Adaptive survey designs accounting for uncertainty and gradual change in survey design parameters

Barry Schouten
Statistics Netherlands, The Hague, The Netherlands – bstn@cbs.nl

Nino Mushkudiani
Statistics Netherlands, The Hague, The Netherlands – nmsi@cbs.nl

1. Introduction

Over the last two decades, there has been a strongly increasing interest in survey data collection monitoring, analysis, and intervention or adaptation. The main causes for this are the diversification of data collection that followed the emergence of online communication, the lack of predictability of survey response rates 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) see Kreuter (2013).

The lack of predictability of data caused different streams of data collection. One of these consists of adaptive or responsive survey designs that adapt or tailor strategies and effort to known and relevant characteristics of sampled units from the target population, see Groves and Heeringa (2006), Wagner (2008) and Schouten, Calinescu and Luiten (2013). In order to adapt, accurate estimates of survey design parameters are not just needed at the overall population level, but also at the deeper level of population subgroups.

Adaptive survey design leans heavily on estimated survey design parameters like contact propensities, participation propensities and costs per sample unit. Such parameters are needed to determine effective data collection strategies and to optimize survey quality under constraints on costs. Estimates for survey design parameters may, however, be inaccurate due to sampling variation and gradual change in time.
A Bayesian analysis of survey data collection may be profitable when expert knowledge and/or historic survey data from the same or similar surveys are available. This knowledge and data may then be employed to set informative prior distributions to coefficients in regression models for survey design parameters and for survey variable outcomes. During or after data collection posterior distributions may be derived for the same parameters, but also for overall quality and cost measures. Even when survey design parameters change gradually in time or change from one survey to the other, including such informative priors in the model deduce posterior distributions that are more informative than without the prior knowledge.
We demonstrate how a Bayesian analysis may be implemented and analyzed in monitoring survey data collection. Furthermore, we discuss the optimization and adaptation of survey design using the posterior distributions for survey design parameters, and quality and cost measures. We do so using two case studies. In the studies the choice of survey modes plays an important role.

2. Adaptive strategies and design parameters

To set up a general model that would describe all possible data collection designs in relation to available covariates from frame data, administrative data and paradata, is too complicated a goal. Here we intend to present models that have many features of designs but in their simplest forms. The included features are more than one data collection phase, baseline covariates as well as paradata, cost functions, and dependence on actions/decisions in earlier phases.

First we introduce some notations. Let the survey design consist of a maximum of T phases that are labelled $t = 1,2, \dots, T$. We define $\mathcal{S}_t$ as the collection of all possible actions in phase t and let s_t represent the action in phase t. For different phases, the collections of actions may be different. The action sets may contain $s_{\emptyset}$, which, if selected, implies that no attempt is made to obtain a response. We define the collection of survey strategies

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

and let $s_{1,T} \in \mathcal{S}_{1,T}$ denote one possible strategy, i.e. sequence of actions. For a strategy $s_{1,T}$, we denote the actions in phase i til j by the vector $s_{i,j}$.

For a subject i, we let x_i be the vector of auxiliary variables that is linked from frame data, administrative data or paradata, x_i consists of the following entries

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

where $x_{0,i} = (x_{0,1,i}, \dots, x_{0,m_0,i})'$ contains the auxiliary variables available at the start of data collection, and $x_{t,i} = (x_{t,1,i}, \dots, x_{t,m_t,i})'$ are the auxiliary variables that are observed for the fielded sample units in phase t. In the optimization of the adaptive survey designs (ASD), actions in phase t can only be chosen based on $x_{0,i}$ to $x_{t-1,i}$.

The design of each survey has a range of features, e.g. advance letter, contact protocol, screener interview, number of phases, reminder protocol, use of incentive, mode of administration (web, telephone, face-to-face, mail), interviewer, refusal conversion procedure and type of questionnaire. The total of choices made for the design features (e.g. incentive, phases, first web mail then telephone interview) will define the data collection strategy or simply strategy. In non-adaptive surveys, these features are implemented uniformly over the whole sample. In adaptive surveys, part of the design features may be implemented differently for different sample units, i.e. there is a set of strategies, see Groves and Heeringa (2006), Wagner (2008), Coffey, Reist and White (2013).

