

Responsive and Adaptive Design for Survey Optimization

*Asaph Young Chun*¹, *Steven G. Heeringa*², and *Barry Schouten*³

We discuss an evidence-based approach to guiding real-time design decisions during the course of survey data collection. We call it responsive and adaptive design (RAD), a scientific framework driven by cost-quality tradeoff analysis and optimization that enables the most efficient production of high-quality data. The notion of RAD is not new; nor is it a silver bullet to resolve all the difficulties of complex survey design and challenges. RAD embraces precedents and variants of responsive design and adaptive design that survey designers and researchers have practiced over decades. In this paper, we present the four pillars of RAD: survey process data and auxiliary information, design features and interventions, explicit quality and cost metrics, and a quality-cost optimization tailored to survey strata. We discuss how these building blocks of RAD are addressed by articles published in the 2017 JOS special issue and this special section. It is a tale of the three perspectives filling in each other. We carry over each of these three perspectives to articulate the remaining challenges and opportunities for the advancement of RAD. We recommend several RAD ideas for future research, including survey-assisted population modeling, rigorous optimization strategies, and total survey cost modeling.

Key words: Responsive design; adaptive design; survey errors; survey costs; optimization; paradata; auxiliary data; total survey error.

1. Introduction

For over a century, survey researchers have faced continual change in the essential conditions under which surveys are designed and data collection is conducted. Time and again, survey practitioners have “responded” or “adapted” to arising challenges and opportunities they faced using innovations in the statistical design and analysis of their studies or in the methods used to collect study data. Many of the changes in the essential conditions of surveys have been clearly recognized as transformational in the field: the idea that samples could represent a population; a theory of inference based on probability samples; the formalization of household sampling and survey methods; the advent of RDD telephone surveys; computerization of survey design, data collection and analysis;

¹ U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C., U.S.A. Email: Asaph.Young.Chun@census.gov.

² University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, U.S.A. Email: sheering@umich.edu.

³ Statistics Netherlands, Henri Faasdreef 312, CBS-weg 11, 2492 JP The Hague, The Netherlands. Email: jg.schouten@cbs.nl.

Acknowledgments: We would like to extend our thanks to the contributors to this Special Section on Responsive and Adaptive Design for their dedicated work and to all peer reviewers for their critical review. We are also so grateful to the JOS editorial staff who provided essential support.

Disclaimer: The views expressed are the authors’ and not necessarily those of the institutions with which the authors are affiliated.

and internet and web-based data collection. In addition to benefitting from these transformational opportunities, survey researchers have been increasingly challenged by several inter-related and chronic trends – the increasing number and complexity of surveys, rising costs of traditional survey designs and methods, and steady decline in respondent participation rates.

In light of these changes and opportunities to seize, the remainder of Section 1 describes the genesis and evolution of responsive and adaptive design (RAD), including the four pillars of RAD that are essential to survey design. Section 1 closes by presenting the overarching research questions addressed by articles published in the 2017 JOS special issue and this special section. Section 2 presents three perspectives, discussing how core elements of RAD are addressed by these articles. Section 3 outlines what open research questions and challenges remain, following each of three perspectives as presented in Section 2. Sections 4 and 5 include a discussion of future directions for RAD and conclusions. This discussion article is a companion to the article by [Chun et al. \(2017\)](#) providing an overview of the 2017 JOS special issue devoted to RAD.

1.1. The Notion and Evolution of Responsive and Adaptive Design

The idea of responsive and adaptive design (RAD) arose in response to the survey challenges that were enumerated at the outset of our discussion. By way of definition, RAD is a data-driven scientific approach to controlling survey design features in real-time data collection by monitoring explicit costs and errors of survey estimates that are informed by auxiliary information, paradata, and multiple sources of data; RAD works toward a goal of survey optimization based on cost-error tradeoff analysis and evidence-driven design decisions, including the most efficient allocation of resources to survey strata. The concept of using RAD for conducting surveys is not new. Some clear antecedents include survey practices like replication (i.e., phases) in sample release, the embedding of experiments in survey data collections, and double sampling (two-phase sampling). These practices are inclusive of subsampling for nonresponse ([Hansen and Hurwitz 1946](#)), sequential analysis or adaptive trials ([Wald 1947](#)), and total survey error and total quality management ([Morganstein and Marker 1997](#)).

[Groves \(2011\)](#) described the three decades spanning 1960–1990 as the “Era of Expansion” in the application of survey methods. He cites Eleanor Singer, who labeled this same period the Golden Era of survey research. Many of us who lived and worked through the day-to-day challenges of that time might view it more as “gilded and shiny” than truly golden; it was an era of stability – scientifically designed surveys, using highly standardized and uniform methods, resulted in high-quality surveys with relatively high response rates and acceptable cost structures. Late in this period of relative stability, the challenge of rising costs for large scale in-person data collections was buffered by the advent of new telephone survey methodology – a buffer that remained effective through at least the late 1990s.

By the new millennium, the continual change in key survey conditions presented survey researchers with new challenges. Scientific and government surveys became more complex and often posed great uncertainty in design parameters and operational features. Survey populations’ resistance to survey participation continued to increase. Survey cost

structures were becoming even more dependent on decisions being made in the field or data collection centers, often with no evidentiary basis to measure or respond to cost fluctuations. Cost metrics, which are inherently multi-dimensional, remained as elusive as data quality. As these challenges grew, new opportunities also arose. Due to advances in computing and technology, there was improved access to sample frames, administrative data sources, global positioning (GPS), and geographic information systems (GIS).

