



The Leverhulme Trust

Network Grant

MANCHESTER
1824

The University of Manchester

Bayesian Adaptive Survey Design Network

Ref: IN-2014-046

Annual Report

January 12th, 2015 – January 11th, 2016

Natalie Shlomo

University of Manchester

**The Leverhulme Trust International Network Grant: Bayesian Adaptive
Survey Designs (BADEN) Ref: IN-2014-046**

Annual Progress Report January 12th, 2015 - January 11th, 2016

Natalie Shlomo, University of Manchester

The activities for the first year of the Leverhulme Trust International Network Grant: Bayesian Adaptive Survey Design (BADEN) (Ref: IN-2014-046) went exactly according to the original work plan as set out in the grant application. Below are details of the network events and meetings with a brief summary from the start of the project on January 12th, 2015 to January 11th, 2016.

The website for the project is hosted by the University of Manchester and a URL was purchased: www.badennetwork.com

Network activities, events and meetings

- The **Kick-start meeting** was held at the United States Census Bureau in Washington DC (26-27 February 2015).

At this meeting, each partner provided a presentation on their institutional state-of-the-art (see: <http://www.cmist.manchester.ac.uk/research/projects/baden/documents/> for summaries). In addition, a research agenda for the network was widely discussed and was finalized in April 2015. The research agenda document is in Appendix 1.

- A **Bi-lateral meeting** between CBS Netherlands, University of Manchester and Statistics Sweden was held in the Netherlands (28 - 29 May 2015).

At this meeting, theoretical topics around the Bayesian framework for adaptive survey designs were widely discussed and clarified. The framework is decomposed into three separate components: modelling design parameters (response propensities, cost functions); learning and updating posterior distributions of design parameters based on information coming in from the current survey; optimisation of quality and cost criteria that also have probability distributions. In addition, the different types of adaptive survey design strategies handled by partner institutions were classified: (1) predetermined strategies (eg. recruitment letters, mixing data collection modes); (2) follow-up strategies (eg. interventions during data collection, stopping criteria); (3) hybrid of both strategies. Other discussions related to partnerships and PhD students, joint research and possible sharing of data. The first theoretical paper is led by researchers at CBS Netherlands.

- The **Intermediary meeting** with all partners was held at Statistics Sweden in Stockholm (10 – 11 September 2015)

At the meeting, the US Census Bureau partner presented an overview of a study comparing the many indicators presented in the literature that have been used for adaptive survey designs. Initial results show that many of these indicators measure the same thing, eg. variation in response propensities, which is indicative of a lack of representativeness. Each of the partners then presented their institutional case study for the elaboration of the Bayesian framework for adaptive survey designs. These case studies will form the basis of the work within the network. Discussions were held about collaborations, possibility to share data and proposed papers for the first network dissemination at the 71st Annual American Association of Public Opinion Research (AAPOR) Conference in May. CBS Netherlands partners provided an overview of the first theoretical paper on Bayesian Adaptive Survey Designs which was widely discussed. Finally, all partners discussed planning for the 4th International Workshop: Advances in Adaptive and Responsive Survey Design to be held at the University of Manchester, (9 - 10 November 2015).

- October, 2015: The PI successfully proposed an **invited panel session** to the 71st Annual American Association of Public Opinion Research (AAPOR) conference, Austin Texas, May 12-15, 2016. The proposal is in Appendix 2.
- **4th International Workshop:** Advances in Adaptive and Responsive Survey Design at the University of Manchester (9 - 10 November 2015)

There were 40 participants to the international workshop from the following countries: United States, Italy, the Netherlands, United Kingdom, Germany, Canada, Norway, Sweden and Australia. The Leverhulme International Network grant funded two invited speakers: Prof. Rod Little from the University of Michigan and Prof. Carl Sarndal from Stockholm University. Additional funds were obtained through the ESRC National Centre for Research Methods to fund Prof. Roger Tourangeau from Westat, United States. In addition, Prof. Peter Lynn from the United Kingdom was also funded from internal funding. There is no doubt that the participation of these leading and prominent Professors greatly enhanced the quality of the discussions and feedback on the papers that were presented at the workshop.

The workshop was held at the University of Manchester, Manchester Meeting Place Conference Hall, and was a highly successful event. The workshop agenda is in Appendix 3.