ASD 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. Three sets of survey design parameters suffice to compute most of the quality and cost constraints:

1. Response propensities per unit $\rho_i(s_{1,T})$ per strategy;
2. Expected costs per sample unit $C_i(s_{1,T})$ per strategy;
3. Adjusted mode effects per unit $D_i(s_{1,T})$ per strategy;

In this paper we restrict attention to nonresponse error. The last set of parameters, the adjusted mode effects, are not considered here. There are two options in defining and modeling survey design parameters. Design parameters can be detailed to subgroups or to individual cases. Below we focus on individual design parameters.

3. Modeling survey design parameters

We introduce basic models for response propensities and costs. Therefore, we break down these parameters into their basic components, like the contact and participation propensities. For these basic components we will, first, make some general assumptions. 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.

Some side remarks are in place. Below we introduce prior distributions for parameters that are shared by multiple design parameters. As a result, the propensities and costs of sample units may become dependent because they share the same underlying parameters. The assumptions then read as independence given the values of these parameters and likelihood functions can still be factorized as the product of individual likelihoods conditional on the parameters. Second, the three assumptions essentially ignore any impact of scale on data collection; it is assumed that there is a stable workload. Importantly, we allow for associations between contact propensities over phases, between participation propensities over phases, and between cost functions over phases.

We define our approach only for contact propensity, a part of response propensity. Models for the response propensity and cost functions can be defines similarly.

Let $\kappa_{t,i}(s_{1,t})$ be the propensity of a contact in phase t under strategy $s_{1,t}$ given that the unit did not respond in earlier phases and is eligible for follow-up. We assume that design features in subsequent phases have no impact on making contact. The outcome(s) of the previous phase(s) can be included in the auxiliary vector, when contact propensities are considered to be dependent on whether there was a noncontact or a refusal. $\lambda_{t,i}(s_{1,t})$ is the propensity of a participation in phase t of subject i under strategy $s_{1,t}$ given contact (and given that the unit did not respond in earlier phases and is eligible for follow-up). Then the response propensity in phase t of a subject i under strategy $s_{1,t}$, $\rho_{t,i}(s_{1,t})$, is

$$\rho_{t,i}(s_{1,t}) = \kappa_{t,i}(s_{1,t}) \cdot \lambda_{t,i}(s_{1,t}).$$

When in subsequent phases all nonresponse receives a follow-up, then

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

We model the propensities using a probit model, i.e. using a binomial link function. Each sample unit has a certain contactability represented as a latent variable $Z_{t,i}^{C}(s_{1,t})$ and contact is obtained when this latent variable is larger than zero and $Z_{t,i}^{C}(s_{1,t}) \sim N(\mu(s_{1,t}), \sigma(s_{1,t}))$, for some $\mu_{t,i}(s_{1,t}), \sigma_{t,i}(s_{1,t})$ so that

$$\kappa_{t,i}(s_{1,t}) = P(Z_{t,i}^{C}(s_{1,t}) > 0).$$

For $m \le m_k$, let $\alpha_{t,k,m}(s_{1,t})$ 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, $\alpha_{t,k,m}(s_{1,t}) = 0$, when $k > t$. The model could be written as

$$Z_{t,i}^{C}(s_{1,t}) = \sum_{k=0}^{t} \alpha_{t,k}(s_{1,t}) x_{k,i} + \varepsilon_{t,i}^{C},$$

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

To be able to include dynamic adaptive survey designs, we need to include paradata. However to keep the model simple, we assume that 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 consider the dependence on past actions. It is unrealistic to assume there is no such dependence in most settings. Past actions could be included by introducing a fixed or random effect per possible history. We add the history as a random effect here. Third, since we suggest to add a dependence on the history of actions as a random effect, the regression coefficients become necessarily dependent on the phase and not on the past. The model becomes

$$Z_{t,i}^{C}(s_{1,t}) = \begin{cases} \alpha_{t,0}(s_t) x_{0,i} + \varepsilon_{t,i}^{C} + \delta_t^{C}(s_{1,t-1}), & t \le t_1, \\ \alpha_{t,0}(s_t) x_{0,i} + \alpha_{t,1}(s_t) x_{t_1,i} + \varepsilon_{t,i}^{C} + \delta_t^{C}(s_{1,t-1}), & t > t_1, \end{cases} \quad (1)$$

