: IN-2014-046

Introduction to the BADEN Network

Introduction

- Aim: to develop theory and practical implementation of adaptive survey designs; in particular, the development of a Bayesian framework to learn and update key input parameters to these designs
- The learning and updating of critical design parameters in a Bayesian framework needs to be done mostly in production environments under theoretical formulations
- Theory to be tested using simulation studies and applications that are subject to real-world constraints in multiple institutions and contexts

Introduction

Objectives:

- To bring together researchers on a periodic basis and to speed up theoretical development and practical
- implementations of adaptive survey designs;
- To establish a cross-institute research agenda for the main research aim;
- To design and implement joint simulation studies;
- To support discussion on theory and the exchange of empirical results;
- To disseminate work to a larger public in order to advocate ideas and to get feedback on feasibility and utility;
- To lay the groundwork for joint papers and other forms of collaboration;
- To assist implementation of adaptive survey designs.

Introduction

- The network will provide the opportunity to share current research and establish a coordinated research agenda across member institutes
- Institutes to focus on: (1) theoretical developments, (2) data analysis and simulation, (3) implementing experimental designs
- Theoretical activities include methods for specification of priors which may be informed using existing data
- Experiments will be planned to test results across multiple institutes and across multiple sources of data, including the testing of prior specifications and their predictive accuracy
- The network to create synergies and collaborations across institutions

Practical Information

- Network Facilitator: Laura Mitchell
- Collaboration Agreement: In brief, each institution will contribute to the annual and end-of-year reporting; be responsible for their own research in their own institution (i.e., not sub-contracted to the Univ. of Manchester); each person/institution owns their own intellectual property; there is no financial agreement since all expenses from the network grant via the network facilitator; 3 years from Jan. 12, 2015
- All travel/accommodation via Laura. We can claim per-diem expenses of 30 pounds a day – just need to sign the form
- Website: http://www.cmist.manchester.ac.uk/research/projects/baden/

Practical Information - Meetings

Year 1:

Kick-off meeting 26-27th Feb 2015 in Washington

Bi-lateral meeting Netherlands – May 2015 (Natalie and Peter to visit Barry to discuss research and planning)

Intermediary Network meeting Statistics Sweden, Stockholm – July/August 2015 (discuss development of analysis data sets and proposed simulation studies)

3rd International Workshop Advances in Adaptive and Responsive Survey Design– 9 - 10th November 2015, University of Manchester

Practical Information - Meetings

Year 2:

Bi-lateral meeting US Census Bureau – March 2016 (Natalie and Barry to visit Stephanie to discuss research and planning)

Network meeting and Invited paper session linked to AAPOR 2016, Texas – May 2016

Intermediary Network meeting University of Southampton – November/December 2016 (discuss results, planning, dissemination)

Practical Information - Meetings

Year 3:

Bi-lateral meeting University of Manchester– March 2017 (Barry and US member (?) to visit Natalie to discuss research and planning)

Network meeting and Invited paper session linked to JSM 2017, Baltimore – July 2017

4th International Workshop Advances in Adaptive and Responsive Survey Design- US institution (?)

State-of-the-Art

Ongoing work at the University of Manchester

related to

Adaptive Survey Design

Topics Covered

- Ongoing Projects
  - New version of SAS code and manual, version 2.1
  - Adaptive Survey Designs Using R-indicators
  - Assessing Optimal Strategies to Reduce Non-response in Longitudinal Studies
  - Population based R-indicators for Monitoring and Improving Representativeness of Response
- Future Work

New version of SAS, R code and Manual – 2.1

Vincent de Heij, Barry Schouten and Natalie Shlomo

- Previous software (R and SAS) versions included:

$$CV(X) = \frac{1 - R(X)}{2\bar{\rho}}$$ with variance estimated as:

$$Var(CV(X)) \cong \frac{S^2(\hat{\rho}_X)}{\hat{\bar{\rho}}^2}\left[\frac{S^2(\hat{\rho}_X)}{n\hat{\bar{\rho}}^2} + \frac{S^2}{S(\hat{\rho}_X)^2}\right] = \frac{S^2}{\hat{\bar{\rho}}^2} + \frac{S^4(\hat{\rho}_X)}{n\hat{\bar{\rho}}^4}$$

where $S^2(\hat{\rho}_X)$ = variance of response propensities under full model and S^2 the estimated variance of $S(\hat{\rho}_X)$ as calculated for the R-indicator

New code in R and SAS added in version 2.1:

$$CV_U(X_k) = \frac{P_U(X_k)}{\hat{\bar{\rho}}}$$

$$CV_C(X_k) = \frac{P_C(X_k)}{\hat{\bar{\rho}}}$$

- Variable-level unconditional CV
- Variable-level conditional CV

and similarly for the category levels

with estimated variances, eg. for unconditional CV:

- $$Var(CV_U(X_k)) \cong \frac{S^2}{\hat{\bar{\rho}}^2} + \frac{P_U^2(X_k)S^2(\hat{\rho}_X)}{n\hat{\bar{\rho}}^4}$$

S^2 estimated variance of the unconditional (conditional) partial R-indicator

Adaptive Survey Design Using R-indicators

Barry Schouten and Natalie Shlomo (submitted)

Paper includes:

- Introduction and motivation for using R-indicators to monitor data collection
- Review of R-indicators and partial R-indicators and presents new theory on bias corrections and standard errors
- Using R-indicators to build and evaluate profiles of nonresponse
- Optimization of an intervention
- Simulation Study to investigate impact of sample size on intervention decisions
- Real dataset example: Dutch Crime Victimisation Survey.

