Journal of Survey Statistics and Methodology (2018) **6**, 431–464

A BAYESIAN ANALYSIS OF DESIGN PARAMETERS IN SURVEY DATA COLLECTION

BARRY SCHOUTEN
NINO MUSHKUDIANI
NATALIE SHLOMO
GABI DURRANT
PETER LUNDQUIST
JAMES WAGNER

In the design of surveys, a number of input parameters such as contact propensities, participation propensities, and costs per sample unit play a decisive role. In ongoing surveys, these survey design parameters are usually estimated from previous experience and updated gradually with new experience. In new surveys, these parameters are estimated from expert opinion and experience with similar surveys. Although survey institutes have fair expertise and experience, the postulation, estimation, and updating of survey design parameters is rarely done in a systematic way. This article presents a Bayesian framework to include and update prior knowledge and expert opinion about the parameters. This framework is set in the context of adaptive survey designs in which different

Barry Schouten is with Statistics Netherlands, PO Box 24500, 2490 HA The Hague, the Netherlands and University of Utrecht, Padualaan 14, 3584 CH Utrecht, the Netherlands. Nino Mushkudiani is with Statistics Netherlands. Natalie Shlomo is with the Social Statistics Department, School of Social Sciences, University of Manchester, Humanities Bridgeford Street, Manchester M13 9PL, UK. Gabi Durrant is with the Department of Social Statistics and Demography, School of Social Sciences, University of Southampton, Building 58, Highfield, Southampton SO17 1BJ, UK. Peter Lundquist is with Statistics Sweden, Box 24300, 104 51 Stockholm, Sweden. James Wagner is with the Survey Research Center, Institute of Social Research, University of Michigan, 500 S. State St., Ann Arbor, MI 48109, USA.

This work was supported by the United Kingdom Leverhulme Trust International Network Grant (reference number: IN-2014–046): Bayesian Adaptive Survey Design Network (BADEN) (January 2015 to April 2018). The participating institutions are the Universities of Manchester (coordinator), Michigan, and Southampton, and RTI, Statistics Sweden, the US Census Bureau, and Statistics Netherlands.
*Address correspondence to Natalie Shlomo, Social Statistics Department, School of Social Sciences, University of Manchester, Humanities Bridgeford Street, Manchester M13 9PL, UK; E-mail: natalie.shlomo@manchester.ac.uk.

doi: 10.1093/jssam/smy012 Advance access publication 11 July 2018

population units may receive different treatment given quality and cost objectives. For this type of survey, the accuracy of design parameters becomes even more crucial to effective design decisions. The framework allows for a Bayesian analysis of the performance of a survey during data collection and in between waves of a survey. We demonstrate the utility of the Bayesian analysis using a simulation study based on the Dutch Health Survey.

KEYWORDS: Adaptive survey design; Gibbs sampler; Nonresponse; Response propensities; Survey costs.

1. INTRODUCTION

Over the last two decades, there has been an increasing interest in survey data collection monitoring, analysis, and intervention or adaptation. The main drivers for this are the diversification of data collection that followed the emergence of online communication, the lack of predictability of survey response propensities despite years of research into survey design, the gradual increase in costs per respondent when response rates are kept at traditional levels, and the availability of a wide range of data collection process data (termed paradata). For instance, see Kreuter (2013) to get a view on research aiming at a deeper understanding of data collection processes. In a changing survey data collection environment with unpredictable and, hence, only partially controllable outcomes, a close watch of the progress of data collection is imperative. For this purpose, timely and accurate estimates of survey design parameters, such as contact propensities, participation propensities, costs per contact, and costs per interview, are needed.

These developments paved the way for survey designs that adapt or tailor strategies and effort to known and relevant characteristics of sampled units from the target population. Such designs we term "adaptive," (see Groves and Heeringa 2006; Wagner 2008; and Schouten, Calinescu, and Luiten 2013). In order to adapt, accurate estimates of survey design parameters are needed at the deeper level of population subgroups. This higher resolution puts pressure on the accuracy of such parameter estimates. Burger, Perryck, and Schouten (2017) analyzed the performance of adaptive survey designs under inaccurate design parameter estimates and concluded—not surprisingly—that biased parameter estimates may lead to suboptimal and, consequently, inefficient designs.

A natural approach to evaluate inaccuracy of survey design parameters and to account for the uncertainty in the optimization of survey design is through Bayesian analysis (Gelman, Carlin, Stern, and Rubin 2014). In such an analysis, survey design parameters are treated as random variables and are assigned prior distributions, which are then updated and transformed to posterior distributions during data collection. These posterior distributions may be used as prior distributions in new waves of the same survey. The added benefit is that prior distributions may also be elicited from expert judgment, so that historic survey data in one survey may also be reused in other surveys. Other advantages of using a Bayesian approach is that we are able to account for the uncertainty in survey design parameters and any function of these parameters. In addition, we are able to obtain distributional estimates, such as percentiles, which can be used to develop other types of quality indicators for monitoring adaptive survey designs.

Despite these advantages, Bayesian analysis of survey data collection is rare, and literature is very thin. An exception is Schafer (2013). Reasons for this absence may be that a Bayesian analysis is not straightforward conceptually or computationally, that the added value may be unclear to survey designers, and that the elicitation of prior distributions is complex and cumbersome in practice. Therefore, we have the following objectives:

(1) to set up (sufficiently general) models for survey design parameters
(2) to introduce a Bayesian analysis of survey design parameters
(3) to introduce a Bayesian analysis of quality and cost indicators based on survey design parameters
(4) to evaluate under what conditions a Bayesian analysis has added value

We show that a Bayesian analysis can be set up and applied to a variety of survey designs. Our focus on overall quality and cost indicators is motivated by the desire to ultimately adapt data collection strategies to different population subgroups. To reach this goal, we include linked auxiliary data and paradata in models for survey design parameters. The main objective here, however, is to show that Bayesian analysis can be effective in monitoring and analyzing survey data collection and also demonstrate how a Bayesian approach can be used through a particular set of model assumptions. We emphasize the generalizability of the methodology where other model assumptions can be used.

A Bayesian analysis of data collection is by itself not novel. There is a vast literature in biostatistics and medical statistics that presents methodology to monitor and optimize treatments using prior knowledge or beliefs. There is a close resemblance to dynamic treatment regimes and continual re-assessment methods in clinical trials (e.g., O'Quigley and Shen 1996; Heyd and Carlin 1999; Murphy 2003; Scharfstein, Daniels, and Robins 2003; Schulte, Tsiatis, Laber, and Davidian 2014). An application to survey data collection is, however, novel and introduces three specific elements: a multi-dimensionality of survey target variables, a very explicit focus on data collection costs, and a multitude of quality indicators describing different survey errors. In the analysis, we assume that survey data collection consists of a series of phases for which costs and quality are evaluated separately and cumulatively, including all outcomes, up to a current phase. Per phase, survey design parameters are defined and are assigned prior distributions. A Gibbs sampler with data augmentation is applied to derive posterior distributions.

The explicit presentation of uncertainty in survey design parameter values is an advantage of Bayesian analysis by itself. However, a Bayesian analysis has even more added value when the inclusion of prior knowledge from historic survey data and expert knowledge provides stronger guidance to design decisions than a non-Bayesian analysis. In order to prove this, we take two approaches in a simulation study linked to the Dutch Health Survey. We gradually misspecify the locations of the prior distributions and gradually increase the variance of the prior distributions. We do this to analyze if and when the prior information is too weak to be of use and may just as well be replaced by a fully noninformative prior distribution.

A natural subsequent step is to adapt survey design within the Bayesian analysis framework. In order to be able to do so, a range of strategies needs to be randomized and available. We discuss randomization of strategies in the Bayesian analysis, and we distinguish contact and participation in obtaining response, but we leave actual optimization to future articles.

In this article, we focus on adaptive survey design with the objective to minimize nonresponse error. Survey design parameters associated with measurement error (e.g., the adjusted mode effect) are out of scope in this article. We refer to Calinescu (2013) and Calinescu and Schouten (2015) for designs incorporating both types of survey errors.

This article has three main sections: In section 2, we describe the various adaptive survey design strategies and link them to response propensities and costs, the survey design parameters of interest. In section 3, we break down the response propensities and costs into their basic components (e.g., contact propensities and cost components per call, present models for these components) and assign prior distributions. Apart from the survey design parameters themselves, we also consider a number of functions of these parameters like the response rate, overall costs, and coefficient of variation of the response propensities. In section 4, we investigate the utility of Bayesian analysis through a simulation study on the Dutch Health Survey and we close with a discussion in section 5.

2. ADAPTIVE STRATEGIES AND DESIGN PARAMETERS

In this section, we provide the necessary background for the models of section 3.

2.1 Types of Strategies and Notation

The design of each survey has a range of features, for example, sample design, advance letter, contact protocol, screener interview, number of phases, reminder protocol, use of incentive, mode of administration, interviewer, refusal conversion procedure, and type of questionnaire. The aggregate of choices made for the design features is called a data collection strategy or simply strategy. In nonadaptive survey designs, these features are implemented uniformly over the whole sample (i.e., there is one strategy). In adaptive survey designs, part of the design features may be implemented differently for different sample units (i.e., there is a set of strategies). For examples, see Groves and Heeringa (2006); Wagner (2008); Coffey, Reist, and White (2013); and Schouten, Calinescu, and Luiten (2013).

Different sample units may be distinguished based on linked auxiliary data from the sampling frame or administrative data, from paradata obtained during data collection, or from survey data from previous waves in longitudinal or panel settings. By analogy to clinical trials (e.g., Scharfstein, Daniels, and Robins 2003 and Murphy 2003), an adaptive strategy based on auxiliary data available at the start of data collection is called *static* and an adaptive strategy based (also) on auxiliary data collected during data collection is called *dynamic*. In our implementation, it is decided beforehand what strategy is applied to each sample case when all auxiliary data is available. Hence, at the start of data collection, we may not know the exact strategy since it depends on paradata, but once it becomes available, the action is fixed. As an example, we may make the decision *a priori* that young males that do not respond to the web questionnaire should receive a face-to-face interview only if there was an attempt to respond to the web questionnaire. We only know if there was such an attempt after the start of the data collection.

For a subject i in the sample, let x_i be the vector of auxiliary variables,

$$x_i = \left(x_{0,1,i}, \ldots, x_{0,m_0,i}, \ldots, x_{T,1,i}, \ldots, x_{T,m_T,i}\right)',$$

where $x_{0,i} = \left(x_{0,1,i}, \ldots, x_{0,m_0,i}\right)'$ contains the m_0 auxiliary variables available at the start of data collection, and $x_{t,i} = \left(x_{t,1,i}, \ldots, x_{t,m_t,i}\right)'$ are the auxiliary variables that are observed for the sample units in phase t. Typical auxiliary variables available at the start of data collection are those variables linked to the sample frame from a register, such as age and gender. Auxiliary variables that are observed for sample units in phase t are typically paradata that have been collected to inform the data collection process, such as the number of contact attempts or whether the respondent "broke-off" connection when responding to a web questionnaire.