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

The analysis become Bayesian by assigning prior distributions to the regression coefficients and random effects in (1). Our aim is the derivation of the posterior distributions of the individual response propensities $\rho_i(s_{1,T})$ and the individual cost parameters $C_i(s_{1,T})$ per strategy given observed data. These overall parameters are, in general, complex functions of the underlying survey design parameters per phase. We derived expressions for the posterior distributions of the regression coefficients and random effects when it was possible, otherwise derived these numerical approximations and applied Markov Chain Monte Carlo methods to generate draws from the posterior distributions.

Data collection may apply randomization in order to learn about multiple strategies simultaneously. Here, we assume that the observed data may contain randomization over strategies but that randomization is only at the outset. Hence, strategy allocation probabilities may depend on auxiliary information known at the start of data collection, but not on paradata coming in during data collection. So in addition to the outcomes, costs and auxiliary vectors, we observe the series of actions, or simply strategy, that were applied per sample unit $s_{1,T}^{i}$;

In the following, we use $\rho(s_{1,T})$ and $C(s_{1,T})$ 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. Note that x may in fact be a matrix, when the auxiliary variables are a vector by themselves. With $\{s_{1,T,i}\}$ we denote the vector of used strategies for all sample units. To shorten expressions, we use $\alpha, \beta, \delta, \gamma, \sigma^2$ for the vectors of regression slope parameters, random effects and regression dispersion parameters over phases and actions, but elaborate when needed. For the sake of 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. A straightforward solution is to perform a Gibbs sampler to the joint density of the regression parameters $\alpha, \beta, \delta, \gamma, \sigma^2$

$$p(\alpha, \beta, \delta, \gamma, \sigma^2 | u_t, c_0, c_R, c_{NR}, x, \{s_{1,T}^i\}). \quad (2)$$

A Gibbs sampler for (2) requires repeated draws from the conditional densities of each regression parameter given the observed data and the other regression parameters, the so-called full conditionals. Literature provides a range of options to sample from these conditional distributions, see Albert and Chib (1993) and Gelman et al (2003).

In the 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 for the sake of brevity, the response rate, the total costs and the coefficient of variation of the response propensities; the analysis of other functions can often be done in an analogous way.

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

$$RR(s_{1,T}) = \frac{1}{N} \sum_{i=1}^{n} d_i \rho_i\, (s_{1,T}), \quad (3)$$

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

$$B(s_{1,T}) = \sum_{i=1}^{n} c_i\, (s_{1,T}), \quad (4)$$

and the coefficient of variation, *CV*, is

$$CV(X, s_{1,T}) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{n} d_i (\rho_i(s_{1,T}) - RR(s_{1,T}))^2}}{RR(s_{1,T})}. \quad (5)$$

For the *CV*, 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 response rate and total costs do not depend on the choice of X. Obviously, 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. They have even more complex forms than the individual response propensities and cost parameters. However, they can again be approximated as a by-product of the Gibbs sampler. For every draw of the individual response propensities and cost parameters, we compute (3) to (5).

4. A simulation study

In the simulation study, we investigate the impact of prior distribution specification and of survey sample size on the shape of posterior distributions. Furthermore, we explore the convergence properties of the Gibbs sampler.

4.1 Design of the simulation study

To evaluate the utility of a Bayesian analysis is, 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 that we compare to are 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 and we view this choice as a “non-Bayesian” analysis, despite the use or prior and posterior distributions. Thus, we still make use of the benefit of a Bayesian analysis in that it 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 10000, 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 and/or a 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.
We quantify this gain by the root mean square error (RMSE) of the posterior distribution relative to the simulation model values. Let $p_\pi(\Theta|u_t, c_0, c_R, c_{NR}, x, \{s_{1,T}^i\})$ 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

$$\mathrm{RMSE}(\Theta;\pi) = \sqrt{(E_{p_\pi}(\Theta) - \Theta_0)^2 + \mathrm{var}_{p_\pi}(\Theta)}, \qquad (6)$$

where Θ_0 is the simulation model value.