The BADEN network presented the following papers at the workshop:

Organization	Speaker	Title of Presentation
CBS Netherlands	Barry Schouten	Nonresponse Adjustment by Design?
	Nino Mushkudiani	Assessing the Impact of Inaccurate Survey Design Parameters on the Performance of Adaptive Survey Designs
	Lisette Bruin	A Bayesian Analysis of Survey Design Parameter
University of Michigan	James Wagner	Prior Specification for Bayesian Estimation of Response Propensity Models During Data Collection
Statistics Sweden	Anton Johansson	Modelling length and final outcome of call sequences in the Swedish Labour Force Survey
US Census Bureau	Stephen Kaputa	Investigating Nonresponse Subsampling in an Establishment Survey Through Embedded Experiments
	Stephanie Coffee	Evaluating Some Indicators for Nonresponse Bias Under a Variety of Design Conditions
	Peter Miller	The U.S. Census Adaptive Design Research Agenda
	Michael Thieme	Draft paper title: Intra-Application Distributed Processing: <i>Computing Speed and Agility for Adaptive Design</i>
	Gina Walejko	Prioritizing cases to increase sample representativeness in the Survey of Income and Program Participation
University of Manchester	Alexandru Cernat	Impact of mixed modes on measurement errors and estimates of change in panel data
University of Southampton	Jamie Moore	Are trajectories of dataset representativeness during survey data collection generalizable? Evidence from the 2011 Census Non-Response Link Study
RTI International	Dan Pratt	Examination of interventions during data collection to increase response and sample representativeness: a field test experiment and simulation

Journal Articles

- Schouten, B. and Shlomo, N. (2015) Selecting Adaptive Survey Design Strata with Partial R-indicators, *International Statistical Review*, Article first published online: 11 DEC 2015 | DOI: 10.1111/insr.12159
- Schouten B., Cobben, F., Lundquist, P. and Wagner, J. (Forthcoming) Does Balancing Survey Response Reduce Nonresponse Bias? *Journal of the Royal Statistical Society Series A*.
- Nishimura, R., Wagner, J. and Elliott, M. (Forthcoming) Alternative Indicators for the Risk of Nonresponse Bias: A Simulation Study. *International Statistical Review*.
- Laflamme, F. and Wagner, J. (Forthcoming) Responsive and Adaptive Collection Designs. In Sage *Handbook of Survey Methodology*, eds. Christof Wolf, Dominique Joye, Tom W. Smith and Yang-Chih Fu.
- Moore, J., Durrant, G.B. and Smith, P.W.F. (2015) Dataset representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice, submitted to the *Journal of the Royal Statistical Society, Series A*.
- Correa, S., **Durrant, G.B.** and Smith, P.W. (2015) Assessing Nonresponse Bias using Call Record Data with Applications to a Longitudinal Study, submitted to *Public Opinion Quarterly (Special issue)*.
- Plewis, I. and Shlomo, N. Statistical Guidance on Optimal Strategies to Reduce Non-response in Longitudinal Studies. Under Revisions and to be submitted to the Journal of Official Statistics.
- Bianchi, A., Shlomo, N. Schouten, B., Da Silva, D. and Skinner, C. (2015) Estimation of response propensities and Indicators of Representative Response Using Population-Level Information (Under Revisions and to be submitted to the Annals of Applied Statistics).

Conference Presentations (excluding the papers presented at the Adaptive Survey Design Workshop, November 2015)

- Sanders, H. L., Wagner J., McCarthy, J, Qi, J. and Kreuter, F. (2015) Reducing Bias and Sampling Error: Using Simulation to Identify Effective Adaptive Design Strategies for the Crops Agricultural Production Survey. Paper presented at the *2015 International Total Survey Error Conference*, Baltimore, September 2015.
- Lundquist, P. and Sarndal, C.E. (2015) Design, Data Collection and Estimation in an Age of High Survey Nonresponse. Paper presented at the *International Statistical Institute 60th World Statistics Conference*, Rio, July 2015.
- Moore, J., Durrant, G.B, Smith, P.W. (2015) Are trajectories of dataset representativeness during survey data collection generalisable? Evidence from the 2011 Census Non-Response Link Study. Paper presented at the *International Statistical Institute 60th World Statistics Conference*, Rio, July 2015.
- Coffey, S., Reist, B. and Zotti, A. (2015) Static Adaptive Design in the NSCG: Results of a Targeted Incentive Timing Study. Paper presented at the American Statistical Association Joint Statistical Meeting, Seattle , August 2015.

Working/Draft Papers

- US Census Bureau:
 - Implementing Multimode Dynamic Adaptive Design at the Census Bureau: An Operational Proof of Concept. To be submitted to Survey Practice.
 - Incentive Targeting Using Static Adaptive Design: Results from an Incentive Timing Experiment. To be submitted to the Journal of Survey Statistics and Methodology.
 - Expanding the Use of Dynamic Design in a Multimode Survey: Larger Sample Sizes and Larger Interventions. To be submitted to Journal of Official Statistics or another appropriate journal.