Let the survey design consist of a maximum of T phases, labelled $t = 1,\ 2, \ldots, T$. (The use of t to denote phase does not necessarily indicate that it is related to time.) We define $\mathcal{S}_t$ as the collection of all possible actions in phase t and let s_t represent the action taken in phase t. Possible actions may include moving from a web questionnaire to a face-to-face interview or stopping contact attempts. We define the total collection of possible actions: $\mathcal{S} := \cup_{t=1}^{T} \mathcal{S}_t$. The action sets may contain $s_\varnothing$, which, if selected, implies that no attempt is made to obtain a response. We define the collection of survey strategies from phase one to T as

$$\mathcal{S}_{1,T} := \{(s_1, \ldots, s_T) : s_t \in \mathcal{S}_t, t = 1,\ 2, \ldots, T\},$$

and let $s_{1,T} \in \mathcal{S}_{1,T}$ denote one possible strategy (i.e., sequence of actions from phase one through phase T).

2.2 Survey Design Parameters

Adaptive survey designs either maximize a quality objective subject to cost constraints and other quality constraints or minimize a cost objective subject to quality constraints. The quality and cost constraints depend on the setting in which the survey is conducted but may concern any survey error. Three sets of survey design parameters suffice to compute most of the quality and cost constraints:

(1) response propensities, $\rho_{T,i}(s_{1,T})$, per unit i and strategy $s_{1,T}$
(2) costs, $C_{T,i}(s_{1,T})$, per unit i and strategy $s_{1,T}$
(3) strategy-specific bias on a specified key survey outcome variable, $D_i(s_{1,T})$, per unit i and strategy $s_{1,T}$ relative to a benchmark strategy

In this article, we concentrate on nonresponse and costs, and to keep the scope manageable, we do not model key survey variables. For this reason, we do not consider the third set of parameters, the strategy-specific biases, relative to a benchmark strategy. We leave this to future research.

There are two options in defining and modeling survey design parameters: the stratum or subgroup level and the individual sample unit level. The first option implies that the average response and costs in a stratum are modeled (i.e., addressing variation within such strata), whereas the second option implies that models for individual units are created. The two options, essentially, represent two main approaches in adaptive design, stratum allocation, and case prioritization (e.g., Peytchev, Riley, Rosen, Murphy, and Lindblad 2010; Wagner, West, Kirgis, Lepkowski, Axinn et al. 2013; Rosen, Murphy, Peytchev, Holder, Dever et al. 2014; Luiten and Schouten 2013; Särndal and Lundquist 2014; and Schouten and Shlomo 2017). In this article, we model individual design parameters, since it offers more flexibility. Any stratification may still be applied afterwards, and stratum design parameters may be derived from the individual propensities and cost functions.

3. MODELING SURVEY DESIGN PARAMETERS

We construct hierarchical Bayes models for response propensities and costs per sample unit and assign prior distributions to the parameters in these models.

3.1 Decomposition of Survey Design Parameters

We give basic models for response propensities and costs. We break down these parameters into their basic components: contact propensities and costs and participation propensities and costs. Outcomes other than contact, noncontact, and refusal/participation are possible and can be included in a relatively straightforward way. However, the number of parameters to be estimated increases with each outcome that we include.

We make two general assumptions: First, we assume that making contact, obtaining participation, and the costs associated with an individual sample unit are independent of contact, participation, and the costs of any other individual sample unit. Thus, we ignore any effect that clustering of sample units may have on response or costs. Second, we assume that there is a stable workload for data collection across interviewers (i.e., we ignore any impact of differential sample sizes).

However, we do allow for associations between contact propensities over phases, between participation propensities over phases, and between cost functions over phases.

Let $\kappa_{t,i}(s_{1,T})$ be the propensity of a contact in phase, $t = 1, 2, \ldots, T$, under strategy $s_{1,T}$ given that the unit did not respond in earlier phases and is eligible for follow-up. Let $\lambda_{t,i}(s_{1,T})$ be the propensity of a participation in phase t of subject i under strategy $s_{1,T}$ given contact.

We assume that propensities do not depend on the future actions after phase t, that is, for all $s_{t+1,T}$ and an alternative $\tilde{s}_{t+1,T}$,

$$\kappa_{t,i}\left(s_{1,t}, s_{t+1,T}\right) = \kappa_{t,i}\left(s_{1,t}, \tilde{s}_{t+1,T}\right), \ t = 1, 2, \ldots, T,$$

$$\lambda_{t,i}\left(s_{1,t}, s_{t+1,T}\right) = \lambda_{t,i}\left(s_{1,t}, \tilde{s}_{t+1,T}\right), \ t = 1, 2, \ldots, T,$$

and we omit the dependence on the future actions in the notation. Furthermore, $s_{1,1} = s_1$.

The response propensity through phase t of a subject i under strategy $s_{1,t}$, is denoted by $\rho_{t,i}(s_{1,t})$. When in subsequent phases all nonresponse receives a follow-up, then the response propensity through all T phases of data collection equals

$$\rho_{T,i}\left(s_{1,T}\right) = \kappa_{1,i}\left(s_{1,}\right) \ \lambda_{1,i}(s_1) + \sum_{t=2}^{T}\left(\left(\prod_{l=1}^{t-1}\left(1 - \kappa_{l,i}\left(s_{1,l}\right) \ \lambda_{l,i}(s_{1,l})\right)\right)\kappa_{t,i}\left(s_{1,t}\right) \ \lambda_{t,i}(s_{1,t})\right). \quad (1)$$

When in subsequent phases only noncontacts receive a follow-up, (i.e., refusal conversion is not allowed), then this response propensity is

$$\rho_{T,i}(s_{1,T}) = \kappa_{1,i}(s_1)\ \lambda_{1,i}(s_1) + \sum_{t=2}^{T}\left(\left(\prod_{l=1}^{t-1}\left(1-\kappa_{l,i}(s_{1,l})\right)\right)\kappa_{t,i}(s_{1,t})\ \lambda_{t,i}(s_{1,t})\right). \quad (2)$$

In general, the costs per sample depend on the phase, the sample unit, and the strategy. We define for a sample unit i with auxiliary vector x_i in phase t, following strategy $s_{1,t} \in \mathcal{S}_{1,t}$

- $C_{0,i}(s_{1,t})$ as the cost to make a contact attempt (visit or call)
- $C_{R,i}(s_{1,t})$ as the cost for the response given contact in phase t
- $C_{NR,i}(s_{1,t})$ as the cost for a nonresponse given contact in phase t

For some actions, these functions may be identical to zero (e.g., a response or nonresponse to a web survey). In this article, we make the simplification that cost functions do not depend on the phase and history of actions but only on the current action, i.e.,

$$C_{0,t,i}(s_{1,t}) = C_{0,i}(s_t),\ \ C_{R,t,i}(s_{1,t}) = C_{R,i}(s_t),\ \ \ C_{NR,t,i}(s_{1,t}) = C_{NR,i}(s_t). \quad (3)$$

The cost parameters $C_i(s_{1,T})$ can be written using these components and the contact and participation propensities. Under (3), when all nonresponse receives follow-up, we get

$$\begin{aligned} C_{T,i}(s_{1,T}) = &\ C_{0,i}(s_1) + \kappa_{1,i}(s_1)\ \left(1-\lambda_{1,i}(s_1)\right)C_{NR,i}(s_1) \\ &+ \kappa_{1,i}(s_1)\ \lambda_{1,i}(s_1)\ C_{R,i}(s_1) \\ &+ \sum_{t=2}^{T}\left(\left(\prod_{l=1}^{t-1}\left(1-\kappa_{l,i}(s_{1,l})\ \lambda_{l,i}(s_{1,l})\right)\right)\right. \\ &\left(C_{0,i}(s_{1,t}) + \kappa_{t,i}(s_{1,t})\ \left(1-\lambda_{t,i}(s_{1,t})\right)C_{NR,i}(s_{1,t})\right. \\ &\left.\left.+ \kappa_{t,i}(s_{1,t})\ \lambda_{t,i}(s_{1,t})\ C_{R,i}(s_{1,t})\right)\right). \end{aligned} \quad (4)$$

For self-administered modes, like web or mail, we are unable to discern a noncontact from a refusal in most cases. For this reason, for these modes, we model only one binary indicator for response/nonresponse, and we model only contact costs and response costs.

3.2 Models for Survey Design Parameter Components

General models for $\kappa_{t,i}(s_{1,t})$ and $\lambda_{t,i}(s_{1,t})$ accounting for all possible associations with the full set of actions $s_{1,t}$, the auxiliary vector x_i, and the phase t would be very complicated and cumbersome and may lead to confusion rather than clarification. We refer to Durrant, D'Arrigo, and Steele (2011), Durrant and Steele (2013), Durrant, Maslovskaya, and Smith (2015), who describe multilevel models and hazard rate models for such general settings. We, therefore, make a number of simplifications in the model specification.

We model the contact propensities with a probit model (i.e., using a binomial link function). Each sample unit has a contactability represented as a continuous latent variable $Z^C_{t,i}(s_{1,t})$, and contact is obtained when this latent variable is larger than zero,

$$u^C_{t,i}(s_{1,t}) = \begin{cases} 1, & Z^C_{t,i}(s_{1,t}) > 0, \\ 0, & Z^C_{t,i}(s_{1,t}) \leq 0, \end{cases}$$

where $u^C_{t,i}(s_{1,t})$ is the indicator of contact of subject i in phase t following strategy $s_{1,t}$ and $Z^C_{t,i}(s_{1,t}) \sim N(\mu(s_{1,t}), \sigma(s_{1,t}))$, for some $\mu_{t,i}(s_{1,t}),\ \ \sigma_{t,i}(s_{1,t})$ so that

$$\kappa_{t,i}(s_{1,t}) = P\left(Z^C_{t,i}(s_{1,t}) > 0\right).$$

For $m \leq m_k$, let $\alpha^C_{t,k,m}(s_{1,t})$ be the regression coefficient in phase t corresponding to the m-th entry in the auxiliary vector $x_{k,i}$, given that $s_{1,t}$ is applied to a unit. Obviously, $\alpha^C_{t,k,m}(s_{1,t}) = 0$, when $k \geq t$. Let $\alpha^C_{t,k}(s_{1,t}) = (\alpha^C_{t,k,1}(s_{1,t}), \ldots, \alpha^C_{t,k,m_k}(s_{1,t}))'$ be the coefficients corresponding to $x_{k,i}$ in phase t. The model can be written as follows:

$$Z^C_{t,i}(s_{1,t}) = \sum_{k=0}^{t-1} \alpha^C_{t,k}(s_{1,t}) x_{k,i} + \varepsilon^C_{t,i},$$

where $\varepsilon^C_{t,i} \sim N(0,\ 1)$ is an error term for the uncertainty of contact of the subject.