We base our simulation study on the 2015 Dutch Health Survey (HS). The HS has a sequential mixed-mode survey design with 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 1 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 1250, 2500, 5000 and 10000. The simulation probabilities and costs are given in appendix C. To model contact and participation, we use a probit regression with GenderAge in phase 1 and GenderAge + BreakOff in phases 2 and 3. For phase 1, online data collection, we set participation propensities equal to response and participation costs are set to zero. We do this, because for online surveys costs are only associated with contact and not with interview. For phases 2 and 3, 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.

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.

	Data	*Web*	*F2F short*	*F2F extended*
RR	True	30.2%	57.6%	60.5%
	Misspecified light	32.2%	57.2%	59.7%
	Misspecified medium	35.2%	56.8%	58.8%
	Misspecified strong	40.2%	56.8%	58.2%

CV	True	0.277	0.069	0.102
	Misspecified light	0.260	0.061	0.094
	Misspecified medium	0.238	0.049	0.082
	Misspecified strong	0.208	0.036	0.063
B	True	3.0	15.2	19.4
	Misspecified light	3.0	14.5	19.5
	Misspecified medium	3.0	13.6	19.8
	Misspecified strong	3.0	12.1	20.3

Misspecification was introduced by shifting contact and participation propensities for each subgroup in the same direction. For the online phase 1 they were increased by 2%, 5% and 10%. For the F2F phases, they were decreased by 2%, 5% and 10%. Hence, we mimic an overestimation of online response and an underestimation of subsequent F2F response, which essentially leads to an underestimation of required budget. Table 1 contains the expected response rates, coefficients of variation and costs based on the sets of misspecified priors.

4.3 Simulation results

We discuss wo comparisons to evaluate the utility of the Bayesian analysis: 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 fully non-informative prior. We view the non-informative prior as a non-Bayesian analysis benchmark.

Table 2 shows the RMSE values for the two priors for four sample sizes: 1250, 2500, 5000 and 10000 units. Three variance levels for the misspecified priors are chosen, corresponding to a historic data set of a modest size of 1250 units (V1), a moderate size of 2500 units (V2) and a large size of 10000 units (V3). 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 and/or the true prior variance decreases. However, for a sample size of 10000, 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 the both prior variance and the observed data sample size are smallest. The biggest gap in RMSE is indeed found for a prior with variance V3 and sample size 1250. The RMSE values of this combination are comparable to that of the non-informative prior with sample size 10000. In the analysis, we consider the population as a whole. However, once subpopulations are of interest and statistics are detailed to such subpopulations, then, obviously, sample sizes get smaller and the prior distributions will still have added value. Table 2 should then be evaluated as the sample sizes of such subpopulations.

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

Table 2: RMSE 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 1250, 2500, 5000 and 10000. The true priors have a variance corresponding to 1250 (V1), 2500 (V2) and 10000 (V3) historic sample units.