Appendix 1: Research Agenda

April 28, 2015

BADEN: A research agenda

Stephanie Coffey, Gabriele Durrant, Chandra Erdman, Peter Lundquist, Andy Peytchev, Barry Schouten, Natalie Shlomo and James Wagner

Summary: This note describes the objectives and intended results of the BADEN network. BADEN is funded by the Leverhulme Trust International Network Grant (reference number: IN-2014-046) for the period January 2015 to December 2017. In this 3-year period, a number of joint papers will be produced and the institutions will additionally produce separate papers that are closely linked but are set in their own context. The network will, furthermore, produce common tools in R and SAS and will create a number of test data sets. All products will be made available on the website. Finally, the network will organise two workshops and organise two conference sessions.

Motivation

BADEN is short for Bayesian Adaptive Survey Design Network. The network is funded by the Leverhulme Trust International Network Grant (reference number: IN-2014-046) for a period of three years (2015 – 2017) and consists of the universities of Manchester (network coordinator), Michigan and Southampton, the national statistical institutes of the Netherlands and Sweden, RTI international, and the US Census Bureau. In the following, the institutions are abbreviated to MAN, MICH, SOTON, RTI, SN, SSB and USC. A detailed description of the network can be found in its Leverhulme project proposal. The network goals are the exchange of plans and results within the consortium, the development of theory for a Bayesian approach towards adaptive survey design, the testing and implementation of theory through practical case studies, the communication of the results to the large public through conference sessions and workshops, and the proclamation of adaptive survey design in general. The national statistical institutions in the network all are investigating, planning and testing adaptive survey designs.

Adaptive survey designs employ different strategies or design features to different population strata. The strata are identified by auxiliary data from administrative data and/or

from paradata. The employment of the strategies may take place in the form of interventions during data collection and/or through the optimisation of design in between waves of the survey. The interventions and optimisation are based on estimated design parameters like stratum contact propensities, stratum participation propensities, stratum mode coverage propensities and stratum costs parameters. These parameters are estimated using a combination of expert knowledge, prior data and current data. As a consequence, the estimated parameters are subject to inaccuracy. Furthermore, the parameters will, generally, change gradually over time.

A Bayesian approach towards interventions and optimisation is natural as it is a tractable and convenient way to mix and weight expert knowledge, prior data and current data, to account for the resulting uncertainty in the design parameters, and to allow for gradual change. The Bayesian component to adaptive survey design consists of 1) prior distributions to parameters in models for nonresponse and measurement and 2) posterior distributions used as the basis for decision rules in interventions and optimisation of design.

In the following three sections, we describe the common research agenda, the activities and roles behind the agenda, and a time frame.

A common research agenda

The introduction of Bayesian ideas into adaptive survey design follows four steps:

1. Translation of Bayesian theory in existing literature to the modelling of survey design parameters;
2. Testing and perfecting of modelling of survey design parameters given a number of realistic case studies;
3. Development of Bayesian theory to perform interventions and optimise design;
4. Testing and evaluation of theory on pilots and simulation studies;

The seven BADEN institutions operate in different countries and settings; they cover a wide range of surveys types, from small, one-time only surveys to large, continuous surveys. As a consequence, the implementation of adaptive survey design has subtle differences and viewpoints. These differences are most clear in the optimisation of design features and the choice of interventions; in some settings the focus is more strongly on timing and number of calls, whereas in other settings it is more on mode switches. There are also clear commonalities: 1) the quality and cost objectives are the same, 2) the proxy measures used to describe these objectives are very similar in nature, and 3) the modelling of design parameters shows strong resemblance. In these crucial ingredients to adaptive survey design, efficiency can be reached by joining effort within the network. Steps 1 and 2 are achieved through joint effort of the whole network, while steps 3 and 4 are achieved through joint effort in smaller groups within the network.

We see the following results for the network:

1. The identification of realistic case studies based on real and/or (partially) synthetic data sets;
2. A joint, theoretical paper on the introduction of prior distributions in models for the estimation of response propensities and design parameters, and on corresponding

posterior distributions for these parameters and for proxy measures for nonresponse error;

3. Tools/script in R/SAS/other statistical software for the derivation of posterior distributions;
4. A joint, practical paper evaluating the posterior distributions for the identified case studies;
5. A joint, theoretical paper on the weighting of prior and current data in time, allowing for gradual shifts in design parameters;
6. A separate, theoretical paper on the introduction of measurement error;
7. Separate papers on intervening and optimising designs within a Bayesian setting, including case studies;

Activities and roles

The network organises two workshops (November 2015 in Manchester, and November 2017 in the USA, exact location not yet decided) and two invited sessions (AAPOR 2016 and JSM 2017). Preliminary ideas and findings for results 1 and 6 will be presented at the first workshop, and the case studies underlying result 4 will form the invited session for AAPOR 2016. The papers under result 7 will form the invited session for JSM 2017. At the second workshop the network will present an evaluation of the case studies and a view on the future.