Now, the number of coefficients in all contact propensity models is $\sum_{k=0}^{T} (T-k)m_k$ for one specific strategy $s_{1,T}$. The total number of coefficients depends on the sizes $|\mathcal{S}_t|$ of the action sets. It is clear that this number can become very large—too large to be feasible in estimation. Hence in practice, models are usually simplified by lowering the number of coefficients.

We evaluate a model that has all important features for adaptive survey designs, but that is as simple as possible. First, to be able to include dynamic adaptive survey designs, we need to include paradata. To keep the model simple however, we assume there is just one phase, say t_1, in which paradata is collected. Up to phase t_1, only the auxiliary variables in $x_{0,i}$ can be used to model the propensities. After phase t_1, the auxiliary variables obtained in phase t_1 can also be included in the model. Second, we account for dependence of success in a certain phase on past actions, which can be included by introducing a fixed or random effect per possible history. We add the history as a random effect here. Since we add a dependence on the history of actions, the regression coefficients become necessarily dependent on the phase. The model becomes

$$Z_{t,i}^{C}\left(s_{1,t}\right)=\begin{cases}\alpha_{t,0}^{C}(s_t)x_{0,i}+\varepsilon_{t,i}^{C}+\ \delta_t^{C}\left(s_{1,t}\right),\ t\leq t_1,\\ \alpha_{t,0}^{C}(s_t)x_{0,i}+\alpha_{t,1}^{C}(s_t)x_{t_1,i}+\varepsilon_{t,i}^{C}+\ \delta_t^{C}\left(s_{1,t}\right),\ t>t_1,\end{cases} \tag{5}$$

where $\delta_t^{C}\left(s_{1,t-1}\right)$ is a random effect.

The model for the participation propensity can be derived analogously and is not given. The $\alpha_{t,k}^{C}$, $\varepsilon_{t,i}^{C}$ and $\delta_t^{C}\left(s_{1,t}\right)$ are replaced by $\alpha_{t,k}^{P}$, $\varepsilon_{t,i}^{P}$ and $\delta_t^{P}\left(s_{1,t}\right)$.

We make the simplification that cost functions do not depend on the phase and design features in previous phases but only on the current phase and design features. A model for the costs functions is

$$C_{g,i}(s)=\gamma_g(s)x_i+\varepsilon_{g,i}(s),\ \ s\in\mathcal{S} \tag{6}$$

where $g\in\{0,\ R,\ NR\}$, $\gamma_g(s)$ are regression parameters allowing for interaction between the current action s and the auxiliary vector x_i, and the $\varepsilon_{g,i}(s)$ are error terms that again allow for an interaction with the current action. The error terms are modeled as independent normal, but other distributions may be considered depending on the application when costs components are skewed or attain values close to zero:

$$\varepsilon_{g,i}(s)\sim N\left(0,\sigma_g^2(s)\right). \tag{7}$$

3.3 A Bayesian Analysis of Survey Design Parameters

We make the analysis Bayesian by assigning prior distributions to the regression coefficients and random effects of section 3.2. Our aim is the derivation of the posterior distributions of the individual response propensities $\rho_{T,i}\left(s_{1,T}\right)$ and the individual cost parameters $C_{T,i}\left(s_{1,T}\right)$ per strategy given observed data. These propensities and costs are, in general, complex functions of the underlying survey design parameters per phase. We derive expressions for the full conditional distributions of the regression coefficients and random effects, but we propose to rely on numerical approximations and Markov Chain Monte Carlo (MCMC) methods to generate draws from the posterior distributions.

3.3.1 Prior distributions. We assign prior distributions to the model parameters in (5), (6), and (7). We assume that regression slope parameters and dispersion parameters are independent over different data collection phases, but they may be dependent within a phase.

For the regression slope parameters and random effects in contact and participation models, we choose normal prior distributions. Despite being based on normal distributions themselves, probit models do not allow for conjugate pairs of prior and posterior distributions for the regression parameters (e.g., Albert and Chib 1993). The normal distributions are obvious choices (see Gelman et al. 2014) but may also be replaced by other distributions. Let $h \in \{\text{"C"}, \text{"P"}\}$. The contact and participation regression slope parameters are modeled as

$$\alpha_t^h(s) \sim N\left(\mu^h(s), \Sigma^h(s)\right), \tag{8}$$

and the contact and participation random effects are modeled as

$$\delta_t^h \sim N\left(0, \ \left(\tau_t^h\right)^2\right), \tag{9}$$

where $\left(\tau_t^h\right)^2$ are specified covariance matrices.

The models for the cost functions are linear. Here, conjugate prior-posterior pairs are possible. We choose normal distributions for the regression slope parameters and inverse Gamma for the regression dispersion parameters. Inverse Gamma distributions are suggested for random effect variance parameters, see Gelman (2006), as they lead to conditionally conjugate prior-posterior pairs. Let again $g \in \{0, \ R, \ NR\}$.The cost regression slope parameters are modeled as

$$\gamma_g(s) \sim N\left(\mu_g(s), \Sigma_g(s)\right), \tag{10}$$

and the cost error term variances are modeled as

$$\sigma_g^2(s) \sim \Gamma^{-1}\left(a_g(s), b_g(s)\right). \tag{11}$$

The contact, participation, and costs models are hierarchical Bayes because different individuals share parameters and because random effects spread out over phases t and actions s_t. The normal and inverse Gamma probability distributions for the regression parameters and random effects are then called hyperpriors. The hyperparameters in (8) to (11) need to be elicited from historic survey data or expert knowledge.

3.3.2 Posterior distributions. The aim is to derive the posterior distributions of response propensities and cost parameters given the observed data. The observed data consists of the following:

- The response outcome per phase per sample unit: $u_{t,i}$
- The realized costs per phase per sample unit: $c_{0,t,i}$, $c_{R,t,i}$ and $c_{NR,t,i}$. Per phase $c_{R,t,i}$ or $c_{NR,t,i}$ is observed only when contact is made. Contact costs are always observed in every phase. Since we do not model variation in costs for

contact, response and nonresponse when the same action is applied in multiple phases, we average realized costs over all phases that employed the same actions.

- The complete auxiliary vector: x_i

Data collection strategies may be randomized in order to learn about multiple strategies simultaneously. There is a vast literature on efficient randomization in adaptive or dynamic treatment regimes (e.g., Murphy 2003; Chakraborty and Murphy 2014; and Laber, Lizotte, Qian, Pelham, and Murphy 2014). In general, designs are called sequential multiple assignment randomized trial (SMART) (e.g., Lei, Nahum-Shani, Lynch, Oslin, and Murphy 2012), when the randomization is independent of future outcomes and, hence, allows for disentangling the outcomes for different strategies. As mentioned in section 2.1, we assume that randomization is only at the outset; strategy allocation probabilities may depend on auxiliary information known at the start of data collection but not on paradata. In addition to the above, we also observe the series of actions, or simply strategy, that were applied per sample unit: $s^i_{1,T}$.

In the following, we use $\rho(s_{1,T})$ and $C(s_{1,T})$ as shorthand for the vector of response propensities and cost parameters over all sample units for a particular strategy. In the same fashion, we use u_t, c_0, c_R, c_{NR}, and x to denote the vectors of outcomes, realized costs components, and auxiliary variables over sample units. With $\{s_{1,T,i}\}$, we denote the vector of used strategies for all sample units. To shorten expressions, we use $\alpha, \delta, \gamma, \sigma^2$ for the vectors of regression slope parameters, random effects, and regression dispersion parameters over phases and actions. For convenience, we use p to express joint and marginal density functions; we omit the reference to the random variables to which they apply and ignore differences between discrete and continuous probability distributions. Finally, in the density functions, we omit the dependence on the hyperparameters.

The joint posterior distribution of interest is

$$p\left(\rho\left(s_{1,T}\right),\ C\left(s_{1,T}\right)\middle|u_t, c_0, c_R, c_{NR}, x, \{s^i_{1,T}\}\right). \tag{12}$$

This joint density follows from integration over all possible combinations of regression parameters $\alpha, \beta, \delta, \gamma, \sigma^2$ and cannot be written in closed form. A straightforward solution is to apply a Gibbs sampler to the joint density of the regression parameters $\alpha, \beta, \delta, \gamma, \sigma^2$:

$$p\left(\alpha, \delta, \gamma, \sigma^2\middle|u_t, c_0, c_R, c_{NR}, x, \{s^i_{1,T}\}\right). \tag{13}$$

An approximation to the joint density in (12) comes as an important by-product of a Gibbs sampler applied to (13); per draw the response propensities and cost parameters can be computed by (5) and (6) and inserting them into (1) or (2) and (4).

A Gibbs sampler for (13) involves repeated draws from the conditional densities of each regression parameter, given the observed data and the other regression parameters, the so-called full conditionals. Appendix A contains expressions for the full conditionals per regression parameter. There is a range of options for sampling from these conditional distributions (see for example Albert and Chib 1993 and Gelman et al. 2014). We choose the approach proposed by Albert and Chib (1993), where we include draws of the latent variables for contact and participation in the Gibbs sampler as a form of data augmentation. In order to carry out the data augmentation, we programmed the Gibbs sampler in R and did not make use of standard libraries in R (e.g., mcmc or gibbs.met) or SAS (e.g., PROC MCMC). The code is available upon request.

3.4 A Bayesian Analysis of Functions of Survey Design Parameters

For monitoring and optimization of data collection, the focus is on functions of the design parameters that correspond to overall quality or cost objectives. We consider three such functions here: the response rate, the total costs, and the coefficient of variation of the response propensities; other functions can often be analyzed in an analogous way. See Nishimura, Wagner, and Elliott (2016) for a discussion of indicators.

Let d_i represent the design or inclusion weight for sample unit i, $i = 1, 2, \ldots, n$. The weighted response rate, RR, for strategy $s_{1,T}$ can be written as

$$RR(s_{1,T}) = \frac{1}{\sum_{i=1}^{n} d_i} \sum_{i=1}^{n} d_i \rho_{T,i}(s_{1,T}), \qquad (14)$$

the total costs, or required budget, B, associated with s_1^T are

$$B(s_{1,T}) = \sum_{i=1}^{n} c_{T,i}(s_{1,T}), \qquad (15)$$

and the coefficient of variation of the response rate, CV, is

$$CV(X, s_{1,T}) = \frac{\sqrt{\frac{1}{\sum_{i=1}^{n} d_i} \sum_{i=1}^{n} d_i (\rho_{T,i}(s_{1,T}) - RR(s_{1,T}))^2}}{RR(s_{1,T})}, \qquad (16)$$

where $\sum_{i=1}^{n} d_i = N$ for many customary sampling designs.