Size	*Prior*	*RR*	*CV*	*B*

		Web	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*
1250	Non-informative	0.014	0.019	0.015	0.046	0.045	0.037	0.010	0.316	0.374
	True V1	0.010	0.012	0.010	0.021	0.023	0.018	0.010	0.218	0.273
	True V2	0.008	0.009	0.008	0.014	0.015	0.012	0.010	0.178	0.223
	True V3	0.004	0.005	0.005	0.007	0.008	0.008	0.010	0.116	0.142
2500	Non-informative	0.010	0.010	0.010	0.012	0.055	0.041	0.009	0.239	0.298
	True V1	0.008	0.008	0.008	0.010	0.035	0.025	0.009	0.204	0.247
	True V2	0.007	0.007	0.007	0.008	0.025	0.017	0.009	0.181	0.217
	True V3	0.004	0.004	0.004	0.006	0.009	0.007	0.009	0.128	0.148
5000	Non-informative	0.006	0.007	0.007	0.009	0.023	0.016	0.009	0.156	0.183
	True V1	0.006	0.006	0.006	0.008	0.018	0.013	0.008	0.143	0.166
	True V2	0.005	0.006	0.006	0.007	0.015	0.010	0.008	0.134	0.157
	True V3	0.004	0.004	0.004	0.005	0.007	0.006	0.009	0.105	0.121
10000	Non-informative	0.005	0.005	0.005	0.007	0.008	0.009	0.010	0.120	0.135
	True V1	0.004	0.005	0.005	0.006	0.008	0.009	0.009	0.114	0.130
	True V2	0.004	0.005	0.005	0.006	0.008	0.010	0.009	0.111	0.125
	True V3	0.003	0.004	0.004	0.005	0.008	0.009	0.010	0.010	0.108

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 a non-Bayesian analysis benchmark.

Table 3 contains the RMSE values for non-informative and misspecified priors estimated using the Gibb sampler. Again, we have chosen three variance levels, corresponding to a historic dataset of 1250 (V1), 2500 units (V2) and 10000 units (V3). Furthermore, we evaluate four sample sizes: 1250, 2500, 5000 and 10000. Recall from table 1 that, for phase 1, 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 2 and 3 are affected only little, but the coefficient of variation is underestimated. The cumulative costs after phase two are underestimated, but after phase 3 they are slightly overestimated.

The main observation from the RMSE values is that a misspecified prior can outperform a non-informative prior, but misspecification should be modest and/or the variance of the prior should be 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% 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% and 10% shifts under misspecified medium 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.

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; misspecified knowledge beats no knowledge. 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%.

Table 3: RMSE 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 1250, 2500, 5000 and 10000. The misspecified priors have a variance corresponding to 1250 (V1), 2500 (V2) and 10000 (V3) historic sample units.