Groves and Heeringa (2006) coined the term “Responsive Design,” following the oft-quoted advice of their mentor, Leslie Kish, who said, “If you want recognition for an idea, put a proper name on it” (e.g. “Design Effect”). Leslie’s advice in this regard was sage. The term “responsive design” stuck as did the companion term, “adaptive survey design.” As a catalyst for the initial growth of RAD, foremost among the new opportunities emerging in the first years of the new millennium were sophisticated new “real time” systems for sample management, data acquisition, and paradata capture in interviewing systems.

New developments in survey designs and methods can often take a decade or more to develop and mature with respect to research, publication, and applications in the field. This was certainly true for RAD. Initial RAD applications often emphasized simple nonresponse subsampling features to address the cost and effort of “end game” data collection at the conclusion of the survey period (Tourangeau et al. 2016). With time, the breadth and sophistication of RAD research and applications have expanded, resulting in the diverse body of knowledge and experience that is in evidence in publications such as the 2017 JOS special issue, this 2018 JOS special section, and a book recently published by Schouten et al. (2017).

RAD, in youth slang, means wonderful, fantastic, or extraordinary. We observe the rise of variants of RAD ideas, turning these ideas into survey practice in various contexts. RAD seems to be coming of age since the explicit implementation of responsive design or adaptive design during the mid-2000s (e.g., Groves and Heeringa 2006; Wagner 2008).

1.2. Four Pillars of Responsive and Adaptive Design

Though precedents and variants of RAD have been embedded in survey practice over decades (Groves and Heeringa 2006; Wagner 2008; Calinescu and Schouten 2016), we argue that RAD has four pillars for constructing survey design: 1) use of auxiliary information to stratify the heterogeneous population under study, 2) design features and interventions to adapt treatment, 3) explicit quality and cost metrics and functions to evaluate the efficacy of adaptation to strata, and 4) a quality-cost optimization strategy to find optimal allocations of treatments to strata. RAD is essentially a form of adjustment by *design* in the data collection as opposed to adjustment by estimation, that is adjustment introduced in the design and data collection stage in contrast to adjustment in the estimation stage. As a consequence, similar to estimation, auxiliary data should relate to nonresponse and other sources of survey errors under investigation, as well as to the key survey variables. Design features should be effective in reducing survey errors for the relevant strata. Quality and cost functions quantifying effort and errors should be properly defined and measurable, but, above all, should be accepted by the stakeholders involved.

The quality-cost optimization strategy should be transparent, reproducible, and easy to implement.

The first two pillars – auxiliary data and design features – emphasize data collection and a behavioural social sciences component, whereas the other two pillars – cost/quality metrics and optimization – are tied to estimation and statistics component. Between 2000 and 2015, there was renewed interest in paradata, or auxiliary data coming from the data collection process (e.g., [Kreuter 2013](#)). For example, call record data, audit trails, and interviewer observations were increasingly used in dashboards to monitor data collection. This might have resulted from increasing digitization of communication. The real-time paradata were instrumental to developing evidence-driven models to understand the process of response and nonresponse and to creating statistical interventions to control for potential nonresponse bias.

Survey design features obviously go as far back as surveys themselves; however, there has been renewed interest in mixed-mode surveys with the emergence of online devices (e.g., [Dillman et al. 2014](#); [Klausch 2014](#)). The survey mode appears to be the strongest quality-cost differential of all design features. Between 2005 and present, various articles have been published about indicators for nonresponse (e.g., Chapter 9 in [Schouten et al. 2017](#)). It has been declining response rates, we observe, that drove the development of alternative indicators; not necessarily to replace response rates but to supplement them and to provide a more comprehensive picture of data quality. Notable in data quality metrics is the development of response propensity measures (e.g., [Chun 2009](#); [Chun and Kwanisai 2010](#); [Toureangeau et al. 2016](#)). It is unfortunate that efforts to develop and implement cost metrics remain quite limited – probably due to practical constraints of quantifying or modelling cost parameters.

Optimization strategies remain an underexplored area. This may be, in part, because they are the final step of RAD. In other words, they require that choices in the other elements have been made and implemented. For instance, a consensus is necessary on quality and cost indicators. We observe, however, that it is also because optimization requires accurate estimates of survey design parameters, such as response propensities and survey costs. Survey cost metrics are multi-dimensional like data quality; optimization strategies, therefore, remain incomplete as long as cost estimates as input variables are neither reliable nor valid indicators of survey costs.

1.3. Overarching Research Questions

We present that the overarching research questions addressed by the 2017 JOS special issue and the 2018 JOS special section are as follows: 1) what approaches can be used to guide the development of cost and quality metrics in RAD and their use over the survey life cycle? 2) which methods of RAD are able to identify phase boundaries or stopping rules that optimize responsive designs? and 3) what would be best practices for applying RAD to produce high quality data in a cost-effective manner?

In response to these core questions of RAD, the JOS special issue and special section sought to address the following topics of adaptive design: theoretical contributions and applications, innovations, and comparisons of different methods of adaptive design that leverage the strengths of administrative records, big data, census data, and paradata as

well as survey response data. For instance, what cost-quality tradeoff paradigm can be operationalized to guide the development of cost and quality metrics and their use around the survey life cycle? Under what conditions can administrative records or big data be *adaptively* used to supplement survey data collection? How are paradata in multi-mode data collection conceptualized, pretested and collected to inform survey design decisions?