Introduction to paper:

Steps of Adaptive Survey Designs:

1. Choose proxy measures for survey quality;
2. Choose a set of candidate design features, e.g. survey modes or incentives;
3. Define cost constraints and other practical constraints;
4. Link available frame data, administrative data and paradata;
5. Form strata with the auxiliary variables for which design features can be varied;
6. Estimate input parameters (e.g. contact and participation propensities, costs);
7. Optimize the allocation of design features to the strata;
8. Conduct, monitor and analyse data collection;
9. In case of incidental deviation from anticipated quality or costs, return to step 7;
10. In case of structural deviation from anticipated quality or costs, return to step 6;
11. Adjust for nonresponse in the estimation

Introduction to paper:

Research questions:

- Can partial R-indicators be used to identify (and monitor) strata for adaptive survey designs?
- If so, how to optimize intervention for the strata?
- How to account for the frequency and length of the survey data collection in the optimization?

Building nonrespondent profiles

Can partial R-indicators be used to identify strata for adaptive survey designs?

- Partial R-indicators allow the building of profiles (characteristics) where more or less attention is required in the data collection to reduce the contrast between respondents and non-respondents
- Derivations for bias correction and standard errors for conditional and unconditional partial R-indicators at the variable/categorical levels are presented
- When monitoring data collection use the coefficient of variation (partial CVs also now included in the software and manual on the website)
- Selection of auxiliary variables: independent and cover different dimensions of target population
- Build nonresponse profiles from variables/categories with large and significant unconditional and conditional R-indicators

Optimizing intervention

How to optimize intervention given a set of strata?

- 4 strategies in the literature for optimization of adaptive survey designs:
  - a trial-and-error approach (e.g. Laflamme and Karaganis 2010, Luiten and Schouten 2013);
  - a set of stopping rules (e.g. Lundquist and Särndal 2013);
  - propensity-based prioritization (e.g. Peytchev et al 2010, Wagner 2013, Wagner and Hubbard 2013);
  - a mathematical optimization problem (e.g. Schouten, Calinescu and Luiten 2013).

- Proposed method: Structured trial and error approach:

Accounting for survey type

How to account for the frequency and length of the survey data collection in the optimization?

Non-response profiles built depending on one-off survey or ongoing (panel or longitudinal) survey

General algorithm:

- Inspect variable-level partial R-indicators and select variables for which unconditional and conditional values are significantly different from zero;
- Select all categories of those variables that have a significant positive unconditional value and a significant conditional value;
- Form a stratification by crossing all categories and, possibly, collapse empty or small strata;
- Compute the category-level unconditional partial R-indicator for the new stratification variable and order the strata by their sign and *p*-value;
- Select strata for follow-up based on their rank until *x* cases are selected (x cases can be determined by cost constraints)

Simulation Study of a Adaptive Design

- 1995 Israel Census Sample of Individuals 15+ (N=753711)
- Population response propensities calculated according to variables: child indicator, income from earnings groups, age groups, sex, number of persons in household and three types of localities
- Based on the probabilities, generate a response indicator as the dependent variable in a logistic regression model using explanatory variables (with interactions)
- Predictions from this model serve as ‘true’ response propensities

Simulation Study of a Adaptive Design

- Overall response rate 69.2%
- Three samples: 1:50 sample (sample size of 15074), 1:100 sample (sample size of 7537) and 1:200 sample (sample size of 3769)

R-indicators and Coefficient of Variation (with confidence intervals) for the three sample before and after targeted follow-up assuming 50% response

Sample	Response Rate Original	Original Sample		Response Rate Final	With Targeted Follow-up Non-response (Response Rate 50%)	
		R-indicator	Coefficient of variation		R-indicator	Coefficient of variation
1:50 n=15,074	69.6%	0.871 (0.857-0.886)	0.093 (0.082-0.103)	72.3%	0.904 (0.890-0.919)	0.066 (0.056-0.076)
1:100 n=7,537	69.0%	0.854 (0.834-0.875)	0.105 (0.090-0.120)	71.9%	0.886 (0.866-0.907)	0.079 (0.065-0.094)
1:200 n=3,769	70.1%	0.843 (0.813-0.872)	0.112 (0.091-0.133)	72.8%	0.871 (0.842-0.901)	0.088 (0.068-0.109)

Simulation Study of a Adaptive Design

- Build a profile of individuals to target for adaptive design:

  Variable level Partial R-indicators ('' for 5% significant level) for the samples before and after targeted follow-up assuming 50% response rate*

Variable	Original Sample			With Targeted Follow-up (50% Response rate)		
	1:50	1:100	1:200	1:50	1:100	1:200
Unconditional Variable Partial R-indicators						
Persons in HH	0.032*	0.040*	0.051*	0.027*	0.034*	0.048*
Type of Locality	0.011*	0.014*	0.020*	0.010*	0.011	0.019*
Age Group	0.047*	0.054*	0.055*	0.033*	0.035*	0.039*
Children in HH	0.030*	0.033*	0.036*	0.014*	0.017*	0.021*
Income Group	0.018*	0.031*	0.027*	0.011*	0.026*	0.021*
Sex	0.019*	0.012*	0.013	0.010*	0.018*	0.015*
Conditional Variable Partial R-indicators						
Persons in HH	0.029*	0.033*	0.047*	0.029*	0.032*	0.046*
Type of Locality	0.011*	0.013*	0.021*	0.009*	0.010*	0.020*
Age Group	0.046*	0.050*	0.052*	0.037*	0.037*	0.041*
Children in HH	0.017*	0.017*	0.014	0.008*	0.009	0.004
Income Group	0.005	0.022*	0.016*	0.007	0.022*	0.017*
Sex	0.017*	0.011*	0.010	0.009*	0.017*	0.016*

- Table distinguishes the variables: number of persons in the household, type of locality, age group, child indicator and sex

Simulation Study of a Adaptive Design

- Inspect categories of these variables and determine which categories have a significant negative unconditional partial R-indicator (under-represented) and a significant conditional value

