The Impact of Targeted Data Collection on Nonresponse Bias in an Establishment Survey: A Simulation Study of Adaptive Survey Design

Jaki McCarthy[1], James Wagner[2], and Herschel Lisette Sanders[3]

Nonresponse rates have been growing over time leading to concerns about survey data quality. Adaptive designs seek to allocate scarce resources by targeting specific subsets of sampled units for additional effort or a different recruitment protocol. In order to be effective in reducing nonresponse, the identified subsets of the sample need two key features: 1) their probabilities of response can be impacted by changing design features, and 2) once they have responded, this can have an impact on estimates after adjustment. The National Agricultural Statistics Service (NASS) is investigating the use of adaptive design techniques in the Crops Acreage, Production, and Stocks Survey (Crops APS). The Crops APS is a survey of establishments which vary in size and, hence, in their potential impact on estimates. In order to identify subgroups for targeted designs, we conducted a simulation study that used Census of Agriculture (COA) data as proxies for similar survey items. Different patterns of nonresponse were simulated to identify subgroups that may reduce estimated nonresponse bias when their response propensities are changed.

Key words: Adaptive survey design; nonresponse bias; establishment surveys.

1. Introduction

Survey nonresponse is a ubiquitous phenomenon that has been an ongoing concern for survey practitioners (De Leeuw and De Heer 2002; Atrostic et al. 2001; Curtin et al. 2005; Brick and Williams 2013). Concern for nonresponse and the potential bias it may introduce includes establishment surveys (IGEN 1998; Petroni et al. 2004) and is not limited to household surveys. Any growth in nonresponse rates leads to concerns about the quality of the survey estimates. Further, the rising cost of surveys and limited budgets has led to the developments of new techniques for managing the risk of nonresponse error within fixed budgets. An initial technique for managing the uncertainty associated with changing patterns of nonresponse was known as "responsive design" (Groves and Heeringa 2006). These designs proceed in phases, where each phase is designed to be complementary by recruiting respondents who were less likely to respond in the previous phases. Later elaborations of responsive design included "adaptive designs" that employ data collection protocols that are targeted at specific subgroups of the sample and seek to

[1] USDA, National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A. Email: Jaki.McCarthy@nass.usda.gov
[2] University of Michigan, Institute for Social Research, 426 Thompson St. Room 4050, Ann Arbor, MI 48104, U.S.A. Email: jameswag@umich.edu
[3] RTI International, 701 13th St NW, Suite 750, Washington, DC 20005, USA. Email: hsanders@rti.org

minimize the risk of nonresponse bias (Wagner 2008; Couper and Wagner 2011; Calinescu et al. 2013). Adaptive designs, while not tailored responsively during field periods, serve the same goals by applying data collection protocols selectively to reduce nonresponse bias or contain costs.

These responsive and adaptive designs have all focused on producing response that is balanced with respect to indirect measures of nonresponse bias. These indirect measures (e.g., the R-Indicator, see Schouten et al. 2009) are based on covariates that are available on the sampling frame or from paradata (i.e., process data from survey data collections, see Couper 1998; Couper and Lyberg 2005). Very little is known about the impact of these responsive and adaptive designs on nonresponse bias. A recent simulation study, using data from household surveys, found that improvements in these indirect measures of nonresponse bias were correlated with reductions in nonresponse bias of adjusted estimates (Schouten et al. 2016).

The simulation study described in this article explores the potential impact of changing the response behavior of different identifiable subgroups in an establishment survey on final, nonresponse-adjusted estimates in a repeat cross-sectional survey. The study makes use of data from the Census of Agriculture, which includes data on a large proportion of agricultural establishments and can be used as a proxy for reported survey data. The subgroups that are targeted by adaptive designs should be chosen based on the joint criteria of 1) a posited approach that could improve response rates for the subgroup, and 2) changes in response for this subgroup are likely to lead to reductions in nonresponse bias of adjusted estimates. In our study, we will select cases for which there could be a plausible approach to increasing their response propensities. Then we will simulate the results of a targeted intervention with these cases to see if it accomplishes the second task of changing adjusted estimates.

2. Background

Nonresponse in establishment surveys is the result of processes that may differ from nonresponse in household surveys and thus it is unclear whether research results from household surveys can be applied to establishment surveys (Tomaskovic-Devey et al. 1994; Phipps et al. 1995; Willimack et al. 2002; Rivière 2002). Establishment surveys face a number of unique challenges. These include the need to gather information from multiple persons within some organizations, the need to have access to records in order to obtain the requested information, and finding persons with authority to release the requested information to the data collection organization. Establishment surveys also differ from household surveys in that size matters. In some industries, very large firms can be responsible for a large proportion of all production. In estimating total production, nonresponse from large firms can have more impact on bias and variance of final estimates. Therefore, it is possible to differentiate units on the basis of their importance for final estimates by a measure of size. Surveys of farming operations are unique in that larger operations may behave like establishments, while smaller operations are often establishments within households.

In establishment surveys, the motivation to respond is a complex function of internal and external pressures. Establishments are often in competitive environments and may

fear that releasing information will lead to competitive disadvantages. Thus sampled establishments with a predisposition to refuse may be much more difficult to convert to respondents than potential household survey refusals. On the other hand, emphasizing the mandatory nature of federally-sponsored establishment surveys has been shown experimentally to motivate response (Groves et al. 1997; Tulp et al. 1991; Dillman et al. 1996). Willimack and colleagues (2002) categorize influences on the motivation to respond into those under the control of the survey designers and those not.