The articles included in the JOS special issue and special section address *interdisciplinary* dimensions of adaptive design, which encompass the following survey drivers: cost, response burden, data quality such as representativeness and response propensity, multiple sources of data, multiple modes of data collection, paradata, and new technologies. For instance, what indicators of data quality can be combined to monitor the course of the data collection process? Under what scenarios can the rules of switching from one mode to another be cost-effective? What stopping rules of data collection can be used across major phases of the survey life cycle?

We reiterate that the JOS articles involve experimental designs or simulations of adaptive design in household surveys, business surveys, and censuses. For instance, how could adaptive design be effectively designed and executed, especially in surveys involving multiple data sources and mixed modes of data collection? How could adaptive design guide web surveys while controlling for multiple sources of survey errors, such as nonresponse, measurement errors, and sampling errors?

2. Critical Pillars of RAD Addressed in the JOS Special Issue and Special Section

In this section we present three perspectives, discussing how the four pillars of RAD that are essential to survey design are addressed by articles published in the 2017 JOS special issue and in this 2018 JOS special section. Perspective A presents discussion points by leveraging the four pillars of RAD above. Perspective B articulates five key elements of RAD, or variants of the four pillars of RAD, to make a coherent discussion. Perspective C focuses on elaborating on cost measures and cost modeling, the missing half of cost-quality tradeoff analysis and optimization strategy, as tied to the third and fourth pillars of RAD.

2.1. Perspective A

Looking at the special issue and special section, overall, the articles lean towards the more statistical pillars: indicators and optimization. This is understandable as it is less reliant on costly experiments or pilots and more on feasible simulations. These articles are well written, useful, and creative in articulating contributions that introduce new perspectives and approaches to the existing literature. As mentioned, considering the optimization strategy, more specifically the translation of quality-cost to intervention and adaptation, there has been a gap in the literature. The advances in this direction are very welcome in the special issue and section. It is, to be sure, a pleasant surprise to see scholars from a variety of survey settings work on these methods.

Nonetheless, the greater statistical focus is also somewhat of a missed chance. This is for three reasons. First, being more statistical, the articles often employ simulations to demonstrate utility rather than real applications, which will bring problems of their own. In many cases the simulations do have a link to real surveys, but not always. Such case

studies may not convince survey designers and data collection departments, as they ignore practical and logistical constraints. Furthermore, the outcomes of case studies appeal less to experience about achievable quality and costs. Second, in the end, we need effective auxiliary variables – design feature combinations to find differentiation and leverage to adapt. RAD works only if we manage to find strategies that work better than what we have done before or proof that for some strata we need to spend more. And third, the more statistically oriented articles tend to be further removed from data collection personnel and, as such, are more difficult to implement. They need translation to daily practice. This holds true for the article by [Burger et al. \(2017\)](#) about the robustness of RAD to inaccurate design parameters. A few articles are exceptions, especially the article authored by the US Census Bureau researchers about the Annual Survey of Manufactures ([Thompson and Kaputa 2017](#)) and the article by NASS about the Crops APS survey ([McCarthy et al. 2017](#)); these articles are much closer to implementation. They do use realistic auxiliary variable – design feature combinations and seem to be driven by a practical need to improve uniform survey designs.

Thus, the special issue and special section represent strong articles that advance the statistical foundation, but are somewhat removed from survey practice. In Chapter 12 of [Schouten et al. \(2017\)](#), nine areas are enumerated and discussed that require progress and further research. The first area is that of empirical evidence that RAD works, or in other words, success stories. These are still relatively thin; see also the discussion of [Tourangeau et al \(2016\)](#). In the special issue, two of the original authors, [Brick and Tourangeau \(2017\)](#), explore and discuss how success stories may be achieved. It would be of great benefit if the authors of the other articles return with follow-up articles describing the spin-off and results of their future work. It is strongly recommended that they do.

2.2. Perspective B

At the risk of misclassification or over-simplification, the articles published in the 2017 JOS special issue and included in this 2018 supplement address five key elements of RAD, or variants of the four pillars as articulated above.

A first and foundational element of any RAD approach is the recognition by the research team that the survey population is *heterogeneous* with respect to the orientation to the survey topic, incentive for survey participation and preference concerning the timing and mode of data collection. Equally important to acknowledging this heterogeneity is the ability to predict where it occurs in the population so that RAD features can be tailored accordingly either prior to or during the survey data collection. In the 2017 special issue, the articles by [Kaminska and Lynn \(2017\)](#); [Durrant et al. \(2017\)](#); [Thompson and Kaputa \(2017\)](#) and [McCarthy et al. \(2017\)](#) address topics related to this element.

The feasibility and effectiveness with which features can be operationalized and actively managed during data collection represent a second key feature of any RAD survey. An “elegant” design may have tremendous appeal but is of little use if it does not work when put into operation. In the 2017 special issue, the articles by [Vandenplas et al. \(2017\)](#); [Early et al. \(2017\)](#); [Plewis and Shlomo \(2017\)](#), and [Burger et al. \(2017\)](#) address challenges in operationalizing and managing desired features of RAD designs. In this 2018 special section, the article by Walejko and Wagner and the article by Murphy and his

coauthors both address operationalization and survey management and monitoring issues encountered in tests of RADs for tailored designs: the former for the 2020 U.S. Census tests and the latter for the U.S. Energy Information Agency's Residential Energy Consumption Survey.