Category level (Unconditional and Conditional) Partial R-indicators ('' 5% significance) for the 1:50 original sample*

Variable	Category	Uncond. Partial	Cond. Partial	Variable	Category	Uncond. Partial	Cond. Partial
Children in HH	None	-0.015*	0.012*	Locality Type	Type 1	-0.010*	0.009*
	1+	0.026*	0.013*		Type 2	0.005*	0.004*
Age Group	15-17	0.020*	0.005*		Type 3	0.001	0.005
	18-21	-0.017*	0.021*	Sex	Male	-0.014*	0.013*
	22-24	-0.015*	0.013*		Female	0.013*	0.012*
	25-34	-0.016*	0.011*	Persons in HH	1	-0.007	0.012*
	35-44	-0.005	0.011*		2	-0.015*	0.008*
	45-54	0.005	0.007*		3	0.007	0.007*
	55-64	0.002	0.009*		4	0.025*	0.022*
	65-74	0.018*	0.020*		5	-0.003	0.008*
	75+	0.026*	0.026*		6+	-0.005	0.008*

- Categories: males, age between 18 and 34, 2-person households, no children in hh and the first type of locality.

Simulation Study of a Adaptive Design

- 32 strata defined by cross-classifying the following sets: {males, females}×{aged 18-34, other}×{2 persons, other}×{no children, has children}×{locality type 1, other}.
- Unconditional categorical partial R-indicators calculated for new strata and sorted by their p-value
- For the 1:50 sample, the high and significant p-values on the under-represented strata were obtained for the following sets in order of significance: {males, 18-34, 2 persons, no children, type 1}; {males, 18-34, 2 persons, no children, not type 1}; {males, 18-34, not 2 persons, no children, type 1}; {males, 18-34, not 2 persons, no children, not type 1}.
- The number of non-respondents to target for follow-up in these four strata are 838 (5.6%), 421 (5.6%) and 188 (5.0%) for the 1:50, 1:100 and 1:200 samples respectively

Simulation Study of a Adaptive Design

- Assume 50% response rate on targeted follow-up:
- Overall response rate increased by 2.7% with significant increase in R-indicator and decrease in the CV assuming a 50% response rate for follow-up
- Reduction in variable level partial R-indicators, although some collinearity remained
- In the 1:200 sample size, sex has gone from non-significant to significant, following the targeted response for both conditional and unconditional partial R-indicators.
- For the categorical level partial R-indicators there is an overall reduction following the targeted response with many categories non-significant
- Conclusion: even with a small increase of response rate, albeit targeted to those non-respondents contributing to the lack of representativity, we are able to improve the representativeness of the data

2011 Crime Victimisation survey

- A sample of 8800 persons randomly assigned to one of four sequential mode strategies: Web followed by face-to-face, mail followed by face-to-face, telephone followed by face-to-face and face-to-face followed by face-to-face.
- Both respondents and nonrespondents to the first phase (Web, mail, telephone or face-to-face) received the second phase (face-to-face) in which the first key sections of the CVS questionnaire were repeated
- Consider two strategies: Web to face-to-face and mail to face-to-face.
- Evaluation of representativeness and construction of strata for the second phase is done using six socio-demographic registry variables: gender (male, female), age (15-25, 25-35, …, 65-75, 75+), employment (yes, no), urbanization of residence (not, little, moderate, strong, very strong), income in Euro's (<3K, 3-5K, 5-10K, 10-15K, …, 25-30K, >30K), ethnicity (native, western non-native, non-western non-native), and registered landline phone number (yes, no

2011 Crime Victimisation survey

Response rate, R-indicator, coefficient of variation and costs for various strategies in the 2011 CVS experiment. Standard error approximations are given within brackets. Costs are given in 1000′s of Web sample unit costs.

Strategy	Response rate	R-indicator	CV	Cost
Web	28.7% (1.0%)	0.806 (0.019)	0.368 (0.034)	2.2
Web → F2F	57.9% (1.1%)	0.829 (0.022)	0.168 (0.019)	49.1
Web scenario 1	39.7% (1.0%)	0.808 (0.021)	0.267 (0.026)	20.0
Web scenario 2	43.6% (1.1%)	0.846 (0.021)	0.206 (0.025)	29.1
Mail	49.0% (1.1%)	0.738 (0.020)	0.283 (0.020)	8.8
Mail → F2F	66.0% (1.0%)	0.812 (0.021)	0.157 (0.016)	42.3
Mail scenario 1	54.1% (1.1%)	0.855 (0.022)	0.159 (0.020)	18.7
Mail scenario 2	59.5% (1.1%)	0.878 (0.022)	0.129 (0.019)	26.8
F2F → F2F	67.9% (1.0%)	0.801 (0.021)	0.160 (0.015)	91.3

- Quality Target : F2F-F2F with CV of 0.160
- Costs for approaching one CVS sample person through mail is approximately four times higher than through Web and the costs for F2F are approximately 30 times higher

2011 Crime Victimisation survey

- Suppose that the available budget is one third of the expensive F2F to F2F strategy, i.e. 30.4. This implies there is budget to allocate 940 cases to F2F after a Web first phase and 720 cases after a mail first phase.