For the *CV* (Schouten, Cobben, and Bethlehem 2009; De Heij, Schouten, and Shlomo 2015), we explicitly denote the dependence on the covariate vector *X*; for any other choice of auxiliary variables, it will generally attain a different value.

The prior and posterior distributions for these three functions are determined by the prior and posterior distributions of the components of the response propensities and cost functions. The posteriors have even more complex forms than the posteriors for individual response propensities and cost parameters. However, (14) through (16) can again be approximated as a by-product of the Gibbs sampler in section 3.3.2.

3.5 Model and Prior Specification

We have specified a sufficiently general Bayesian model for nonresponse and costs over a sequence of data collection phases. The model choices consist of the types of nonresponse, the covariates included in the various nonresponse and cost models, the potential inclusion of interactions between covariates in these models, the link function to transform the latent propensity to a binary nonresponse outcome, the normality of the cost error terms, the random effects to introduce dependency on historic actions, and, the prior specifications of the various regression parameters. The efficacy of the analysis depends heavily on these specifications, so that exploratory model checks and data analysis are imperative (Gelman et al. 2014, Chapters 6 and 7). For example, some costs components may be skewed so that a log transformation is needed to justify a normal error model. Some costs components may also attain values close to zero so that a distribution with support $[0, \infty)$ may need to be favored over a normal distribution.

It goes beyond the scope of this article to discuss nonresponse and cost model specification in full detail. We refer to Groves (1989), Bethlehem, Cobben, and Schouten (2011), and Kreuter (2013). However, we recommend keeping models parsimonious and only including elements that are, or may be, varied in (adaptive) survey design. More specifically, include cost components that vary between design choices and consider nonresponse type, covariate, and design feature combinations that are known to be effective and matter to the key survey variables.

The Bayesian modelling has two elements: the hierarchy/levels and the prior distributions. Hierarchy introduces dependence between parameters such as the random effects in our models, because parameters are known to be associated and for parsimony. We recommend following empirical results in the literature and performing model checking when historic survey data are available. The choice of prior distributions and their associated hyperparameters can be influential. We investigate the sensitivity of the specification of the prior distributions in section 4. Sensitivity analyses, such as in this article, are imperative to find the right level of prior variance when experts are consulted and to select the right amount of historic survey data when parameters change over time.

4. A SIMULATION STUDY TO INVESTIGATE THE UTILITY AND SENSITIVITY OF BAYESIAN ANALYSIS

In the simulation study, we investigate the impact of prior distribution specification and of survey sample size. Specifically, we look at the added value of the prior information and the sensitivity to prior specification. More generally, we aim to demonstrate the use of the Bayesian approach and the importance of a sensitivity analysis. First, we present the specifics of the simulation study and discuss how we attempt to prove the efficacy of a Bayesian analysis. Next, we show results of the simulation study and discuss the conditions under which the analysis is useful.

4.1 Design of the Simulation Study

To evaluate the utility and sensitivity of a Bayesian analysis, we compare posterior distributions of response rates, coefficients of variation of response propensities, and total costs starting from different prior distributions for the survey design parameters, more specifically, for the regression slope and dispersion parameters in contact, participation, and cost models per data collection phase. The prior distributions are compared against fully non-informative priors, which have (arbitrary) large variances and expectations that are the same for all population subgroups. These priors conform to lack of knowledge at the start of data collection which we view as a benchmark choice. The Bayesian analysis still allows for an easy display of uncertainty during and after data collection. We make two comparisons that both start from “true” priors. The true priors have expectations that exactly match the simulation model and have variances that correspond to the standard errors for a historic dataset of sample size 10,000 (i.e., as if we have already observed a fairly large and unbiased realization of the survey). In the first comparison, we gradually misspecify expectations of the true priors in order to mimic bias due to time change or change of survey design. However, the variances of the priors remain the same. In the second comparison, we gradually increase variances but keep expectations constant in order to mimic imprecision. The two comparisons allow us to see how much gain comes from the prior knowledge. Note that in our implementation of adaptive survey design, the information in the prior determines only the optimization of the choice of actions in between waves but not during data collection, even if the design is dynamic.

We quantify this gain by the square root of the posterior mean of $(\Theta - \Theta_0)^2$, which we call the root mean square error (RMSE) of the posterior distribution relative to the simulation model values. Let $p_\pi(\Theta | u_t, c_0, c_R, c_{NR}, x, \{s^i_{1,T}\})$ be a posterior for a data collection quality or cost indicator Θ of interest (e.g., the response rate, CV, or total costs), using prior π. The RMSE for this indicator and prior is then defined as,

$$\mathrm{RMSE}(\Theta; \pi) = \sqrt{\left(E_{p_\pi}(\Theta) - \Theta_0\right)^2 + \mathrm{var}_{p_\pi}(\Theta)}, \tag{18}$$

where Θ_0 is the simulation model value.

We base our simulation study on the 2015 Dutch Health Survey (HS) (CBS 2018). The HS has a sequential mixed-mode survey design with the web, followed by face-to-face interviewing, i.e., non-respondents to a web survey invitation are re-allocated to interviewers. We consider three data collection phases: web, short face-to-face, and extended face-to-face. The extended face-to-face corresponds to an additional round of face-to-face visits for those sample units that have not been contacted or that are soft refusals after three face-to-face visits. Two auxiliary variables, gender and age, are linked from administrative data, and one variable, web break-off, is added from phase one paradata. Gender and age are crossed to form six strata, {0–29 years, 30–59 years, 60 years and older} × {female, male}. Web break-off is a binary indicator for a broken-off web response; it is not crossed with the gender-age variable but added as a main effect. We refer to the variables as “GenderAge” and “BreakOff.” From 2015 HS data, contact propensities, participation propensities, and costs per sample unit are derived for the three phases and used to simulate analysis data sets of sample size 1,250, 2,500, 5,000, and 10,000. The simulation probabilities and costs are given in Appendix C. To model contact and participation, we use a probit regression with GenderAge in phase one and GenderAge + BreakOff in phases two and three. For phase one (online data collection), we set participation propensities equal to response and participation costs are set to zero. We do this because online surveys costs are only associated with the contact and not with the interview. For phases two and three, we do distinguish contact and participation propensities. To model costs, we use a linear regression with GenderAge in all phases. Table 1 gives simulation response rates, coefficients of variation, and total costs cumulatively for all phases based on the true simulation model values in the top row of each section.

We chose prior distributions as specified in section 3.3.1 and applied the Gibbs sampler of section 3.3.2. We refer to Appendix B for details about prior elicitation.

Misspecification was introduced by shifting contact and participation propensities for each subgroup in the same direction. For the online phase one, they were increased by 2 percent, 5 percent, and 10 percent. For the face-to-face

Table 1. Expected Response Rates (RR), Coefficients of Variation (CV) and Total Costs (B) Cumulatively Based on the 2015 HS Simulation Model, and Based on the Three Misspecified Priors (Missp light; Missp medium; Missp strong).

	Data	*Web*	*Face-to-face short*	*Face-to-face extended*
RR	True	30.2%	57.6%	60.5%
	Missp light	32.2%	57.2%	59.7%
	Missp medium	35.2%	56.8%	58.8%
	Missp strong	40.2%	56.8%	58.2%
CV	True	0.277	0.069	0.102
	Missp light	0.260	0.061	0.094
	Missp medium	0.238	0.049	0.082
	Missp strong	0.208	0.036	0.063
B	True	3.0	15.2	19.4
	Missp light	3.0	14.5	19.5
	Missp medium	3.0	13.6	19.8
	Missp strong	3.0	12.1	20.3

phases, they were decreased by 2 percent, 5 percent, and 10 percent for the remaining nonrespondents. We denote the 2 percent as "Missp light," the 5 percent as "Missp medium," and the 10 percent as "Missp strong." Hence, we mimic an overestimation of online response and an underestimation of subsequent face-to-face response, which essentially leads to an underestimation of required budget. Table 1 also contains the expected response rates, coefficients of variation, and costs based on the three sets of misspecified priors. The convergence properties of the Gibbs Sampler are presented in Appendix D.

4.2 Simulation Results

We evaluate the utility of the Bayesian analysis by assuming the effect of increasing the variances of prior distributions and shifting their expectations.

4.2.1 Variance of the prior distributions. In the first evaluation, we focus on the variance term of the RMSE of the posterior distributions and vary the sample size of the observed data. The true prior is compared with the relatively non-informative prior.

Table 2 shows the RMSE for the non-informative and the true priors for four sample sizes: 1,250, 2,500, 5,000, and 10,000 units. Three variance levels are used to misspecify the true prior according to its scale as follows: the variance is obtained corresponding to a historic data set of a modest size of 1,250 units leading to a large prior variance (denoted "V large"); the variance is

Table 2. RMSE*1000 for Fully Non-informative and True Priors for Response Rates (RR), Coefficients of Variation (CV) and Costs (B) Cumulatively after Each Phase and for a Dataset of Sample Sizes 1,250, 2,500, 5,000 and 10,000. The true priors have a variance corresponding to historic sample units of size 1,250 (V large), 2,500 (V mod) and 10,000 (V small). F2F and F2FE are short for face-to-face in phase two and extended face-to-face in phase three.

Size	*Prior*	*RR*			*CV*			*B*		
		Web	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*
1,250	Non-informative	14	19	15	46	45	37	10	316	374
	True V large	10	12	10	21	23	18	10	218	273
	True V mod	8	9	8	14	15	12	10	178	223
	True V small	4	5	5	7	8	8	10	116	142
2,500	Non-informative	10	10	10	12	55	41	9	239	298
	True V large	8	8	8	10	35	25	9	204	247
	True V mod	7	7	7	8	25	17	9	181	217
	True V small	4	4	4	6	9	7	9	128	148
5,000	Non-informative	6	7	7	9	23	16	9	156	183
	True V large	6	6	6	8	18	13	8	143	166
	True V mod	5	6	6	7	15	10	8	134	157
	True V small	4	4	4	5	7	6	9	105	121
1,0000	Non-informative	5	5	5	7	8	9	10	120	135
	True V large	4	5	5	6	8	9	9	114	130
	True V mod	4	5	5	6	8	10	9	111	125
	True V small	3	4	4	5	8	9	10	10	108

designed to correspond to a historic data set of a moderate size of 2,500 units leading to a moderate prior variance (denoted "V mod"); the variance is obtained corresponding to a historic data set of a large size of 10,000 units leading to a small prior variance (denoted "V small").

We note that the RMSE depends on the scale of the population parameters of interest; RMSE values for costs are, therefore, larger.

The RMSE values under the true priors are always lower than for the non-informative prior, as expected. The gap gets larger when the sample size decreases or the true prior variance decreases. However, for a sample size of 10,000, the added value of prior information is already quite small. For even larger sample sizes, it will not make much difference whether the prior knowledge is added or not.