Even in today's world where RAD concepts are widely accepted and practiced, we as survey practitioners find it hard not to focus all possible efforts on maximizing the response to the survey. It is in our genes. Consequently, we struggle with the RAD concept that a rigorously applied survey protocol will ultimately reach a phase capacity at which additional effort and expenditures will not add any significant information content to the data that have been collected. Even more foreign to our traditional view is the idea that the design itself has reached the point at which further effort should be stopped. But if we can get over that hurdle, how do we as practitioners decide when a phase has reached capacity or the RAD study data collection should stop? Lewis (2017) and Paiva and Reiter (2017) both in the 2017 special issue describe quantitative tools that can guide phase transition or stoppage in RAD data collections.

As noted above, a principal aim of RAD is to achieve an optimal balance of cost and errors in survey populations where individuals' orientation to the survey request, incentive to participate, or data collection preferences vary. To fully achieve this aim, there must be reliable metrics for assessing both costs and errors. Nonresponse and associated selectivity in the composition of the observed population sample are a potentially important source of bias in the survey data. But nonresponse bias can be difficult to quantify, especially for surveys where the sample frame provides little information on the characteristics of respondents and nonrespondents. In the 2017 special issue, Brick and Tourangeau (2017) present models for survey nonresponse and investigate just how effective responsive designs might be at attenuating the bias associated with those nonresponse mechanisms. Särndal and Lundquist (2017) investigate whether actively controlling the "balance" of the observed sample during the RAD data collection should be preferred to standard methods in which post-survey calibration weighting adjustments are used. Closely related to the topic of weighting calibration using sample frame and administrative data is the option to use large scale administrative data sources as a substitute for direct survey or census enumeration. In this special section, Keller et al. describe a U.S. census investigation into the metrics for evaluating when external administrative data may be a suitable substitute for assigning vacancy status to addresses in the forthcoming 2020 Census enumeration.

2.3. Perspective C

Going forward, the litmus test of RAD success depends heavily on the extent to which the third and fourth pillars of RAD are properly assembled and tested against the pressure of total survey errors and total survey costs – both anticipated and unanticipated. The critical gap remaining in these two pillars of RAD is more due to under-development of the framework of cost metrics and lack of its implementation in real-world survey applications. Cost-quality optimization, by definition, would suffer inasmuch as cost metrics are not properly implemented. The underlying questions to ask include but are not limited to the following fundamentals: What are measurable survey costs? What would be desirable properties of

cost modeling? What methods are available and feasible to measure survey costs, inform cost-quality tradeoff analysis, and develop cost-quality optimization strategies?

Costs and errors are reflections of each other; increasing one tends to reduce the other (Groves 1989). Thus cost-quality optimization strategies would be neither feasible nor complete unless there is rigorous development and examination of the cost functions of various survey designs that offer error properties (Groves 1989; Chun 2012; Mulry and Spencer 2012). It bears reminding that a viable cost model is a function of fixed costs and variable costs as follows (Groves 1989):

Total Cost = Fixed Costs + Variable Stratum Costs

$$C = C_0 + \sum_1^H C_h n_h$$

Where C = total cost;

C_0 = fixed cost, incurred regardless of selected sample size;

C_h = variable stratum cost of the n_h sample cases in the h th stratum, namely cost of selecting, measuring, and processing each of the n_h sample cases in the h th stratum.

Fixed costs are costs that remain fairly constant in a survey, such as costs for survey system design, IT, and survey management. Variable costs are costs that vary as a function of the sample cases in various strata. Variable costs may include costs of frame construction, interviewing, nonresponse followup, data entry, and editing, which incur over the survey life cycle.

In practice, the pragmatic cost models need to be inclusive of *nonlinear, discontinuous and stochastic* properties of survey costs (Felligi and Sunter 1974; Groves 1989). They deserve discussion. Groves observes that existing cost models tend to be linear functions of survey parameters like the number of interviews, although nonlinear cost models often apply to practical survey administration. Most cost models are continuous in those parameters; however, he points out that discontinuities in costs often arise when administrative changes accompany certain design changes. While cost models tend to be deterministic, costs can vary extensively because of chance occurrences in probability sample selection, or choice of interviewers. Groves argues that because closed-form solutions to complex design problems do not exist, simulation approaches are useful to measure the sensitivity of results to changes in various design, cost, and error parameters. He offers several simulation examples of cost and error models that demonstrate that gathering better cost data must be given priority in order to develop reasonable cost models accounting for cost-error tradeoff.

The cost models proposed by Groves remain useful and viable today. Cases in point are the articles by Paiva and Reiter (2017) and Kaminska and Lynn (2017) in the 2017 JOS special issue and by Murphy and his colleagues in this special section. Using data from the 2007 U.S. Census of Manufactures, Paiva and Reiter show how to compute and compares measures of cost for various sample sizes by applying the traditional cost model. Kaminska and Lynn provide and test explicit cost metrics to determine pros and cons of alternative methods for allocating sample elements to data collection protocols, particularly in a longitudinal survey setting. Extending the cost model by Groves,

Kaminska and Lynn demonstrate how variants of adaptive and non-adaptive designs can be appraised in terms of relative costs as well as multiple measures of data quality for each proposed scenario of RAD. In a discussion of adaptive, responsive, and tailored (ART) design principles, Murphy and his colleagues make a smart move of presenting relative cost per case by interview protocol. They also provide data visualization of percentage of cases requiring editing, one that is tailored to the needs of cost metrics in an energy consumption survey sponsored by the U.S. Energy Information Administration. None of these articles, however, has taken a major step yet towards nonlinear, discontinuous, and stochastic properties of cost modeling.