*Variable-level unconditional and conditional partial R-indicators for various strategies in the 2011 CVS experiment. (p-value: * = below 0.1%, † = below 1% , # = below 5%).*

		Unconditional		Conditional	
		Phase 1	Phase 1 and 2	Phase 1	Phase 1 and 2
Gender	Mail	0.024 #	0.014	0.040 *	0.024 #
	Web	0.020 #	0.003	0.001	0.007
Ethnicity	Mail	0.077 *	0.058 *	0.043 *	0.033 *
	Web	0.039 *	0.047 *	0.022 †	0.021 #
Income	Mail	0.067 *	0.056 *	0.056 *	0.047 *
	Web	0.077 *	0.046 *	0.053 *	0.032 †
Urbanization	Mail	0.026 #	0.026 #	0.014	0.015
	Web	0.015	0.053 *	0.014	0.034 *
Age	Mail	0.087 *	0.051 *	0.064 *	0.037 *
	Web	0.061 *	0.036 *	0.041 *	0.022 #
Phone	Mail	0.038 *	0.027 †	0.016	0.011
	Web	0.029 *	0.046 *	0.016	0.026 †

2011 Crime Victimisation survey

- First scenario (one off survey and no time to perform a full F2F phase 2) - four categories turned up for both Web (income groups 10-15K and 15-20K, age group >75 years and non-western non-natives) and for mail (males, age groups 15-25 years and 25-35 years and non-western non-natives)
- From these categories stratifications were formed and strata with significant negative unconditional values were selected for follow-up; 594 cases for Web and 329 for mail.
- Second scenario (continuous survey with full F2F phase 2) - four categories for Web (income group >30K, natives, persons with a registered phone and persons living in little or non-urbanized areas) and five categories for mail (income group >30K, natives, persons with a registered phone and age groups 55-65 years and 65-75 years)
- From these categories stratifications were formed and strata that did not have significant negative unconditional values were deselected for follow-up; leaving a total of 896 for Web and 601 for mail for follow-up

2011 Crime Victimisation survey

- Under scenario 1, for Web the second phase does not improve the R-indicator while for mail the second phase leads to an enormous increase in the R-indicator.
- The coefficient of variation for mail has become similar to the target from the F2F to F2F design while for Web it is still higher
- Under scenario 2, the R-indicator increases for both Web and mail and are significantly higher than for strategy F2F to F2F
- Because of the lower response rate the coefficient of variation for Web is still higher than the target but for mail it is significantly lower

2011 Crime Victimisation survey

*Unadjusted response means for five CVS survey variables and coefficient of variation for designs with a Web or mail first phase. (p-value for test against phase 1 response only: * = below 0.1%, † = below 1% , # = below 5%).*

Web	Phase 1	Phase 1 and 2	Scenario 1	Scenario 2
Coefficient of variation	0.368	0.168	0.267	0.206
# victimizations per 100	26.6	30.3	26.6	30.1
% victimized	8.1	10.7 †	8.9	10.8 †
Nuisance scale	1.3	1.3	1.4	1.4
% unsafe	25.5	25.4	25.6	26.9
% not satisfied police	45.3	47.1	46.9	47.0

Mail	Phase 1	Phase 1 and 2	Scenario 1	Scenario 2
Coefficient of variation	0.283	0.157	0.159	0.129
# victimizations per 100	17.6	22.3 #	23.4 †	22.4 #
% victimized	8.8	10.0 #	10.6 #	10.3 #
Nuisance scale	1.2	1.3	1.3	1.3
% unsafe	27.1	25.4	26.4	26.2
% not satisfied police	47.8	47.5	48.0	48.1

2011 Crime Victimisation survey

- Victimization variables show significant differences against a Web only or a mail only design at the 5% level, the other variables do not
- Neighbourhood nuisance scale seems to be robust against changes in design.
- There is some indication that decreases in the coefficient of variation coincide with significant changes in the victimization variables; the only design where it did not change significantly, scenario 1 for Web, still had a relatively high coefficient of variation.
- Remarkably, the number of reported victimizations in designs with a Web first phase is a lot higher than those with a mail first phase, although percentages victimized are similar.

2011 Crime Victimisation survey

- Application confirms that building adaptive survey designs based on response to a first phase can be risky
- For mail the second phase allocation turned out right and all indicators improved, but for Web hardly any improvement was found
- The F2F second phase helped raise response rates of some strata but was counterproductive on other strata
- This risk reflects the lack of knowledge about the efficacy of the second phase which is included in the scenario where both phases have been conducted first
- The application shows that it may be fruitful to perform a first pilot wave in which some investment is made in learning if and how a second phase improves response
- After this wave the design can be optimized for subsequent waves and statistics for the first wave (and obviously future waves) can be based on the optimized design

Accessing Optimal Strategies to Reduce Non-response in Longitudinal Studies

Ian Plewis and Natalie Shlomo (revisions)

Paper includes:

- Introduction and motivation
- Sample maintenance strategies and their costs and potential benefits
- Response propensity models and two sets of measures derived from them, one set based on the variability of the predicted probabilities of responding (R-indicators) and the other on receiver operating characteristic (ROC) curves
- The UK Millennium Cohort Study, with illustrative results
- Implications of the findings and the challenges they present to some of the assumptions made about how best to conduct longitudinal studies

Introduction

- In this description, I cover only parts of the paper related to R-indicators, although research related to ROC curves may be an interesting area to pursue for targeting non-respondents in adaptive survey designs
- Longitudinal studies need to retain sample members over time to remain representative of target population

  - How effective are strategies to retain sample members in a longitudinal study?

  - Which sample members should be the targets of intervention to improve the quality of response?

- Use R-indicators to partially address these questions

UK Millennium Cohort Study

- Wave one sample: 18,818 babies in 18,552 families born in the UK over a 12 month period during the years 2000-2001 and living in selected UK electoral wards at age 9 months
- Sample frame: child birth register
- Boost samples in areas with high proportions of Black and Asian Families, disadvantaged areas and three smaller UK countries over represented
- Design weights with respect to the sample size range from 2.0 (England advantaged stratum) to 0.23 (Wales disadvantaged stratum)
- First 4 waves: 9 months, 3, 5 and 7 years old

UK Millennium Cohort Study

- Face to face interviewing, partners interviewed where possible, data collected from cohort members and siblings
- Standard practice – reissue all eligible cases at wave t conditional on their being in the observed sample at wave 1
- Cases ineligible: emigration, institutional care or child death
- Exceptions: hard refusals not reissued and majority of eligible cases that were unproductive at waves 2 and 3 not reissued at wave 4
- Overall wave 1 response rate was 72%:

England – Advantaged 76%, Disadvantaged 72%, Ethnic 66%
Wales – Advantaged 79%, Disadvantaged 72%
Scotland – Advantaged 77%, Disadvantaged 74%
N. Ireland– Advantaged 68%, Disadvantaged 64%

Response Propensity Model

- Explanatory variables predictors of non-response taken at wave 1 (so representativeness is with respect to wave 1 and not the target population):

Family income (8), Ethnic group of cohort child (6), Accommodation type (2), Tenure (3), Mother’s age (2), Mother’s education (7), Child breast fed (2), Mother long term illness (2), Family status (2), Mother voted in last election (2), Mother gave consent to record linkage (2), Provided a stable address at wave 1 (2), Change of address between waves 1 and 2 (2), Interactions with Tenure and Accommodation type

- Take into account sample design (disproportionate stratification and clustering) using SAS Proc SurveyLogistic
- Use predicted probabilities to estimate R-indicators and partial R-indicators

Re-issuing Strategies

Strategy	Explanation
Standard Practice	
S2	reissue all eligible cases at wave 2 conditional on being observed at wave 1
S3	reissue all eligible cases at wave 3 conditional on being observed at wave 1
S4	reissue all eligible cases at wave 4 conditional on being observed at wave 1
Hypothetical strategy - only reissue productive cases from previous waves	
P3.2	Only reissue at wave 3 cases that were productive at wave 2
P4.23	Only reissue at wave 4 cases that were productive at waves 2 and 3
P4.3	Only reissue at wave 4 cases that were productive at wave 3
Hypothetical strategy - of not reissuing refusals from previous waves	
C3.2	Only reissue at wave 3 cases that were not refusals from wave 2
C4.32	Only reissue at wave 4 cases that were not refusals in waves 2 and 3
C4.3	Only reissue at wave 4 cases that were not refusals from wave 3
Hypothetical strategy of not reissuing at wave 4 cases that were not productive at wave 2 but were productive at wave 3	
W4	

Re-issuing Strategies

Row label	Cases lost	Percentage of actual productive sample
S2	n.a.	-
S3	n.a.	-.
P3.2	1,444	9.5%
C3.2	473	3.1%
S4	n.a.	-
P4.23	1,668	12.0%
P4.3	639	4.6%
C4.23	536	3.9%
C4.3	216	1.6%
W4	1,029	7.4%

- Sample size at wave 1 is 18,552 but some cases omitted due to item nonresponse at wave 1
- Sample sizes after omitting ineligible cases: 18,148, 17,990, 17,819 for waves 2 to 4

Results of R-Indicators

Strategies	R-indicator (CI)	Difference from Standard Practice (CI*)
S2	0.781 (0.763 – 0.799)	-
S3	0.794 (0.777 – 0.811)	-
P3.2	0.715 (0.694 – 0.735)	-0.079* (-0.094- -0.065)
C3.2	0.771 (0.752 – 0.789)	-0.023* (-0.031- -0.015)
S4	0.740 (0.720 - 0.760)	-
P4.23	0.666 (0.644 - 0.687)	-0.074* (-0.088 - -0.060)
P4.3	0.715 (0.694 - 0.736)	-0.024* (-0.032 - -0.016)
C4.23	0.721 (0.700 - 0.741)	-0.019* (-0.028 - -0.010)
C4.3	0.735 (0.715 - 0.756)	-0.004 (-0.009 - 0.001)
W4	0.692 (0.671 - 0.712)	-0.048* (-0.060 - -0.036)

* CI of difference calculated by bootstrapping

Results of R-Indicators

- S2 and S3 show slight increase in representativeness with respect to wave 1 as indicated by the increase in the R-indicators.
- S4 shows lower representativeness at wave 4 compared to the first two waves using standard reissuing practice
- P show that representativeness falls if only productive cases from previous waves are reissued and R-indicators are all significantly lower than for the standard practice in each wave
- The R-indicator for P4.3 is significantly higher to the R-indicator P4.23 ($p<0.001$) and is closer to the R-indicator of the standard practice S4

Results of R-Indicators

- Comparing C3 to S3 suggests that representativeness is less compromised if refusals from previous waves are not reissued (R-indicator more similar to standard practice)
- Comparing strategies of cases that refused just at wave three not being reissued at wave four (C4.3) to those that refused at either wave two or wave three (C4.23), there is a significant increase in representativeness for C4.3 ($p<0.001$)
- C4.3 similar representativeness compared to standard practice in S4 with a non-significant difference in R-indicator
- Strategy C4.3 had the lowest number of dropped cases
- W4 shows that representativeness is significantly reduced if wave non-respondents are not reissued

Results of Partial R-Indicators – Education Qualifications

Strategies	Unconditional Partial R-indicator (CI)	Difference from Standard Practice (CI)
S2	0.060 (0.052 – 0.067)	0.015 (0.007 – 0.023)
S3	0.060 (0.053 – 0.067)	
P3.2	0.083 (0.075 – 0.092)	0.024* (0.020 – 0.027)
C3.2	0.067 (0.059 – 0.074)	0.007* (0.006 – 0.008)
S4	0.077 (0.069 – 0.085)	
P4.23	0.100 (0.091 – 0.109)	0.023* (0.020 – 0.026)
P4.3	0.086 (0.078 – 0.094)	0.009* (0.008 – 0.010)
C4.23	0.083 (0.075 – 0.092)	0.007* (0.006 – 0.008)
C4.3	0.078 (0.070 – 0.086)	0.002* (0.002 – 0.002)
W4	0.091 (0.082 – 0.100)	0.023 (0.014 – 0.031)