During the kick-off four data sets were proposed: the US Consumer Expenditure survey, the US Survey of Income and Program Participation, the European Social Survey, a synthetic data set by RTI, the UK Understanding Society and a Dutch multi-mode data set; In the coming months, these data sets will be evaluated for their utility and possibly other data sets will be added. It seems natural that around each data set a case study be created and a small team formed to conduct the case study and present results at the AAPOR. Manchester and Southampton may join one or more team.

Concerning the papers: During the kick-off four institutions explicitly expressed interest in co-authoring the paper under result 2 and producing the corresponding code under result 3: MICH, RTI, SN, USC. The paper under result 4 will be a collective effort. The papers under results 5 and 7 are in the more distant future and co-authoring of these papers will be decided later. For SN a paper under result 6 has some urgency and it will be initiated at the same time as the paper under result 2. Other institutions may decide to co-author this paper at a later stage.

Time frame

2015:

- February: Kick-off meeting;
- May: Research agenda document and website are completed;
- September: Intermediary BADEN meeting about data sets (7 – 8 September), working document for result 2, AAPOR session and 3rd international workshop;
- October: AAPOR session submitted;
- October: First version of paper result 2 (and possibly also result 6) for discussion at 3rd workshop;
- November: 3rd international workshop and meeting BADEN (9 – 10 November);

2016:

- March: Results 2 and 6 - final versions of papers about prior-posteriors for design parameters and tools in SAS/R to compute posterior distributions;
- May: AAPOR session (12 - 15 May in Austin TX) and meeting BADEN about a joint experimental study and follow-up papers;
- September: Result 4 paper about case studies;
- October: JSM session about optimization of designs submitted;
- November: Intermediary BADEN meeting about follow-up papers (results 5 and 7), a joint experimental study, the JSM session, and other topics that have emerged;

2017:

- March: Draft papers for JSM session (result 5 and 7);
- July: JSM session (29 July – 3 August 2017 in Baltimore MD) and BADEN meeting about papers and 4th international workshop;
- September: Final versions of separate papers in JSM session (result 5 and 7);
- November: 4th international workshop and final BADEN meeting;

Appendix 2:

Session Proposal for the 71st Annual American Association of Public Opinion Research (AAPOR) Conference, Austin Texas, May 12-15, 2016

Proposer: Natalie Shlomo

Title: Advances in adaptive survey designs: developing a Bayesian perspective

Justification and description: Adaptive survey designs employ different strategies or design features to different population strata. The strata are identified by auxiliary data from administrative data and/or from paradata. The employment of the strategies may take place in the form of interventions during data collection and/or through the optimization of design in between waves of the survey. The interventions and optimization are based on estimated design parameters such as stratum contact propensities, stratum participation propensities, stratum mode coverage propensities and stratum costs parameters. These parameters are estimated using a combination of expert knowledge, prior data and current data. As a consequence, the estimated parameters are subject to inaccuracy. Furthermore, the parameters will, generally, change gradually in time.

A Bayesian approach towards interventions and optimization is natural as it is a tractable and convenient way to mix and weight expert knowledge, prior data and current data, to account for the resulting uncertainty in the design parameters, and to allow for gradual change. The Bayesian component to adaptive survey design consists of 1) prior distributions to parameters in models for nonresponse and measurement and 2) decision rules in interventions and optimization of design.

The Bayesian Adaptive Survey Design Network (BADEN) funded by the Leverhulme trust in the United Kingdom consists of the universities of Manchester (network coordinator), Michigan and Southampton, the national statistical institutes of the Netherlands and Sweden, RTI international, and the US Census Bureau. For this panel proposal, BADEN aims to sponsor a session focusing on case studies with an emphasis on design features and monitoring to inform adaptive survey designs.