3. What Open Research Questions and Challenges Exist for Implementing RAD?

Following each of the three perspectives as presented above, we turn to discussing what major questions and challenges remain to be addressed for advancing RAD.

3.1. Perspective A

In line with Perspective A as articulated above, the other eight areas of challenges addressed by Schouten et al. (2017) are as follows: 2) best practices for implementation, 3) clear and flexible quality-cost objectives, 4) versatile data collection systems, 5) skillsets for data collection staff, 6) relevant designed paradata, 7) application to longitudinal settings, 8) a total survey error approach, and 9) optimization strategies.

Areas 2 to 6 all relate to prerequisites for actual implementation, some of which are methodological and some of which are logistical and IT-related. They do not translate directly to academic research questions but do pose very interesting challenges for which there is an outlet in more practical journals and in conference proceedings. The real challenge here is to bridge the gap between theory and practice. RAD, perhaps more than ever before, has strong implications for how surveys are actually done. In its most strict form, RAD prescribes what sample units get what treatments to optimize what specified objectives under what decision rules. This has traditionally been the mandate of data collection departments and staff, and not of methodology. In order to get closer to implementation, however, data collection staff need to become co-researchers and co-authors. The case studies need to be driven by real-life issues with decreasing response and increasing costs. Areas 7 and 8 present new settings and applications that are mostly unexplored. Here, research questions could easily be formed and the attendant challenges are very exciting but also complicated.

How to implement RAD in longitudinal settings is a very interesting avenue to explore. The Kaminska and Lynn (2017) special issue article is one of the first to dive into this area. In panels, there are rich data about respondents, obviously, but also new challenges such as attrition, panel refreshment and conditioning. How informative is the previous-wave survey data about participation and measurement data quality? Is their explanatory power strong enough to overcome time lags and attrition? Are panel respondents consistent in terms of participation, costs and measurement, that is do they show the same behaviour in subsequent waves? How can RAD be combined with panel refreshment and could RAD be part of panel refreshment strategies? The measurement and answering behaviour is very

interesting in terms of RAD optimization. Since longitudinal studies are often about change, how can RAD be embedded over multiple waves?

The other broad research area is total survey error. RAD has mostly had a nonresponse perspective, probably because it has been driven by declining response rates and increasing costs. However, the most powerful design feature, the survey mode, has impacts on all survey errors. Nowadays, many survey designs are sequential mixed-mode and go from cheaper to more expensive modes – the rationale being that part of the sample is empathic to surveys and will respond under all strategies. RAD optimization, then, concerns decisions about the allocation of more expensive modes. The obvious questions are whether measurement is equivalent and whether a possible gain in participation is offset by a loss in comparability; and in RAD terms, whether these questions are answered differently for different sample subgroups. This is a discussion that goes beyond that of the mandate of data collection departments, as questionnaire content and survey estimates are typically produced by substantive departments. Similar total survey error impact may come from other design features such as interviewer allocation, split-questionnaire designs, central question follow-up procedures, and mobile devices. This will become even more prominent when survey data are combined with big data or mobile device sensor data into hybrid forms of data collection. RAD must, therefore, have a total survey error view. In such a setting, the number of quality indicators and constraints may increase or may require experimental components such as repeated measurements or randomization in question ordering.

Area 9, optimization strategies, is a key element of RAD. When posing RAD as a mathematical optimization problem, one finds that the number of possible designs quickly grows to a level beyond the reach of naïve/brute force optimization. The large number of options is not necessarily a problem as long as the optimization problem is (nearly) linear, but the most interesting problems are not linear and, even worse, not convex. These are, generally, complex problems to solve, such that clever and intuitive strategies are needed. Another approach may be to accept that suboptimal designs are good enough as long as they are better than uniform designs. Optimization strategies go hand in hand with strategies to learn and update. Mobile device data collection, for example, may be promising but it is a relatively unknown area. How can we optimize design when promising yet new design features emerge? Most survey designers and survey users are averse to constant change in design and with good reason. So how do we include optimization in terms of time series continuity?

3.2. *Perspective B*

RAD is a design tool that researchers can apply to potentially reduce both costs and errors of a one-off survey or a longer term program of surveys. RAD is not a panacea, capable of solving all problems of nonresponse, budget, or other survey errors. Considerable research and empirical work have demonstrated that not all RAD applications will succeed in optimizing the costs and errors of survey data collections.