The most advantageous setting is where both prior variance and the observed data sample size are smallest. The biggest gap in RMSE is indeed found for a prior with variance "V small" and sample size 1,250. The RMSE values of this combination are comparable to that of the non-informative prior with sample size 10,000.

In the analysis, we consider the population as a whole. However, when methods are applied to subpopulations, sample sizes are smaller, and the prior distributions still have added value.

The results of the first evaluation suggest that a Bayesian analysis is advantageous for small to modest size samples of (sub)populations and where historic survey data and expert knowledge lower the variances of the posterior distributions.

4.2.2 Misspecification of the prior distributions. In the second evaluation, we gradually misspecify the prior distributions for the contact and participation regression slope parameters and compare the RMSE to a fully non-informative prior. We view the non-informative prior again as the analysis benchmark.

Table 3 contains the RMSE values for non-informative and misspecified priors estimated using the Gibb sampler. Again, we choose three variance levels, corresponding to a historic dataset of 1,250 ("V large"), 2,500 units ("V mod"), and 10,000 units ("V small"). Furthermore, we evaluate four sample sizes: 1,250, 2,500, 5,000, and 10,000. Recall from table 1 that, for phase one, the misspecification leads to a growing overestimation of the response rate and a growing underestimation of the coefficient of variation, whereas costs are fixed. The cumulative response rates after phases two and three are affected only a little, but the coefficient of variation is underestimated. The cumulative costs after phase two are underestimated, but after phase three they are slightly overestimated.

The main observation from the RMSE values in table 3 is that a misspecified prior can be worse than a non-informative prior, but the misspecified prior will outperform a non-informative prior when the misspecification is modest or the variance of the prior is relatively large. Furthermore, in close analogy to the results in the previous subsection, it holds that the larger the sample of the observed data, the smaller the misspecification must be to outperform the non-informative prior.

The 2 percent shift in propensities under misspecified light is small enough for the CV to get RMSE values that are similar or smaller than those for the non-informative prior. This holds also to some extent for the 5 percent and 10 percent shifts under misspecified moderate and large, when the variance of the prior is large.

For the response rate and costs, RMSE values are almost always larger for the misspecified priors, unless the variance of the prior is relatively large.

As expected, decreasing the sample size of the observed data leads to higher RMSE values for all priors. When sample sizes are lowered, in general, the misspecified priors will ultimately perform better than the non-informative prior according to the RMSE shown in (18), since the infinite variance of the

Table 3. RMSE*1000 for Fully Non-informative and Misspecified Priors for Response Rates (RR), Coefficients of Variation (CV) and Costs (B) Cumulatively after Each Phase and for a Dataset of Sample Sizes 1,250, 2,500, 5,000, and 10,000. The misspecified priors have a variance corresponding to historic sample units of size 1,250 (V large), 2,500 (V mod), and 10,000 (V small). F2F and F2FE are short for face-to-face in phase two and extended face-to-face in phase three.

Size	*Prior*	*RR*			*CV*			*B*		
		Web	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*
1,250	Non-informative	14	19	15	46	45	37	10	316	374
	Missp light V large	12	14	12	19	20	16	10	398	265
	Missp light V mod	14	11	11	12	13	12	10	459	284
	Missp light V small	18	7	9	6	13	15	10	549	356
	Missp medium V large	23	26	26	16	20	16	10	1061	719
	Missp medium V mod	32	27	30	10	14	15	10	1360	964
	Missp medium V small	44	29	36	6	23	28	10	1767	1295
	Missp strong V large	46	10	10	13	15	16	10	1346	689
	Missp strong V mod	63	8	8	8	19	26	10	1768	942
	Missp strong V small	87	5	5	8	33	47	10	2324	1281
2,500	Non-informative	10	10	10	12	55	41	9	239	298
	Missp light V large	8	9	8	9	32	24	9	213	196
	Missp light V mod	10	8	8	8	21	15	9	286	203
	Missp light V small	30	8	5	6	10	14	9	706	543
	Missp medium V large	14	15	17	9	32	22	9	619	441
	Missp medium V mod	22	19	22	8	19	12	9	959	702
	Missp medium V small	53	15	22	5	15	22	9	1787	1379
	Missp strong V large	29	8	8	9	23	15	9	819	440
	Missp strong V mod	45	7	7	9	12	12	9	1251	682

	Missp strong V small	77	8	4	9	28	37	9	2065	1151
5,000	Non-informative	6	7	7	9	23	16	9	156	183
	Missp light V large	7	6	6	8	17	12	9	164	155
	Missp light V mod	9	6	6	8	13	9	9	217	168
	Missp light V small	14	5	6	6	8	9	9	399	267
	Missp medium V large	11	9	10	9	16	11	9	407	308
	Missp medium V mod	17	12	14	9	12	9	9	669	504
	Missp medium V small	33	21	27	8	13	16	9	1323	994
	Missp strong V large	20	6	6	9	12	9	9	520	300
	Missp strong V mod	32	6	6	10	10	11	9	864	488
	Missp strong V small	65	7	4	10	27	31	9	1736	984
10,000	Non-informative	5	5	5	7	8	9	10	120	135
	Missp light V large	5	5	5	7	9	10	9	106	116
	Missp light V mod	5	5	5	6	9	11	9	128	116
	Missp light V small	10	5	6	6	12	13	9	287	196
	Missp medium V large	6	8	8	7	9	10	9	216	176
	Missp medium V mod	10	10	11	7	10	11	10	386	297
	Missp medium V small	24	19	23	7	16	18	9	988	752
	Missp strong V large	11	5	5	7	10	13	9	275	169
	Missp strong V mod	19	5	5	7	13	16	9	500	287
	Missp strong V small	48	4	4	9	26	29	9	1288	739

non-informative prior will dominate over the misspecification. The (pathological) exception is where the expectation of the non-informative prior happens to be close to the true value (e.g., true contact or participation propensities do not vary between subpopulations and are also close to 50 percent).

The results of this second evaluation suggest turning points for the utility of a Bayesian analysis that depend on the size of the misspecification, the size of the sample, and the variance of the prior distributions. This is a complex function that requires further study. However, the results under the current simulation model show that misspecification may be very influential and may quickly reduce the added value of a Bayesian analysis.

To give a further impression, figure 1 shows posterior distributions produced by the Gibbs sampler for a selection of the regression slope parameters over the three phases. Next to the posterior densities, the densities of the selected priors are also shown, as well as the density of the true prior assuming a historic dataset of 10,000 sample units. The posterior densities clearly have different means and variances, depending on the prior specification and the data sample size. However, in most cases, the posterior densities overlap with the true prior and often have very similar support.

5. DISCUSSION

We introduced a Bayesian model for survey design parameters related to response and costs. The model is general in that it describes multiple data collection phases, includes both auxiliary variables that are available before data collection starts and auxiliary variables that become available during data collection (paradata), acknowledges multiple nonresponse outcomes, accounts for dependence on previous actions, and enables the inclusion of randomization over different data collection strategies. Many surveys conducted by statistical institutes can fit into this framework. Furthermore, we constructed an analysis strategy based on a Gibbs sampler in which all model parameters are repeatedly drawn. The Gibbs sampler provides estimates for the posterior distributions of the contact and participation propensities and the costs per sample unit. From the Gibbs sampler, the posterior distributions for overarching quality indicators, such as the response rate or coefficient of variation of the response propensities, and cost indicators can easily be derived as an important by-product. Under this particular model, the computational time of the Gibbs sampler was manageable and sufficiently short to run overnight for a range of scenarios. We are thus able to meet the first three objectives of the article as stated in the introduction to set up a Bayesian analysis for survey data collection monitoring and analysis.

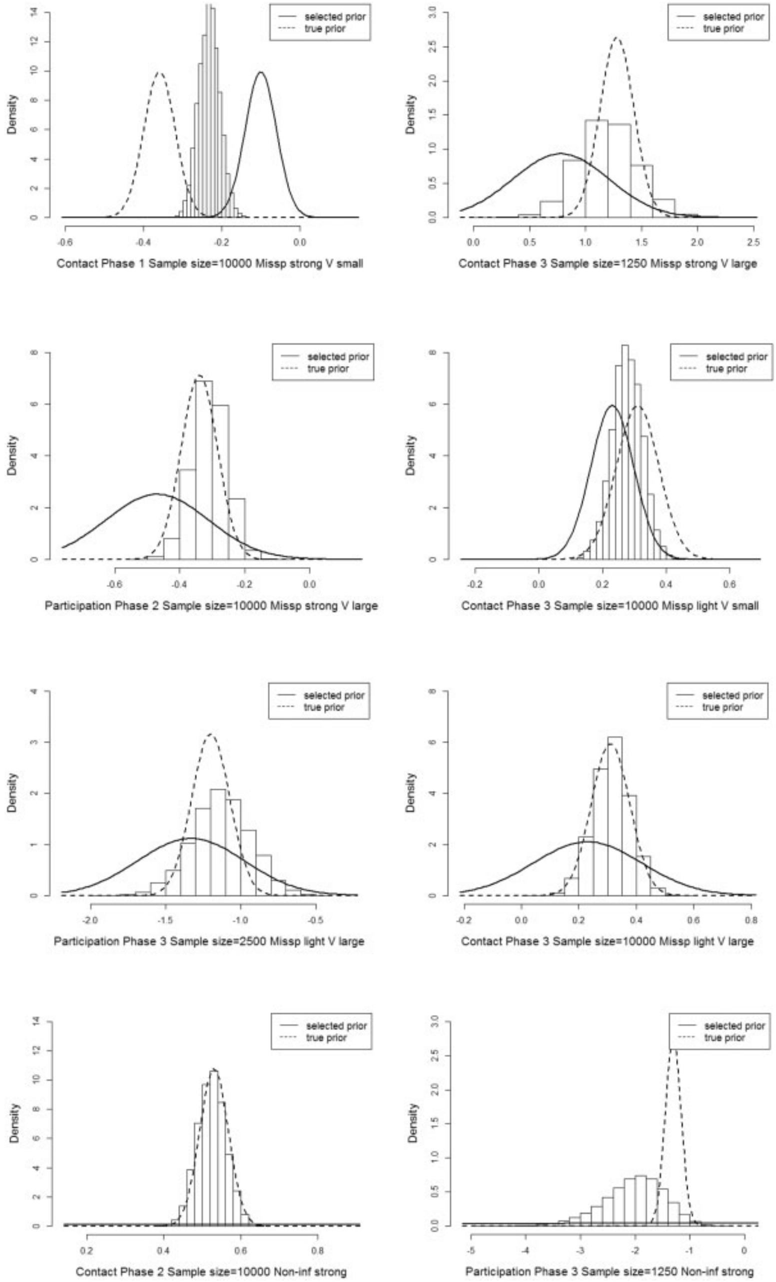

Figure 1. Empirical Posterior Densities for the Slope Parameters for the Gibbs Sampler Runs of the Misspecified and Non-informative Priors for Different Phases and Different Data Sets.