Size	*Prior*	*RR*			*CV*			*B*		

		Web	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*	*Web*	*F2F*	*F2FE*
1250	Non-informative	0.014	0.019	0.015	0.046	0.045	0.037	0.010	0.316	0.374
	Missp light V1	0.012	0.014	0.012	0.019	0.020	0.016	0.010	0.398	0.265
	Missp light V2	0.014	0.011	0.011	0.012	0.013	0.012	0.010	0.459	0.284
	Missp light V3	0.018	0.007	0.009	0.006	0.013	0.015	0.010	0.549	0.356
	Missp medium V1	0.023	0.026	0.026	0.016	0.020	0.016	0.010	1.061	0.719
	Missp medium V2	0.032	0.027	0.030	0.010	0.014	0.015	0.010	1.360	0.964
	Missp medium V3	0.044	0.029	0.036	0.006	0.023	0.028	0.010	1.767	1.295
	Missp strong V1	0.046	0.010	0.010	0.013	0.015	0.016	0.010	1.346	0.689
	Missp strong V2	0.063	0.008	0.008	0.008	0.019	0.026	0.010	1.768	0.942
	Missp strong V3	0.087	0.005	0.005	0.008	0.033	0.047	0.010	2.324	1.281
2500	Non-informative	0.010	0.010	0.010	0.012	0.055	0.041	0.009	0.239	0.298
	Missp light V1	0.008	0.009	0.008	0.009	0.032	0.024	0.009	0.213	0.196
	Missp light V2	0.010	0.008	0.008	0.008	0.021	0.015	0.009	0.286	0.203
	Missp light V3	0.030	0.008	0.005	0.006	0.010	0.014	0.009	0.706	0.543
	Missp medium V1	0.014	0.015	0.017	0.009	0.032	0.022	0.009	0.619	0.441
	Missp medium V2	0.022	0.019	0.022	0.008	0.019	0.012	0.009	0.959	0.702
	Missp medium V3	0.053	0.015	0.022	0.005	0.015	0.022	0.009	1.787	1.379
	Missp strong V1	0.029	0.008	0.008	0.009	0.023	0.015	0.009	0.819	0.440
	Missp strong V2	0.045	0.007	0.007	0.009	0.012	0.012	0.009	1.251	0.682
	Missp strong V3	0.077	0.008	0.004	0.009	0.028	0.037	0.009	2.065	1.151
5000	Non-informative	0.006	0.007	0.007	0.009	0.023	0.016	0.009	0.156	0.183
	Missp light V1	0.007	0.006	0.006	0.008	0.017	0.012	0.009	0.164	0.155
	Missp light V2	0.009	0.006	0.006	0.008	0.013	0.009	0.009	0.217	0.168
	Missp light V3	0.014	0.005	0.006	0.006	0.008	0.009	0.009	0.399	0.267
	Missp medium V1	0.011	0.009	0.010	0.009	0.016	0.011	0.009	0.407	0.308
	Missp medium V2	0.017	0.012	0.014	0.009	0.012	0.009	0.009	0.669	0.504
	Missp medium V3	0.033	0.021	0.027	0.008	0.013	0.016	0.009	1.323	0.994
	Missp strong V1	0.020	0.006	0.006	0.009	0.012	0.009	0.009	0.520	0.300
	Missp strong V2	0.032	0.006	0.006	0.010	0.010	0.011	0.009	0.864	0.488
	Missp strong V3	0.065	0.007	0.004	0.010	0.027	0.031	0.009	1.736	0.984
10000	Non-informative	0.005	0.005	0.005	0.007	0.008	0.009	0.010	0.120	0.135
	Missp light V1	0.005	0.005	0.005	0.007	0.009	0.010	0.009	0.106	0.116
	Missp light V2	0.005	0.005	0.005	0.006	0.009	0.011	0.009	0.128	0.116
	Missp light V3	0.010	0.005	0.006	0.006	0.012	0.013	0.009	0.287	0.196
	Missp medium V1	0.006	0.008	0.008	0.007	0.009	0.010	0.009	0.216	0.176
	Missp medium V2	0.010	0.010	0.011	0.007	0.010	0.011	0.010	0.386	0.297
	Missp medium V3	0.024	0.019	0.023	0.007	0.016	0.018	0.009	0.988	0.752
	Missp strong V1	0.011	0.005	0.005	0.007	0.010	0.013	0.009	0.275	0.169
	Missp strong V2	0.019	0.005	0.005	0.007	0.013	0.016	0.009	0.500	0.287
	Missp strong V3	0.048	0.004	0.004	0.009	0.026	0.029	0.009	1.288	0.739

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.

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 given when data collection starts and auxiliary variables that become available during data collection, 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. The computation times of the Gibbs sampler are manageable and sufficiently short to run overnight for a range of scenarios.

The most important objective is to show the added value of a Bayesian analysis. In the evaluation, we viewed a fully non-informative prior as representing, essentially, a non-Bayesian analysis. In order, to be able to compare, we remain in the Bayesian framework of prior and posterior distributions. The evaluation is based on a simulation study using realistic contact propensities and costs, and participation propensities and costs from a multi-mode 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.

Some limitations to our study: 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. However, we believe that such changes are relatively straightforward given the exposition in this paper. 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 to include, see Schouten, Bruin and Mushkudiani (2016), but does introduce new modeling choices because values of survey variables are unknown for nonrespondents. Third, and strongly related to the previous point, we focused on nonresponse and have not yet considered strategy-dependent measurement biases. In multi-mode surveys, such an extension and broader look is inevitable.

The findings of this paper point at a sensitivity of Bayesian analyses to misspecification in prior distributions. Such a sensitivity may be partially overcome by moderating the strength of historic survey data and expert knowledge 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 and Ibrahim et al 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.

We briefly touched on the elicitation of prior distributions from historic survey data and 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, likely requires some interpolation or proportional fitting of detailed models to marginal distributions that are given by experts. This trade-off holds, especially, for settings where priors are elicited from different, but similar, surveys. When prior distributions are based on historic data from the same survey, then models may be fitted directly.

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