There are several researchable questions that the study team should answer before a RAD is considered. First, given what is known from prior or similar surveys, is the survey population truly heterogeneous with respect to the orientation to the survey topic, their

incentive for survey participation, and timing or mode of data collection? Is it possible to design and implement adaptive or responsive features in the survey design that are matched to these different motivations, incentives, and preferences? Second, if such heterogeneity is present in the survey population and alternative design features can be identified that are responsive to these differences, can a RAD that incorporates these features be successfully operationalized during the field period? For example, attempts at a multi-phase responsive design for a telephone survey that spans a three-week data collection period will be limited by the time available to implement and evaluate the initial phase before transitioning to a second phase with alternative design features. The time constraint imposed by this same survey might be addressed using an adaptive design with pre-allocated features (e.g., contact materials, mode, and incentives). However, to be effective, such an adaptive design will require information from past experiences or experimental testing to guide the presurvey assignment of alternative design features to individual sample members. Finally, before deciding on a RAD for a future survey, the research team should carefully consider the added implementation and management costs for a RAD design that may entail multiple phases, multiple modes and other variations in design features. Will the fixed costs of the RAD implementation be offset by a reduction in the variable costs of collecting the survey responses? The criterion by which a RAD (or any statistical design) should be judged is that it should minimize mean squared error for key statistical aims subject to the budget and time period allocated for the project. It is relatively easy, given a fixed budget, to construct a RAD-like design that will result in increased nominal sample sizes and possibly even higher weighted overall response rates. However, that same design may be subject to large losses in effective sample size or differential selection biases for total sample and subpopulation estimates.

3.3. Perspective C

When it comes to the development and implementation of cost metrics that realize cost-quality tradeoff analysis and optimization, RAD seems to have a long way to go. Cost estimates, to us as survey practitioners, are day-to-day concerns to take into account in building a reasonable cost-quality tradeoff analysis and in developing optimization strategies that are rigorous enough to design optimal allocations of treatments to the population strata under study. Yet good and best practices of survey cost modelling are quite limited. Articles in the JOS special issue and special section move in the right direction of nailing down cost metrics and integrating cost metrics together with data quality metrics. Cost functions, however, remain to be rudimentary, not reaching the pragmatic cost models that need to be inclusive of nonlinear, discontinuous, and stochastic properties of survey cost estimates.

“Survey researchers have given much less attention to survey cost models than to survey error models,” wrote Groves (1989, 79) three decades ago. Unfortunately, we are facing the same issue today even as survey costs are increasingly becoming the major driver of survey design. The U.S. National Children’s Study stopped in large part due to problems with its design and management, as well as a huge survey cost that had already cost USD 1.2 billion by the time of its termination (Altman et al. 2014). With a life-cycle cost of about USD 13 billion, the 2010 U.S. Census was the most expensive U.S. census in

history; it was 56 percent more costly than the 2000 Census in constant 2010 dollars ([Government Accountability Office 2015](#)). Costs are the main driver of design changes for the 2020 U.S. Census, spurring innovations, such as the use of administrative records and third-party data, to materialize notable cost savings and sustain data quality ([U.S. Census Bureau 2017](#)). Unless the missing half of the cost-quality optimization is rigorously examined and fixed, RAD looks to be facing an uphill battle. What questions and challenges remain to be addressed by the community of RADers?

First, we should probably draw lessons from the Total Survey Error framework, based on which the error function has been well specified and developed over the last few decades. Can we conceive of the total survey cost model? Can we pair total survey cost model to Total Survey Error model? We are pointing to the model that may account for the traditional notion of fixed and variable costs and that is adaptive enough to be tailored to the pragmatic needs of cost modeling – nonlinear, discontinuous, and stochastic in terms of survey cost properties. We should work closely with accountants and field managers who may have the first-hand experience in cost estimating, computing, and cost modeling – top down or bottom up followed by converging cost estimates. Second, we as survey practitioners and survey designers need to maintain and archive cost data together with data quality information and make proprietary cost information available, whenever possible, in the collective interests of cost-quality optimization research. Probably the first step is to steadily maintain and share survey cost data based on a reasonable framework of total survey cost in the vein of total survey error. We, as a collective community of survey cost-quality modelers, should make concerted efforts to implement what is measurable and what is collectable when it comes to fixed and variable costs. These efforts should be followed by collective standardizing, if possible, as the American Association for Public Opinion Research has established standards of various response rates. Finally, cost-quality tradeoff analysts and optimization strategists should answer the question of examining the cost implications of designs that offer different survey error properties as echoed by [Groves \(1989\)](#). Several examples of cost analysis were provided by Groves when it comes to sampling error, nonresponse error and a few of the measurement errors associated with mode of data collection. Studies on total survey cost, cost per interview, and relative cost are on the rise as reported by [Wagner et al. \(2016\)](#). However, research on cost modeling remains thin relative to a coherent framework of cost-error properties.

4. Future Directions for Research and Application in RAD

The articles published in the 2017 JOS special issue and those included in this section are early indicators of the directions that future research and applications of RAD for surveys will be taking. The original concept of responsive design presented by [Groves and Heeringa \(2006\)](#) was in many ways overly structured and formulaic for the wide range of RAD applications that we now see described in the pages of JOS and elsewhere. The diversity of design and methodological developments that now fall under the RAD banner is so much larger than the original ideas. Some of these developments have been such major departures that they earned new labels: adaptive survey design (ASD) and, most recently, adaptive, responsive, and tailored (ART) design (e.g., see Murphy, Biemer, and Berry in this issue). Furthermore, a richer application of RAD has been found, for instance,

in a framework of dynamic question ordering (DQO) where question order is adaptively altered to improve response rate and imputation quality (Early et al. 2017).