The fourth and most important objective is to show the added value and sensitivity of a Bayesian analysis. In the evaluation, we viewed a fully non-informative prior as the benchmark in which no historic survey data or expert knowledge is incorporated. The evaluation is based on a simulation study using realistic contact propensities and costs and participation propensities and costs from a multimode survey. The evaluation shows that the Bayesian analysis is sensitive to misspecification in the propensities and costs; shifts in propensities and costs should be relatively modest to outperform an analysis with a non-informative prior. The corresponding turning point does depend on the variance of the informative prior and consequently hints at some form of moderation of historic/expert knowledge. The evaluation also shows that without misspecification the Bayesian analysis is to be favored to a "non-Bayesian" analysis, especially for smaller sample sizes of observed data. More generally, the evaluation provides an appreciation for the Bayesian approach and the need to carry out a sensitivity analysis.

What are the implications of these findings for prior elicitation? In general, it warrants sensitivity analyses. When priors are based on historic survey data, then careful consideration is needed of the timeliness and the amount of historic survey data that are available. Such a sensitivity may be partially overcome by moderating the strength of historic survey data over time (i.e., the more timely the data and knowledge, the more power is attached). Such moderation can be done using so-called power priors (Ibrahim and Chen 2000; Ibrahim, Chen, Gwon, and Chen 2015). However, moderation may also be achieved by adding a hierarchical level to the Bayesian models representing change in time, which comes at the cost of extra model parameters. In retrospective Bayesian analyses, we are currently investigating the use of moderation in time. When priors are based on expert knowledge, then it is to be recommended to vary the prior variance level (i.e., to explore the weight that is given to the experts).

We touched only briefly on the elicitation of prior distributions from expert knowledge. In models with many auxiliary variables, such elicitation may be difficult to conduct. Furthermore, data collection experts will generally not be able to provide values for slope and dispersion parameters in regression models but only for propensities and costs at the subgroup level. An effective elicitation of expert knowledge, therefore, will require some interpolation or proportional fitting of detailed models to marginal distributions that are given by experts. This tradeoff holds, especially for settings where priors are elicited from different but similar surveys. In order to develop effective prior elicitation procedures, we currently apply the Bayesian analysis framework to a broad set of case studies.

We see three conceptual limitations to our study that deserve future research and extension. First, although our model for monitoring of response and costs has general features, it does not fit all possible data collection designs and analyses, and particular designs and analyses may require adaptations of the model. In addition, we assumed normal distributions based on our dataset, although it is straightforward to specify the models according to the distributions found in the data. We believe that such changes are relatively modest given the exposition in this article. Second, we have not yet considered the (key) survey variables. Such variables may be modeled and monitored simultaneously, and design decisions may be based on a mix of overall quality and cost indicators and key survey estimates. Such an extension is fairly easy but does introduce new modeling choices because values of survey variables are unknown for nonrespondents. For this reason, we leave this extension to future research. Third, and strongly related to the previous point, we focused on nonresponse and have not yet considered strategy-dependent measurement biases. Such an extension is clearly worthwhile for multimode surveys.

Ultimately, the Bayesian analysis framework should support adaptive survey design decisions. Such an application means that historic survey data and expert knowledge should be comprised of multiple, possibly randomized, strategies and that observed data may be used to learn and update strategies for which information is weak or missing.

APPENDIX A: THE GIBBS SAMPLER AND CORRESPONDING FULL CONDITIONALS

The Gibbs sampler has the following steps:

(1) Set the random effects for the contact and participation equations to zero, $\delta_t^C = 0$ and $\delta_t^P = 0$, and fit regression models to all contact, participation, and cost equations and use the resulting estimated parameter values as starting values for the regression parameters $\alpha^C, \alpha^P, \gamma$.
(2) For each unit (i) in each phase (t), sample the latent variables $Z_{t,i}^C$ and $Z_{t,i}^P$ from $p\left(Z_{t,i}^C \middle| \alpha^C, \delta, u_{t,i}, x_i, \{s_{1,T}^i\}\right)$ and $p\left(Z_{t,i}^P \middle| \alpha^P, \delta, u_{t,i}, x_i, \{s_{1,T}^i\}\right)$.
(3) For each phase, sample the contact slope parameters α_t^C from $p\left(\alpha_t^C \middle| Z_t^C, \delta_t^C, x_i, \{s_{1,T}^i\}\right)$.
(4) Sample the random effects δ_t^C from $p\left(\delta_t^C \middle| Z_t^C, \alpha_t^C, x_i, \{s_{1,T}^i\}\right)$.
(5) For each phase, sample the participation slope parameters α_t^P from $p\left(\alpha_t^P \middle| Z_t^p, \delta_t^P, x_i, \{s_{1,T}^i\}\right)$.
(6) Sample the random effects δ_t^P from $p\left(\delta_t^P \middle| Z_t^P, \alpha_t^P, x_i, \{s_{1,T}^i\}\right)$.
(7) For the three cost components, sample the variance parameters σ^2 from $p(\sigma^2 | \gamma, c_0, c_R, c_{NR}, x, \{s_{1,T}^i\})$.
(8) For the three cost components, sample the slope parameters γ from $p(\gamma | \sigma^2, c_0, c_R, c_{NR}, x, \{s_{1,T}^i\})$.
(9) Return to step 1.

In the remainder of this appendix, we provide expressions for the full conditionals of the various regression model parameters for response and costs that are sampled in the Gibbs sampler.

A.1 Full conditionals for regression parameters in response propensity models.

Contact propensity and participation propensity models have the following form:

$$Z_{t,i}\left(s_{1,t}\right) = \alpha_t(s_t)x_{t,i} + \varepsilon_{t,i} + \delta_t\left(s_{1,t-1}\right),$$

where $Z_{t,i}(s_{1,t})$ is a latent variable, and $x_{t,i}$ is a column vector of baseline and paradata covariates of length m. The regression parameters are the slope parameters in the vector α_t and the random effect δ_t. Apart from these, also the latent variable $Z_{t,i}$ is updated in the Gibbs sampler, although it is not of direct interest. In a survey, we only observe whether $Z_{t,i}\left(s_{1,t}\right) > 0$ occurs or not. In the main text, the superscripts "C" and "P" are added to distinguish contact and participation models, but for the derivation of full conditionals, this distinction is not important; it is fully analogous.

In the following, the vector of random effects in phase t for all possible histories of actions, $s_{1,t-1}$, is denoted by δ_t. Obviously, each survey sample unit received just one treatment series of actions. We add the subscript i to indicate the strategy that was assigned to unit i (i.e., $s_{i,1,t}$ is the series of actions assigned to unit i in phases one to t). We let $\delta_{t,i}$ be the random effect that applies to unit i.

A.1.1 Slope parameters in contact model. For $\alpha_t(s)$, the prior distribution is normal $\alpha_t(s) \sim N(\mu(s), \Sigma(s))$. The full conditional distribution is also normal, and we denote it as,

$$(\alpha_t(s)|u_t, z_t, \delta_t, x, s_{1,t}) \sim N(\mu_{FULL}(s), \Sigma_{FULL}(s)). \tag{A1}$$

To derive the expectation and covariance of the full conditional distribution for action s, we need to restrict to sample units that reached phase t and for which $s_{i,t} = s$. Let this number be $n_t(s)$. For convenience, we label the units $i = 1,\ 2, \ldots, n_t(s)$. Let $z^*_{t,i} = z_{t,i} - \delta_{t,i}$ and let $z^*_t = (z^*_{t,\ 1}, \ldots, z^*_{t,n_t(s)})^T$. Furthermore, let X be the $n_t(s) \times m$ covariate matrix with sample units as rows.

It follows that the parameters in (A1) can be written as,

$$\Sigma_{FULL}(s) = \left((\Sigma(s))^{-1} + X^TX\right)^{-1}, \tag{A2}$$

$$\mu_{FULL}(s) = \Sigma_{FULL}(s)\left((\Sigma(s))^{-1}\mu(s) + X^Tz^*_t\right). \tag{A3}$$

A.1.2 Random effects. We will now derive the posterior distribution of $\delta_t(s_{1,t})$ for all possible strategies up to phase t, $s_{1,t}$. We assume that the prior distribution is the same for all strategies, $\delta_t(s_{1,t}) \sim N(0,\ \tau_t^2)$. Furthermore, we assume that the random effects are independent. Hence, the full conditional distribution depends only on the outcomes of all sample units that reached phase t and that have exactly the strategy $s_{1,t}$, i.e., $s_{1,t} = s_{i,1,t}$. Let there be $n_t(s_{1,t})$ such units, labeled $i = 1,\ 2, \ldots, n_t(s_{1,t})$. The full conditional distribution,

$$\left(\delta_t(s_{1,t}) \middle| \alpha_t(s_t), u_t, z_t, x, s_{1,t}\right) \sim N(\mu_{FULL}(s), \Sigma_{FULL}(s)). \tag{A4}$$

Let $\tilde{z}_{t,i}(s_{1,t}) = z_{t,i}(s_{1,t}) - \alpha_t(s_t)x_{t,i}$, then it follows that the parameters in (A4) can be written as,

$$\Sigma_{FULL}(s) = \left(\left(\tau_t^2\right)^{-1} + n_t(s_{1,t})\right)^{-1}, \tag{A5}$$

$$\mu_{FULL}(s) = \Sigma_{FULL}(s) \sum_{i=1}^{n_t(s_{1,t})} \tilde{z}_{t,i}(s_{1,t}). \tag{A6}$$

A.1.3 Latent response propensity. The last variables to update for the propensity models are the latent variables $Z_{t,i}$ for the sample units that reached phase t. It holds that $\left(Z_{t,i} \middle| \alpha_t, \delta_t, x, s_{1,t}\right) \sim N(\alpha_t(s_{i,t})x_i + \delta_{t,i}, 1)$ distributed. When $u_{t,i} = 1$, then $Z_{t,i} > 0$ and $\left(Z_{t,i} \middle| u_t, \alpha_t, \delta_t, x, s_{1,t}\right)$ is the normal distribution restricted to the positive real axis. For $u_{t,i} = 0$, $\left(Z_{t,i} \middle| u_t, \alpha_t, \delta_t, x, s_{1,t}\right)$ is the normal distribution restricted to the non-positive real axis.

There are no explicit expressions for this distribution. When the outcome is $u_{t,i} = 1$, draws from $N(\alpha_t(s_{i,t})x_i + \delta_{t,i}, 1)$ are repeated until a draw is positive. For $u_{t,i} = 0$, draws are repeated until a non-negative value is found.

A.2. Full conditionals for regression parameters in costs models.