The proposed papers are:

Paper 1: Lisette Bruin, Nino Mushkudiani and Barry Schouten (Statistics Netherlands): A Bayesian analysis of mixed-mode data collection. Results from four case studies

Paper 2: Anton Johansson, Peter Lundquist, Sara Westling and Gabriele Durrant (Statistics Sweden and University of Southampton): Modelling length and final response outcome of call sequences in the Swedish Labour Force Survey

Paper 3: Stephanie Coffey, Benjamin Reist (United States Census Bureau): Improving Response Propensity Estimation for Adaptive Design Interventions at the U.S. Census Bureau

Paper 4: Daniel Pratt¹, Jeffrey Rosen¹, David Wilson¹, Melissa Cominole¹, Elizabeth Copello¹, and Andrey Peytchev² (RTI and University of Michigan): Interventions during data collection to increase response and sample representativeness: a field test experiment and responsive design simulation

Paper 5: James Wagner: Using Bayesian Methods to Estimate Response Propensity Models During Data Collection

Appendix 3

4th WORKSHOP ADVANCES IN ADAPTIVE AND RESPONSIVE SURVEY DESIGN

November 9th and 10th, 2015

The Harwood Room, Manchester Meeting Place
University of Manchester Sackville Campus, United Kingdom

Final Programme

Day 1:

- 9:00 – 9:30** **Registration**
- 9:30 – 9:45** **Welcome**
- 9:45 – 10:25** **Keynote Address: Roger Tourangeau**
- 10:25 – 11:05** **Keynote Address: Carl Sarndal**
- 11:05 – 11:30** **Coffee Break**
- 11:30 – 1:00** **Indicators for Quality and Costs (Gabi Durrant – chair, Carl Sarndal – discussant)**
Barry Schouten - Nonresponse adjustment by design?

Nino Mushkudiani, Joep Burgera, Koen Perryck and Barry Schouten - Assessing the impact of inaccurate survey design parameters on the performance of adaptive survey designs

Stephanie Coffey - Evaluating Some Indicators for Nonresponse Bias Under a Variety of Design Conditions
- 1:00 – 2:00** **Lunch**
- 2:00 – 3:30** **Tools and Logistics (Dan Pratt – chair, Francois Laflamme – discussant)**
Peter Miller - The U.S. Census Adaptive Design Research Agenda

Peter Lynn - Targeted designs in longitudinal surveys: prospects and definitions

Michael Thieme and Benjamin Reist - Intra-Application Distributed Processing: Computing Speed and Agility for Adaptive Design
- 3:30 – 4:00** **Coffee Break**
- 4:00 – 5:00** **Paradata and Other Auxiliary Data (Natalie Shlomo – chair, Freuke Kreuter - discussant)**
Jamie Moore, Gabriele Durrant and Peter Smith - Are trajectories of dataset representativeness during survey data collection generalizable? Evidence from the 2011 Census Non-Response Link Study

Anton Johansson, Peter Lundquist, Sara Westling and Gabriele Durrant - Modelling Long Call Sequences and Final Outcome in the Swedish Labour Force Survey to reduce the Number of Unproductive Calls
- 5:00 – 5:30** **Reflections, Follow-up**
- 5:30 – 7:00** **Wine Reception**

Day 2:

9:00 – 9:40

Keynote Address: Rod Little

9:40 – 10:40

Models for Response and Measurement (Barry Schouten – chair, Roger Tourangeau – discussant)

Sjoertje Vos, Annemieke Luiten and Nino Mushkudiani - Indicators of response and quality in Dutch person and household surveys

Alexandru Cernat - Impact of mixed modes on measurement errors and estimates of change in panel data

10:40 – 11:00

Coffee Break

11:00 – 12:00

Bayesian Approaches (Stephanie Coffey – chair, Rod Little – discussant)

James Wagner - Using Bayesian Methods to Estimate Response Propensity Models During Data Collection

Lisette Bruin, Nino Mushkudiani and Barry Schouten - A Bayesian analysis of survey design parameters

12:00 – 1:00

Lunch

1:00 – 2:30

Design Features and Interventions (James Wagner – chair, Peter Lynn – discussant) – Part 1

Annamaria Bianchi - Targeted letters in longitudinal surveys

Daniel Pratt, Melissa Cominole, Elizabeth Copello, Andrey Peytchev, Emilia Peytcheva, Jeffrey Rosen, and David Wilson - Examination of interventions during data collection to increase response and sample representativeness: a field test experiment and simulation

Tobias Gummer - Case prioritization in a survey: Comparing different selection criteria

2:30 – 3:00

Coffee Break

3:00 – 4:00

Design Features and Interventions (James Wagner – chair, Peter Lynn – discussant) – Part 2

Gina Walejko - Prioritizing cases to increase sample representativeness in the Survey of Income and Program Participation

Stephen Kaputa and Katherine Thompson - Investigating Nonresponse Subsampling in an Establishment Survey Through Embedded Experiments

4:00 – 4:30

Discussion, Next Steps