Without question, the future of survey research will continue to bring new challenges and opportunities. The past decade's development of RAD approaches to population surveys has prepared us to adapt to expected changes in survey and other population data environments. Although the future demand for surveys and primary data collections will remain high, there will certainly be a growing emphasis on data collections to augment existing systems of administrative systems and other structured and unstructured sources of "big data". "Survey-assisted" population modeling, one that integrates large big data systems and streams with carefully designed survey observations, has long been used in the fields of agriculture, forestry and environmental sciences, and small area estimation. It is rapidly being extended to medicine and epidemiology as well as to economics, demography, and other social sciences. In this integrated role, survey data collection will assist in several ways:

"Model training" – providing timely estimates of models parameters relating the outcomes of interest to the covariate information available in the big data systems;

"Model refinement" – by supplying more complete information on multivariate associations, mediating and moderating effects and chronological or spatial variation in big data models;

"Compensation" – for population noncoverage, nonobservation or missing data in the large data systems;

"Insight" – into the error structure of large scale data systems that can only be obtained through direct survey measurement.

RAD will increasingly be called upon to support systems of model-based estimation, inference and prediction. Research to develop optimal, cost-efficient designs for such applications will need to be developed in the context of the statistical information that is present in existing sources of data and the specific statistical aims of the survey-assisted modeling system. Survey statisticians and methodologists leading this research will need to transfer their general knowledge of RAD principles and total survey error to these problems of survey-assisted modeling. Accountants and field managers proficient in cost estimation and cost modeling should be paired with survey designers and survey methodologists to develop, test and scale up pragmatic cost modeling aligned with a reasonable framework of total survey costs.

5. Conclusions

RAD is not an entirely new concept; tailoring and targeting have been part of survey design for quite some time. But it has never drawn such attention nor possessed such a formal structure. Given the strong cost-quality differential of self-administered versus interviewer-assisted surveys, the rise of all kinds of new devices and forms of communication, the option to form hybrid data collections using sample surveys and big data and the general individualization of societies, we believe it is inevitable that RAD

will be a natural component of many data collections. Perhaps this will be in relatively simple and rudimentary forms, but it is not logical to be inflexible and apply a uniform strategy. Furthermore, innovation in communication technology seems to be accelerating, with the likely consequence that it will be a moving target and it will always be different subpopulations that have adopted older and newer forms of communication. RAD principles may also expand to the big data arena by differentiating what sources are used for whom. We hope that the gap between data collection and methodology will become smaller so that data collection experts will not be misunderstood and they will polish the very promising ideas and anticipated yield of RAD to feasible designs.

The development of cost functions that may explicitly account for complex survey design features remains to be pursued by leveraging simulations and experimental studies that control for different cost properties. Cost-quality optimization strategies of RAD may be realistically developed only if cost functions are rigorously designed, tested, and implemented, while multiple metrics of data quality are being further matured.

6. References

- Altman, R., P. Pizzo, R. Gibbons, K. Hudson, R. Jenkins, B. Lee, M. Lichtveld, M.L. Miranda, C. Perry, H. Zoghbi, and L. Jorgenson. 2014. *National Children's Study*. Working Group NIH Advisory Committee to the Director Final Report. Bethesda, MD: National Institutes of Health. https://acd.od.nih.gov/documents/reports/NCS_WG_FINAL_REPORT.pdf.
- Brick, M. and R. Tourangeau. 2017. "Responsive Survey Designs for Reducing Nonresponse Bias." In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 735–752. Doi: <https://doi.org/10.1515/jos-2017-0034>.
- Burger, J., K. Perryck, and B. Schouten. 2017. "Robustness of Adaptive Survey Designs to Inaccuracy of Design Parameters." In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 687–708. Doi: <https://doi.org/10.1515/jos-2017-0032>.
- Calinescu, M. and B. Schouten. 2016. "Adaptive Survey Designs for Nonresponse and Measurement Error in Multi-Purpose Surveys." *Survey Research Methods* 10(1): 35–47. Doi: <http://dx.doi.org/10.18148/srm/2016.v10i1.6157>.
- Chun, A.Y. 2009. Nonparticipation of the 12th graders in the National Assessment of Educational Progress: Understanding Determinants of Nonresponse and Assessing the Impact on NAEP Estimates of Nonresponse Bias According to Propensity Models. University of Maryland, College Park, USA. <http://hdl.handle.net/1903/9916>.
- Chun, A.Y. 2012. "What Counts as Group Quarters? – A Glimpse of Census Cost-Data Quality Models." Paper presented at the Joint Statistical Meetings, San Diego, U.S.A. (accessed July 30, 2012).
- Chun, A.Y. and M. Kwanisai 2010. "A Response Propensity Modeling Navigator for Paradata." Proceedings of the Survey Research Methods Section of the American Statistical Association, Joint Statistical Meetings, Vancouver, Canada, 356–369. http://ww2.amstat.org/sections/SRMS/Proceedings/y2010/Files/306125_55196.pdf.