We will derive the parameters of the posterior distributions of the parameters in the costs models. There are three such models: one for contact, one for participation, and one for refusal. The derivation of full conditionals is fully analogous so that we omit reference to the specific type of costs. The model has the form,

$$\mathrm{C}_i(s) = \gamma(s)x_i + \varepsilon_i(s), \tag{A7}$$

$$\varepsilon_i(s) \sim N\left(0, \sigma^2(s)\right), \tag{A8}$$

where x_i is the column vector of baseline covariates of length m and $\varepsilon_i(s)$ is action-dependent error term. In this article, the costs depend on the action that is applied but not on the phase in which it is applied. For actions that are applied in multiple phases, we therefore consider the average costs over all phases. The parameters that need to be updated and that require full conditional distributions are the slope parameters $\gamma(s)$ and the dispersion parameters $\sigma^2(s)$.

In updating regression parameters for a specific action s, we need to restrict ourselves to sample units that were treated by action s at least once during the survey. Let there be $n(s)$ such units. The observed c_i then is the average cost for the sample unit over all phases in which the action has been applied. We let c be the column vector of length $n(s)$ containing the values of the sample units.

A.2.1 Slope parameters. The prior distribution for the slope parameters is multivariate normal, $\gamma(s) \sim N\left(\mu(s), \Sigma^2(s)\right)$. The full conditional distribution is also normal, and we denote it as,

$$(\gamma(s)|c, x, \sigma^2) \sim N(\mu_{FULL}(s), \Sigma_{FULL}(s)). \tag{A9}$$

Let X be the $n(s) \times m$ covariate matrix with sample units as rows.

It follows for the parameters in (A9) that

$$\Sigma_{FULL}(s) = \left(\frac{1}{\sigma^2(s)} X^T X + \left(\Sigma^2(s)\right)^{-1}\right)^{-1}, \tag{A10}$$

$$\mu_{FULL}(s) = \Sigma_{FULL}(s)\left(\frac{1}{\sigma^2(s)} X^T c_R + \left(\Sigma^2(s)\right)^{-1} \mu(s)\right). \tag{A11}$$

A.2.2 Variance of error term. The prior distribution is inverse Gamma $\sigma^2(s) \sim \Gamma^{-1}(a(s), b(s))$. The full conditional is also inverse Gamma:

$$\left(\sigma^2(s)\middle|\gamma, c, x\right) \sim \Gamma^{-1}(a_{FULL}(s), b_{FULL}(s)). \tag{A12}$$

Given the notation introduced earlier, we have for the parameters in (A12),

$$a_{FULL}(s) = a(s) + \frac{n(s)}{2} \tag{A13}$$

$$b_{FULL}(s) = b(s) + \frac{1}{2}\ (c(s) - X\gamma(s))^T (c(s) - X\gamma(s)). \tag{A14}$$

APPENDIX B: ELICITATION OF HYPERPARAMETERS IN THE HEALTH SURVEY SIMULATION STUDY.

In the simulation study, we have probit regression models for contact and participation and linear regression models for costs. In phase one, there is only a contact model, whereas in phases two and three, there are also models for participation. Hence in total, there are ten models. For an informative prior, hyperparameters are needed for all regression coefficients in all models. We elicited informative priors by assuming that a historic Health survey data set of sample size $n = 10,000$ was available. Per type of regression coefficient, we explain how we proceeded in constructing priors.

Regression slope parameters in cost models are as follows: The slope parameters $\gamma_C(s)$ and $\gamma_R(s)$ are normally distributed, and a saturated model with variable AgeGender is applied. In a saturated linear model, each parameter is estimated using only the sample units in the corresponding population stratum. Consider the parameter γ_C for a particular stratum and a particular strategy. Based on historic data, the parameter is estimated as the average of the observed individual costs $c_{C,i}$ for sample units in the stratum that received the specified strategy. The average stratum costs are approximately normally distributed with expectation equal to the true γ_C and variance equal to $\frac{\sigma_R^2}{n}$. Given that we simulate data ourselves in this article, we can derive hyperparameters directly from the simulation model values.

Regression dispersion parameters in cost models are as follows: The dispersion parameters $\sigma_C^2(s)$ and $\sigma_R^2(s)$ have an inverse Gamma distribution and are constant over population strata. Consider σ_C^2 for a particular strategy. Given historic data, it is estimated as the sample variance over the observed individual costs $c_{C,i}$ for sample units that received the specified strategy. The sample variance divided over the true variance σ_C^2 and multiplied with $n-1$ is approximately χ_{n-1}^2 distributed. The expectation and variance of a χ_{n-1}^2 distribution are, respectively, $n-1$ and $2(n-1)$. This means that the sample variance has an expectation and variance equal to, respectively, σ_C^2 and $2\sigma_C^4/(n-1)$. An inverse Gamma distribution, $\Gamma^{-1}(\alpha, \beta)$, has expectation and variance equal to, respectively, $\beta/(\alpha-1)$ and $\beta^2/((\alpha-1)^2(\alpha-2))$. Hence, α and β can be derived as $\alpha = 2 + \frac{1}{2}(n-1)/\sigma_C^2$ and $\beta = \sigma_C^2 + \frac{1}{2}(n-1)$. Again, under the simulation model, these hyperparameters can be derived directly from the simulation values.

Regression slope parameters in contact/participation models are as follows: The elicitation of hyperparameters is analogous for contact and participation models. Because of the probit link function, there is no explicit expression for estimators for the regression slope parameters. Given that we include only main effects for baseline covariates, x_0, and paradata, x_1, it is therefore not straightforward how to choose hyperparameters based on historic data in an analytic way. For this reason, we simulated 2,000 datasets of size 10,000 and fitted probit regression models to each dataset. Over the 2,000 fitted vectors of parameters, means and variances were computed for single slope parameters and covariances for pairs of slope parameters. These means, variances and covariances were used as hyperparameters. Somewhat surprisingly, absolute covariances were sometimes quite large, especially between the slope parameters of the two vectors x_0 and x_1. Obviously, this approach can only be applied in a simulation study; for a real historic dataset, another approach is needed.

APPENDIX C: SIMULATION STUDY PROPENSITIES AND COSTS.

Tables C.1 and C.2 present the contact propensities, participation propensities and contact and participation costs, respectively, per phase and AgeGender × Break-off subgroup for the simulation study in Section 4.

Table C.1: Contact and Participation Propensities per Phase and Subgroup.

	Phase 1	*Phase 2*		*Phase 3*	
	Response	*Contact*	*Participation*	*Contact*	*Participation*
15-24, F, no break-off	0.30	0.70	0.58	0.50	0.30
15-24, F, break-off	0	0.70	0.87	0.50	0.59
15-24, M, no break-off	0.28	0.70	0.64	0.50	0.36
15-24, M, break-off	0	0.70	0.93	0.50	0.65
25-44, F, no break-off	0.33	0.74	0.49	0.62	0.22
25-44, F, break-off	0	0.74	0.94	0.62	0.66
25-44, M, no break-off	0.31	0.74	0.44	0.62	0.16
25-44, M, break-off	0	0.74	0.94	0.62	0.66
45-65, F, no break-off	0.35	0.80	0.38	0.90	0.10
45-65, F, break-off	0	0.80	0.95	0.90	0.72
45-65, M, no break-off	0.40	0.80	0.36	0.90	0.08
45-65, M, break-off	0	0.80	0.95	0.90	0.69

Table C.2: Contact and Participation Costs per Unit per Phase and Subgroup. Standard Deviations Are Given within Brackets.

	Phase 1	*Phase 2*		*Phase 3*	
	Response	*Contact*	*Participation*	*Contact*	*Participation*
15-24, F, no break-off	3 (1)	11 (1)	16 (1)	14 (1)	16 (1)
15-24, F, break-off	3 (1)	11 (1)	12 (1)	14 (1)	12 (1)
15-24, M, no break-off	3 (1)	12 (1)	14 (1)	15 (1)	14 (1)
15-24, M, break-off	3 (1)	12(1)	10 (1)	15 (1)	10 (1)
25-44, F, no break-off	3 (1)	10 (1)	21 (1)	13 (1)	21 (1)
25-44, F, break-off	3 (1)	10 (1)	19 (1)	13 (1)	19 (1)
25-44, M, no break-off	3 (1)	10 (1)	20 (1)	13 (1)	20 (1)
25-44, M, break-off	3 (1)	10 (1)	18 (1)	13 (1)	26 (1)
45-65, F, no break-off	3 (1)	8 (1)	26 (1)	11 (1)	18 (1)
45-65, F, break-off	3 (1)	8 (1)	21 (1)	11 (1)	21 (1)
45-65, M, no break-off	3 (1)	9 (1)	24 (1)	12 (1)	24 (1)
45-65, M, break-off	3 (1)	9 (1)	20 (1)	12 (1)	20 (1)

APPENDIX D: CONVERGENCE PROPERTIES OF THE GIBBS SAMPLER.

The Gibbs sampler produces a draw of a Markov chain that has the posterior distribution of interest as its stationary distribution. The Markov chain is initiated from one or more starting values. Under certain conditions, the Markov chain takes time to converge to its stationary distribution. For this reason, usually a burn-in period is discarded where the Markov chain has not yet reached its stationary distribution. After the burn-in period, the Markov chain moves through its parameter space at a certain "speed." This speed is termed the mixing property of the chain and determines the required length of the Markov chain (i.e., the number of iterations in the Gibbs sampler). Both the burn-period and the mixing of the Gibbs sampler cannot be determined with certainty, since the stationary distribution is unknown. However, various diagnostics have been developed to make an empirical assessment. We checked the burn-in period and convergence of the Gibbs sampler using the Raftery Lewis convergence diagnostic (Raftery and Lewis 1992) as implemented in the R package CODA. We required that the 2.5 percent-quantile of the posterior distribution could be approximated with a specified precision. We apply the Raftery and Lewis convergence diagnostic on the first 5,000 iterations to determine the number of iterations that are needed for convergence.

Somewhat surprisingly, we found for all scenarios that the burn-in period is very short and below 100 iterations. Nonetheless, we always discard the first 5,000 iterations. We then carried out the second part of iterations. The starting values here are the parameters obtained after 5,000 iterations in the first run. The number of iterations in the second run is derived from the Raftery and Lewis convergence diagnostic. Convergence was usually reached within 4,500 iterations, however we had some cases where up to 10,000 iterations were required. Computation times in R were roughly 5,000 iterations in twenty minutes. Figures D1, D2, and D3 show Gibbs sampler runs for regression slope parameters in phase one and three for contact model and phase two for participation model under the non-informative, true, and misspecified strong prior. The plots Gibbs iteration (I) are the first 1,000 runs from the originally discarded 5,000 runs under

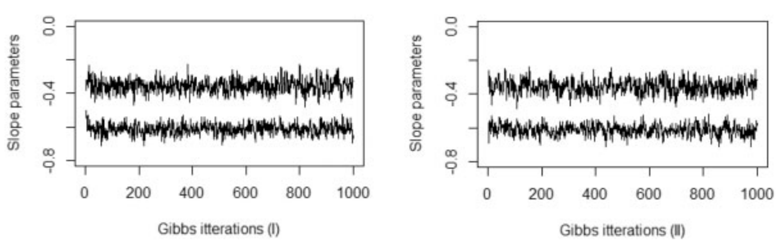

Figure D1. Gibbs Sampler Draws for the Phase One Contact Slope Parameters under a Non-informative Prior.