Results of Partial R-Indicators

- Unconditional partial R-indicators significantly different from zero so highest educational qualification contributes to lack of representativity and highest for P4.32
- Other variables with high unconditional partial R-indicator (not shown): Family income group (8), Family status (4), Tenure (3), Ethnic group (6),
- Conditional partial R-indicators significantly different from zero so educational qualification continues to contribute to lack of representativity conditional on other variables
- Unconditional partial R-indicators larger than conditional partial R-indicators suggesting that the impact of each variable is reduced when controlling for other variables (multicollinearity of auxiliary variables)

Targeting Data Collection

- Considering the reissuing strategy at wave 4, we might decide not to reissue the cases that were unproductive at earlier waves (row P4.23) only if they belonged to majority ethnic group or had some educational qualifications

- The estimate of the R-indicator then increases from 0.666 (0.644 - 0.687) to 0.689 (0.669-0.709) with 481 cases lost and so the strategy is nearly as effective as not reissuing unproductive cases just from wave three (P4.3: 0.715 (0.694 - 0.736))

.

Some comments on ROC Curves

- ROC is the plot of P(+| r = 0) against P(+| r = 1), i.e. the true positive fraction (TPF) against the false positive fraction (FPF) for all c; slope of the ROC for any c is

$$P(\hat{\rho}=c\,|\,r=0)\,/\,P(\hat{\rho}=c\,|\,r=1)$$

- AUC (area under the curve) varies from 1 (when the model for predicting response perfectly discriminates between respondents and non-respondents) down to 0.5, the area below the diagonal (when there is no discrimination between the two categories)
- AUC interpreted as the probability of assigning a pair of cases, one respondent and one non-respondent, to their correct categories, bearing in mind that guessing would correspond to a probability of 0.5
- A linear transformation of AUC (= 2*AUC – 1), often referred to as a Gini coefficient, is commonly used

Some comments on ROC Curves

- Optimum threshold - minimise a cost function for overall cost of non-response per case (TC) (Pepe, 2003)

$$TC = C^{+}_{r=0} TPF(1-\hat{\bar{\rho}}_s) + C^{-}_{r=0}(1-TPF)(1-\hat{\bar{\rho}}_s) + C^{+}_{r=1} FPF(\hat{\bar{\rho}}_s)$$

- The first cost term is the cost of intervening when an intervention is indicated (i.e. the predicted probability of non-response is above the chosen threshold *c*) and when the case is indeed a non-respondent (*r=0*); the second and third cost terms are misclassification costs arising from failing to intervene when the case is a non-respondent and intervening when the case is a respondent; $1-\hat{\bar{\rho}}_s$ is the prevalence of non-response

Some comments on ROC Curves

- Optimum cut point on the response propensity scale determined by minimising *TC* with respect to the threshold
- Implies that the slope of the ROC curve at the optimum threshold is *O*F* where *O* is the odds of being a respondent (and greater than one) and

  $F = C^{+}_{r=1} / (C^{-}_{r=0} - C^{+}_{r=0})$, i.e. *F* is the ratio of the actual cost of intervening when there would have been a response without the intervention (the false positives) to the opportunity cost of failing to intervene for a non-respondent (the false negatives) minus the actual cost of intervening when the prediction to be a non-respondent is correct (the true positives) and the denominator is assumed to be positive
- Based on using this cut-off we will examine how the R-indicators behave (still to do for the revisions to the paper)

Conclusions and Future Work

- Not re-issuing refusals (C) has less of an impact on representativity with respect to wave 1 than re-issuing productive cases (P) since many continue to be refusals at subsequent waves
- Indication that only re-issuing productive cases (P) impacts on representativity with respect to wave 1 but can be mitigated by targeting refusals based on profiles from R-indicators
- All response models and R-indicators (SE's and bias corrections) carried out under the complex survey design
- Further work: Include paradata such as neighbourhood observations ; compare the use of optimal thresholds from the ROC curve below which interventions should be carried out to R-indicators and test for consistencies in the theory

Indicators for Representative Response Based on Population Totals

Annamaria Bianchi, Natalie Shlomo, Barry Schouten, Damiao DaSilva, Chris Skinner (paper to be submitted)

Paper includes:

- Introduction and motivation
- Definition and estimation of population based R-indicators
- Theoretical properties, SE and bias corrections
- Simulation Study to evaluate properties
- Real data application from the business revenue data in tax register of Statistics Netherlands
- Future Work

Indicators for Representative Response Based on Population Totals

Annamaria Bianchi, Natalie Shlomo, Barry Schouten, Damiao DaSilva, Chris Skinner (paper to be submitted)

Research Questions:

How to extend sample-based R-indicators to population-based R-indicators?

What are the statistical properties of population-based R-indicators?

Are the population-based R-indicators practicable in real survey settings?

Some Caveats of Population-based R-indicators

- Can be applied to any setting with missing data on variables of interest and (almost) complete auxiliary data, such as register data
- Population-based R-indicators have weaker conclusions about the nature of response compared to sample-based R-indicators as they are less accurate because they have to discern sampling variation from response variation, i.e. population-based indicators not very useful for small surveys
- Risk of measurement errors when comparing representativeness of survey questions to population statistics, i.e. survey questions must have the same definitions and classifications as population tables (possible to draw erroneous conclusions about the representativeness of response if population statistics are biased)

Some Caveats of Population-based R-indicators

- Options to improve representativeness during data collection for adaptive designs are limited since no individual auxiliary information for non-respondents is available
- Assessments of representativeness may still be useful in the design of advance and reminder letters, in interviewer training and in paradata collection
- Future work to consider hybrid settings where the R-indicator is based on both linked data and population tables

Types of information

Sample-based information: microdata for respondents and non-respondents

Population-based information: microdata for respondents and aggregate data from population

TYPE1 → population cross-products

TYPE2 → population marginal counts

Estimation

Sample-based estimation:

1) Estimate response propensities using logistic/linear regression on sample

$$\hat{\rho}_i = \frac{\exp\left(x_i^T \hat{\beta}\right)}{1+\exp\left(x_i^T \hat{\beta}\right)}$$

2) Replace population means by design-weighted sample means

$$\hat{\rho}_{i,OLS} = x_i^T \left(\sum_s d_i x_i x_i^T\right)^{-1} \sum_r d_i x_i, \qquad i \in s$$

Population-based estimation:

1) Estimate response propensities (only for respondents) by

- TYPE1: $\hat{\rho}_{i,T1} = x_i^T \left(\sum_U x_i x_i^T\right)^{-1} \sum_r d_i x_i$
- TYPE2: $\hat{\rho}_{i,T2} = x_i^T \left(N\hat{S}_{xx} + N\overline{x}_U \overline{x}_U^T\right)^{-1} \sum_r d_i x_i, \qquad i \in r$

2) Replace population means by propensity-weighted response means

Population-based R-indicator

Estimators:

$$\hat{R}_{\hat{\rho}} = 1 - 2\hat{S}_{\hat{\rho}},$$

1) $$\hat{S}^2_{\rho} = \frac{1}{N-1}\sum_r d_i \rho^{-1}_{i,T1}\left(\hat{\rho}_{i,T1} - \hat{\bar{\rho}}_r\right)^2$$

2) $$\hat{S}^2_{\hat{\rho}} = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i \in r} d_i \hat{\rho}_{i,T1} - \left(\frac{1}{N}\sum_{i \in r} d_i\right)^2\right\}$$

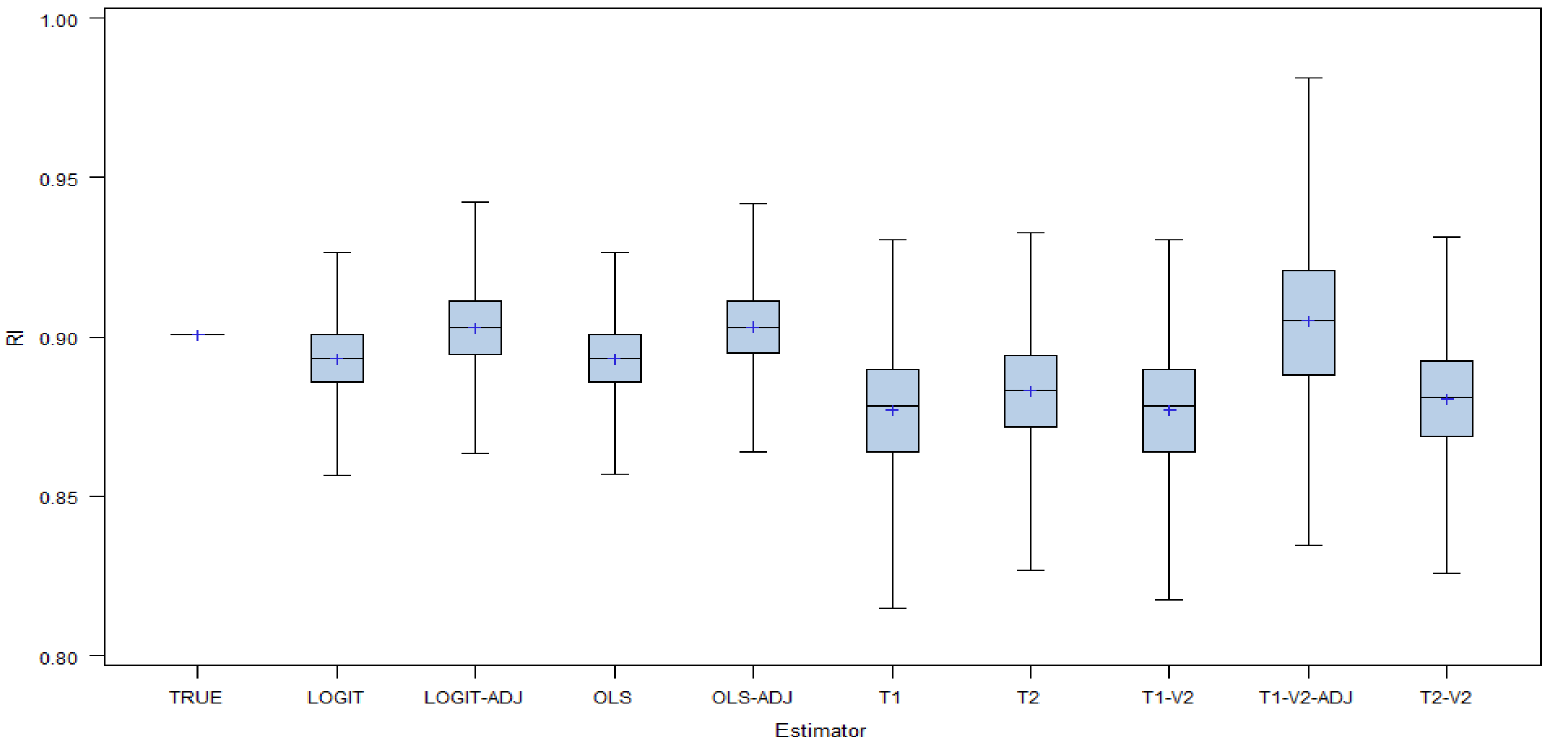
3) $$\hat{S}^2_{\hat{\rho}} = \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i \in r} d_i \hat{\rho}_{i,T1} - \left(\frac{1}{N}\sum_{i \in r} d_i\right)^2\right\} - ADJ$$

Generally, population-based indicators have **ADDITIONAL bias** resulting from the fact that they cannot completely discern sampling variation from response variation