- Chun, A.Y., B. Schouten, and J. Wagner. 2017. JOS Special Issue on Responsive and Adaptive Design: “Looking Back to See Forward – Editorial.” In A.Y. Chun, B. Schouten, J. Wagner (Eds), JOS Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics*, 33(3): 571–577. Available at: <https://www.degruyter.com/downloadpdf/j/jos.2017.33.issue-3/jos-2017-0027/jos-2017-0027.pdf>.
- Dillman, D.A., J.D. Smyth, and L.M. Christian. 2014. *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method* (4th Ed.), Wiley.
- Durrant, G., O. Maslovskaya, and P. Smith. 2017. “Using Prior Wave Information and Paradata: “Can They Help to Predict Response Outcomes and Call Sequence Length in a Longitudinal Study?” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 801–833. Doi: <https://doi.org/10.1515/jos-2017-0037>.
- Early, K., J. Mankoff, and S. Fienberg. 2017. “Dynamic Question Ordering in Online Surveys.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 625–657. Doi: <https://doi.org/10.1515/jos-2017-0030>.
- Felligi, I.P. and A.B. Sunter. 1974. “Balance Between Different Sources of Survey Errors – Some Canadian Experiences.” *Sankhya Series C* Vol. 36(3): 119–142.
- Government Accountability Office 2015. 2020 Census: *Progress Report on Using Administrative Records to Control Enumeration Costs. Testimony before the Subcommittees on Government Operations and Information Technology, Committee on Oversight and Government Reform, House of Representatives, Washington DC.* <https://oversight.house.gov/wp-content/uploads/2015/11/Goldenkoff-GAO-Statement-11-3-Census-2020.pdf>.
- Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley and Sons.
- Groves, R.M. 2011. “Three Eras of Survey Research.” *The Public Opinion Quarterly* 75(5): 861–871. <https://doi.org/10.1093/poq/nfr057>.
- 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: Statistics in Society* 169(3): 439–457. Doi: <https://doi.org/10.1111/j.1467-985X.2006.00423.x>.
- Hansen, M.H. and W.N. Hurwitz 1946. “The Problem of Nonresponse in Surveys.” *Journal of the American Statistical Association* 41(236): 517–529. Doi: 10.1080/01621459.1946.10501894.
- Kaminska, O. and P. Lynn 2017. “The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 781–799. Doi: <https://doi.org/10.1515/jos-2017-0036>.
- Klausch, L.T. 2014. “Informed Design of Mixed-Mode Surveys: Evaluating Mode Effects on Measurement and Selection Error.” PhD Thesis, University Utrecht, The Netherlands. <https://dspace.library.uu.nl/handle/1874/300673>.
- Kreuter, F. 2013. *Improving Surveys with Paradata: Analytic Uses of Process Information*. New York, U.S.A.: John Wiley & Sons.
- Lewis, T. 2017. “Univariate Tests for Phase Capacity: Tools for Identifying When to Modify a Survey’s Data Collection Protocol.” In A.Y. Chun, B. Schouten, and J. Wagner

- (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 601–624. Doi: <https://doi.org/10.1515/jos-2017-0029>.
- McCarthy, J., J. Wagner, and H. Sanders 2017. “The Impact of Targeted Data Collection on Nonresponse Bias in an Establishment Survey: A Simulation Study of Adaptive Survey Design.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 857–871. Doi: <https://doi.org/10.1515/jos-2017-0039>.
- Morganstein, D. and D.A. Marker. 1997. “Continuous Quality Improvement in Statistical Agencies.” In *Survey Measurement and Process Quality* edited by L.E. Lyberg, P.P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin, 475–500. New York, U.S.A.: John Wiley & Sons.
- Mulry, M. and B. Spencer. 2012. “A Framework for Empirical Cost Modeling Relating Cost and Data Quality.” Paper presented at the 2012 International Total Survey Error Workshop (accessed September 3, 2012).
- Paiva, T. and J. Reiter. 2017. “Stop or Continue Data Collection: A Nonignorable Missing Data Approach for Continuous Variables.” 579–599. In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 579–599. Doi: <https://doi.org/10.1515/jos-2017-0028>.
- Plewis, I. and N. Shlomo 2017. “Using Response Propensity Models to Improve the Quality of Response Data in Longitudinal Studies.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 753–779. Doi: <https://doi.org/10.1515/jos-2017-0035>.
- Särndal, C. and P. Lundquist 2017. “Inconsistent Regression and Nonresponse Bias: Exploring Their Relationship as a Function of Response Imbalance.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 709–734. Doi: <https://doi.org/10.1515/jos-2017-0033>.
- Schouten, B., A. Peytchev, and J. Wagner 2017. *Adaptive Survey Design*. Chapman and Hall.
- Thompson, K.J. and S. Kaputa 2017. “Investigating Adaptive Nonresponse Follow-up Strategies for Small Businesses through Embedded Experiments.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 835–856. Doi: <https://doi.org/10.1515/jos-2017-0038>.
- Tourangeau, R., M. Brick, S. Lohr, and J. Li. 2016. “Adaptive and Responsive Survey Designs: a Review and Assessment.” *Journal of the Royal Statistical Society: Series A* 180: 203–223. Doi: <http://dx.doi.org/10.1111/rssa.12186>.
- U.S. Census Bureau. 2017. 2020 Census Operational Plan v3.0: *A New Design for the 21st Century*. Washington D.C. <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan3.pdf>.
- Vandenplas, C., G. Loosveldt, and K. Beullens. 2017. “Fieldwork Monitoring for the European Social Survey: an illustration with Belgium and the Czech Republic in Round 7.” In A.Y. Chun, B. Schouten, and J. Wagner (Eds), Special Issue on Responsive and Adaptive Design, *Journal of Official Statistics* 33(3): 659–686. Doi: <https://doi.org/10.1515/jos-2017-0031>.

- Wagner, J. 2008. "Adaptive Survey Design to Reduce Nonresponse Bias." PhD thesis, University of Michigan, Ann Arbor, USA.
- Wagner, J., K. Olson, and R. Anderson. 2016. "Survey Costs: The Missing Half of the "Cost-Error." Tradeoff, Paper presented at the Joint Statistical Meetings, Chicago, IL, USA.
- Wald, A. 1947. *Sequential Analysis*. Toronto, Canada: General Publishing Co.