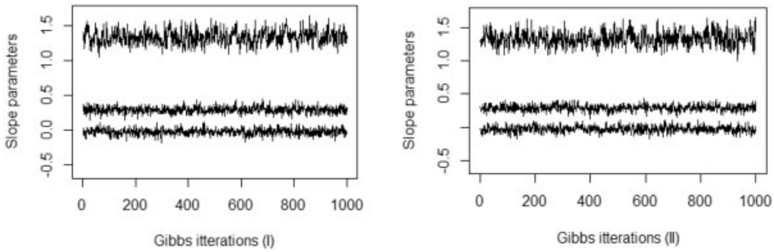

Figure D2. Gibbs Sampler Draws for the Phase Three Contact Slope Parameters under the True Prior.

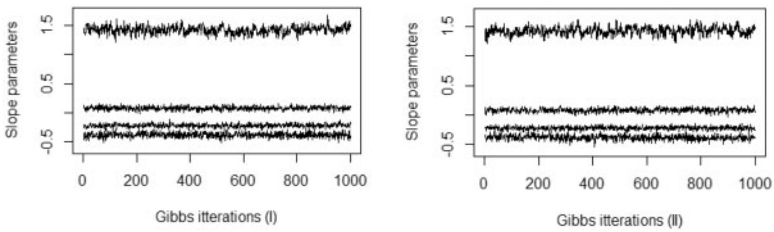

Figure D3. Gibbs Sampler Draws for the Phase Two Participation Slope Parameters for the Misspecified Prior Strong.

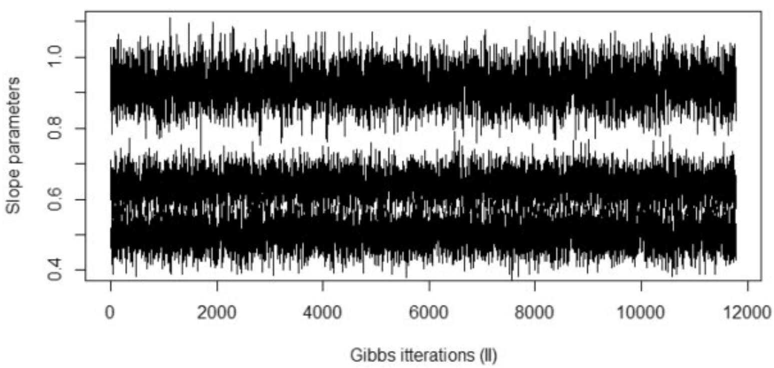

Figure D4. Gibbs Sampler Draws for the Phase Two Contact Slope Parameters under the Non-informative Prior.

the burn-in, and in the plots Gibbs iteration (II), the first 1,000 runs from the second set of runs to convergence. Figure D4 shows the regression parameters in phase two for contact model for the second run to convergence. Here, 10,000 iterations were required by the Raftery and Lewis convergence diagnostic.

REFERENCES

Albert, J. H., and S. Chib (1993), "Bayesian analysis of binary and polychotomous response data", *Journal of the American Statistical Association*, 88, 669–680.

Bethlehem, J., F. Cobben, and B. Schouten (2011), *Handbook of Nonresponse in Household Surveys*, Hoboken, NJ: John Wiley & Sons.

Burger, J., K. H. Perryck, and B. Schouten (2017) "Robustness of Adaptive Survey Designs to Inaccuracy of Design Parameter", Journal of Official Statistics, 33(3), 687–708.

Calinescu, M. (2013), "Optimal Resource Allocation in Adaptive Survey Designs," Ph.D. thesis, VU University Amsterdam, The Netherlands, available at https://research.vu.nl/en/publications/optimal-resource-allocation-in-adaptive-survey-designs (last accessed June 4, 2018).

Calinescu, M., and B. Schouten (2015), "Adaptive Survey Designs to Minimize Mode Effects. A Case Study on the Dutch Labour Force Survey", *Survey Methodology*, 41(2), 403–425.

CBS (2018), Korte beschrijving van de Gezondheidsenquête sinds 2010-2013 (in Dutch), Statistics Netherlands, De Haag, The Netherlands, available at https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/gezondheidsenquete-vanaf-2010-2013 (last accessed June 4, 2018).

Chakraborty, B., and S. A. Murphy (2014), "Dynamic Treatment Regimes", *Annual Review of Statistics and Its Applications*, 1(1), 447–464.

Coffey, S., B. Reist, and M. White (2013), "Monitoring Methods for Adaptive Design in the National Survey of College Graduates", JSM Proceedings, Survey Methods Research Section, Alexandria, VA: American Statistical Association, 3085–3099.

De Heij, V., B. Schouten, and N. Shlomo (2015), "RISQ 2.1 Manual. Tools in SAS and R for the Computation of R-indicators, Partial R-indicators and Partial Coefficients of Variation", Statistics Netherlands, Den Haag, The Netherlands, available at www.risq-project.eu (last accessed June 4, 2018).

Durrant, G. D., J. D'Arrigo, and F. Steele (2011), "Using Paradata to Predict Best Times of Contact, Conditioning on Household and Interviewer Influences", *Journal of the Royal Statistical Society A*, 174(4), 1029–1049.

Durrant, G. D., and F. Steele. (2013), "Analysing Interviewer Call Record Data by Using a Multilevel Discrete Time Event History Modelling Approach", *Journal of the Royal Statistical Society A*, 176(1), 251–269.

Durrant, G. D., O. Maslovskaya, and P. W. F. Smith (2015), "Modelling Final Outcome and Length of Call Sequence to Improve Efficiency in Interviewer Call Scheduling", *Journal of Survey Statistics and Methodology*, 3(3), 397–424.

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models", *Bayesian Analysis*, 1(3), 515–534.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014), *Bayesian Data Analysis* (2nd ed.), London: Chapman and Hall.

Groves, R. M. (1989), *Survey Errors and Survey Costs*, New York, NY: John Wiley & Sons.

Groves, R. M., and S. G. Heeringa (2006), "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs", *Journal of the Royal Statistical Society: Series A*, 169(3), 439–457.

Heyd, J. M., and B. Carlin (1999), "Adaptive Design Improvements in the Continual Reassessment Method for Phase I Studies", *Statistics in Medicine*, 18(11), 1307–3121.

Ibrahim, J. G., and M. H. Chen (2000), "Power Prior Distributions for Regression Models", *Statistical Science* 15(1), 46–60.

Ibrahim, J. G., M. H. Chen, Y. Gwon, and F. Chen (2015), "The Power Prior: Theory and Applications", *Statistics in Medicine*, 34(28), 3724–3749.

Kreuter, F. (2013), *Improving Surveys with Paradata. Analytic Uses of Process Information*, Edited monograph, Hoboken, NJ: John Wiley and Sons.

Laber, E. B., D. J. Lizotte, M. Qian, W. E. Pelham, and S. A. Murphy (2014), "Dynamic Treatment Regimes: Technical Challenges and Applications", *Electronic Journal of Statistics*, 8(1), 1225–1272.

Lei, H., I. Nahum-Shani, K. Lynch, D. Oslin, and S. A. Murphy (2012), “A SMART Design for Building Individualized Treatment Sequences”, *Annual Review of Clinical Psychology*, 8(1), 21–48.

Luiten, A., and B. Schouten (2013), “Adaptive Fieldwork Design to Increase Representative Household Survey Response: A Pilot Study in the Survey of Consumer Satisfaction”, *Journal of Royal Statistical Society, Series A*, 176(1), 169–190.

Murphy, S. A. (2003), “Optimal Dynamic Treatment Regimes”, *Journal of the Royal Statistical Society: Series B*, 65(2), 331–355.

Nishimura, R., J. Wagner, and M. R. Elliott (2016), “Alternative Indicators for the Risk of Nonresponse Bias. A Simulation Study”, *International Statistical Review*, 84(1), 43–62.

O’Quigley, J., and L. Z. Shen (1996), “Continual Reassessment Method: A Likelihood Approach”, *Biometrics*, 52(2), 673–684.

Peytchev, A., S. Riley, J. Rosen, J. Murphy, and M. Lindblad (2010), “Reduction of Nonresponse Bias in Surveys through Case Prioritization”, *Survey Research Methods*, 4, 21–29.

Raftery, A. E., and S. Lewis (1992), “How Many Iterations in the Gibbs Sampler”, in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, pp. 763–773, Offord: Oxford University Press.

Rosen, J. A., J. J. Murphy, A. Peytchev, T. E. Holder, J. A. Dever, D. R. Herget, and D. J. Pratt (2014), “Prioritizing Low-propensity Sample Members in a Survey: Implications for Nonresponse Bias”, *Survey Practice*, 7(1), 1–8.

Särndal, C. E., and P. Lundquist (2014), “Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation”, *Journal of Survey Statistics and Methodology*, 2(4), 361–387.

Schafer, J. L. (2013), “Bayesian Penalized Spline Models for Statistical Process Monitoring of Survey Paradata Quality Indicators”, in *Improving Surveys with Paradata. Analytic Uses of Process Information*, ed. F. Kreuter, pp. 311–340, New YorK: Wiley.

Scharfstein, D. O., M. J. Daniels, and J. M. Robins (2003), “Incorporating Prior Beliefs about Selection Bias into the Analysis of Randomized Trials with Missing Outcomes”, *Biostatistics*, 4(4), 495–512.

Schouten, B., M. Calinescu, and A. Luiten (2013), “Optimizing Quality of Response through Adaptive Survey Designs”, *Survey Methodology*, 39(1), 29–58.

Schouten, J. G., F. Cobben, and J. Bethlehem (2009), “Indicators for the Representativeness of Survey Response”, *Survey Methodology*, 35(1), 101–113.

Schouten, B., and N. Shlomo (2017), “Selecting Adaptive Survey Design Strata with Partial R-Indicators”, *International Statistical Review*, 85(1), 143–163.

Schulte, P. L., A. A. Tsiatis, E. B. Laber, and M. Davidian (2014), “Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes”, *Statistical Science*, 29(4), 640–661.

Wagner, J. (2008), “Adaptive Survey Design to Reduce Nonresponse Bias,” Ph.D. thesis, University of Michigan, Ann Arbor.

Wagner, J., B. T. West, N. Kirgis, J. M. Lepkowski, W. G. Axinn, and S. Kruger Ndiaye (2013), “Use of Paradata in a Responsive Design Framework to manage a Field Data Collection”, *Journal of Official Statistics*, 28(4), 477–499.