Simulation Study: Sampling properties of population-based R-indicators

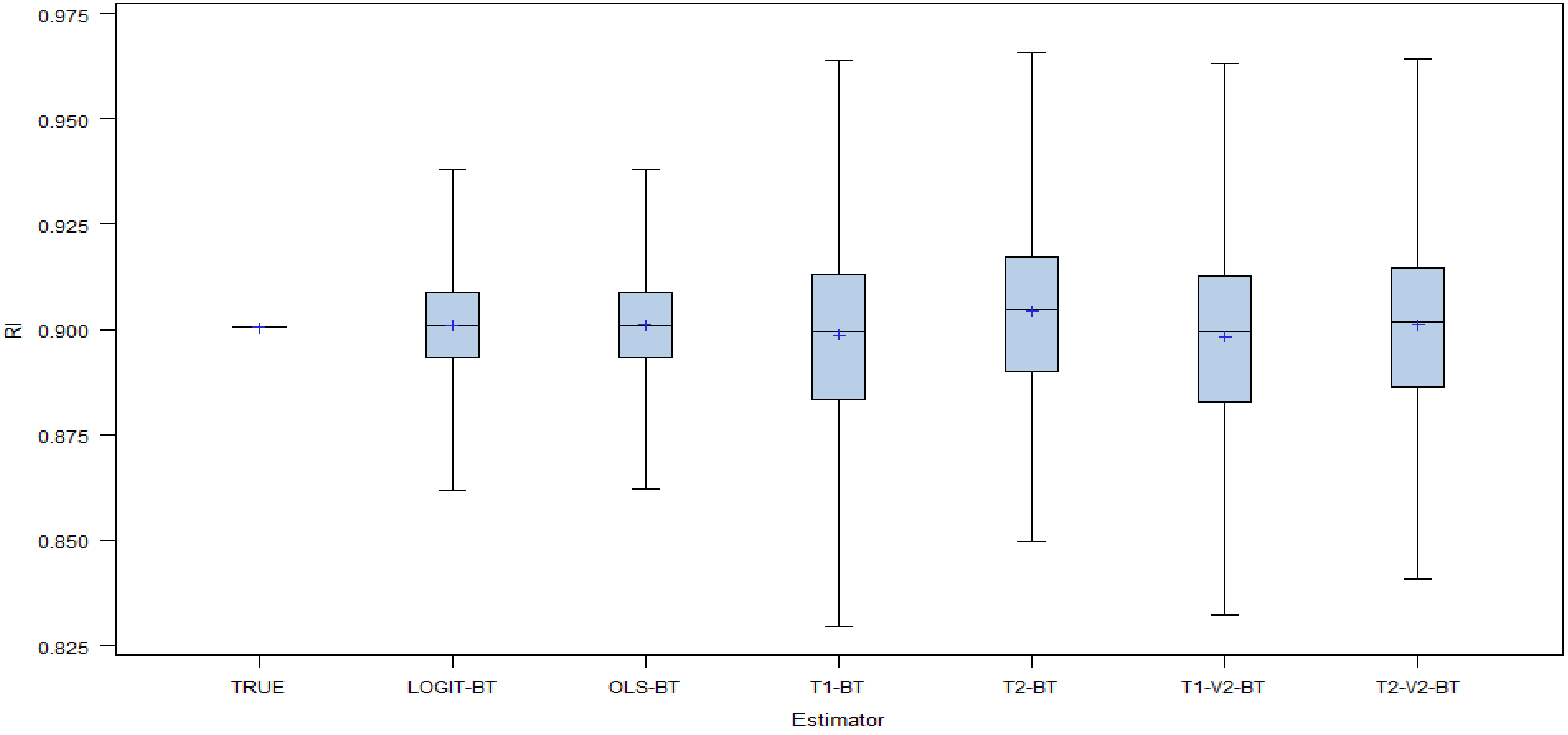
- 500 simple random samples drawn from 1995 Israel Census data on 322,411 households
- Sampling fraction 1:50,1:100 and 1:200
- True population response propensities defined under four different scenarios of response rates
- For each random sample selected, 500 bootstap samples were drawn with replacement to estimate bias and variance
- Analytical expression of bia correction evaluated (Damiao DaSilva)
- Two models: exact model on all explanatory variables and reduced model

R-indicator based on 500 samples (1:50 samples):



- Population-based R-indicators (T1, T2, T1-V2, T2-V2) have greater bias than sample-based counterparts
- Analytical bias adjustment (T1-V2-ADJ) (Da Silva ,et al.) performs well but more variation

Bootstrap bias adjusted R-indicators based on 500 simulations (1:50 samples):



- Bootstrap adjustment can reduce bias
- Variability increases

Preliminary Conclusions

- Population-based R-indicators have larger (downward) bias and larger variance than sample-based indicators and hence allow for weaker conclusions
- In unadjusted form, Type 2 better than Type 1 for the pop-based R-indicator
- Analytical expression for bias correction for Type 1 performs well however the Type 2 bias correction performs less well and generally overcompensates for the bias
- Bootstrap performs well for the bias adjustment and variance estimation
- Sample based R-indicators not influenced by response rates, but they affect pop-based R-indicators with better performance for low response rates (less chance of estimating propensities greater than 1)

Preliminary Conclusions

- In addition, Type 1 estimator does not account for sampling variability since we ‘plug in’ poulation covariance matrix; Type 2 uses response covariance matrix and hence accounts for sampling variation
- Larger RRMSE for reduced model compared to full model but bias adjusted R-indicators similar
- Also looked at combinations of Type 1 and Type 2: composite population and response only covariance matrix; composite population and response only response propensities, where weight ing factor is a function of the reponse rate

Future Work

- Finish current simulations and complete paper
- Partial pop-based R-indicators? Are they of any use?
- Combining population auxiliary information and paradata known for the sample into one response model
- New approach to estimation of population –based R-indicators using a GLM and EM algorithm

Where do we go from HERE?