Nonresponse to federally-sponsored establishment surveys is a continued concern and often only mitigated with increasing cost and effort in data collection (Petroni et al. 2004; IGEN 1998). This has prompted research into new methods for controlling the risk of nonresponse bias while controlling costs. Several research studies have explored the ability of adaptive survey designs to improve data quality with respect to nonresponse bias in household surveys. Peytchev and colleagues (2010) pre-identified cases from whom it was likely to be difficult to obtain response. These cases were worked earlier and with a more expensive protocol than other cases (e.g., incentives for interviewers that complete these difficult cases). They found that this extra effort did not lead to differences in final estimates. They concluded that "it may be more efficient to identify cases with the greatest potential to induce nonresponse bias in selected survey variables" (Peytchev et al. 2010, 27). Wagner and colleagues (2012) explore the impact of "responsive design" features on the composition of who responds to a large face-to-face household survey. They explore re-allocating effort across groups in order to improve the balance of who responds with respect to fully observed covariates. They show experimental evidence that targeted effort can lead to increased response rates among the targeted cases. Luiten and Schouten (2013) present results from an experiment that tailored several features of a mixed-mode (Web-Mail-CATI) household survey design. The assignment of the initial mode was tailored to characteristics of the case. In the CATI mode, both the assignment of the interviewer and the timing/number of the calls was tailored. This led to improved "representativeness," that is, less variability in response rates between subgroups without reducing the overall response rate or increasing costs. Calinescu and colleagues (2013) explore the use of optimization techniques to assign survey mode and effort (numbers of call attempts) differentially to cases with the objective of minimizing the variance of estimated response propensities. It is unclear whether similar results would be obtained for establishment surveys.

Targeting establishments with differential recruitment protocols has been suggested as a method for combatting establishment nonresponse. As Tomaskovic-Devey and colleagues put it, "[b]ecause establishments that are part of larger firms have special constraints, it might be wise to design different survey protocols for these establishments" (Tomaskovic-Devey et al. 1994, 454). While they were mainly considering the design of the questionnaire, this approach has been extended to the recruitment protocol. For example, Groves and colleagues (1997) found an interaction effect between the title of the person who receives the request to complete the survey and the size of the firm on response rates. For example, large firms were more likely to enroll if the "Unemployment Insurance Reporter" was requested, while smaller firms were more likely to enroll when the "Head of Payroll" was the requested respondent. In another example, Earp and McCarthy (2009) proposed a method for targeting subgroups of agricultural establishments for whom an incentive might be more effective.

Determining which cases should be targeted by an adaptive design is different for establishment surveys than for household surveys. One advantage that establishment survey practitioners may have over household survey practitioners is that there is often much more auxiliary information known about sample units in establishment surveys. Often administrative data may be available from other sources, or information from other data collections can be used in adaptive designs. This makes it more feasible to plan targeted data collection in advance, rather than responsively during data collection. For example, NASS collects data from farm and ranch operations in many surveys and the Census of Agriculture (COA). These data may be helpful in allocating recruitment effort for future surveys.

In past research, the National Agricultural Statistics Service (NASS) has used targeted data collection techniques to entice likely nonrespondent establishments to respond to surveys, mainly the Quarterly Agricultural Survey (also known as the Crops Acreage, Production, and Stocks Survey – APS) and the Agricultural Resource Management Survey Phase III (ARMS III). NASS has employed nonresponse propensity scores from decision trees to determine which operations receive targeted treatment efforts (Earp and McCarthy 2011). As Peytchev et al. (2010) discovered, employing targeted data collections for likely nonrespondent operations can be very expensive for very little return. This work focuses on identifying cases that can be targeted because we believe we can increase their probabilities of response assuming a simple reallocation of effort under the existing design, and identifying how changing their probabilities of response in this manner can change fully adjusted estimates. For example, resources can be diverted from cases unlikely to be influenced by alternative data collection procedures (i.e., predicted to refuse cooperation) to those where increased effort might increase response (i.e., those predicted to be hard to contact, but not likely to refuse).

The simulations described here represent alternative strategies that could be pursued by using the nonresponse scores NASS has available. We can first produce a gold standard estimate with no nonresponse. A simulated baseline estimate for comparison can be produced by removing data for Crops APS survey nonrespondents and applying nonresponse adjustments. Any simulated estimate with nonresponse can be compared to the gold standard to measure its bias, once nonresponse adjustments have been made. Our research questions are: Simulation 1: if resources are diverted from the most likely refusals and instead used to increase response among likely noncontacts, is the bias reduced? Simulation 2: if resources are diverted from the most likely to refuse and instead used to increase response from cases with a lower likelihood of refusing, is bias reduced? Simulation 3: if neither of these scenarios results in significant changes in bias, what would be the impact of severely reduced response rates on bias?

3. Data and Methods

3.1. Data

The Crops Acreage, Production, and Stocks Survey (APS) is a quarterly survey conducted by NASS in March, June, September, and December. The data from these surveys are used to calculate production estimates of numerous crops and grain stocks across the U.S.

(on a national and state level). The sample design is MPPS (multivariate probability proportional to size). The data collection period is very brief, lasting only two weeks. A paper copy of the survey is initially mailed to sampled operations. The materials include instructions for submitting response via the web. Nonrespondents are then contacted and interviewed by telephone or by field interviewers. Most of the data are collected over the phone (CATI).

In the Crops APS, classification trees have been developed to identify and flag likely nonrespondent operations in each state (McCarthy et al. 2010; Mitchell and McCarthy 2012). These classification tree models predict Crops APS response using characteristics and information reported by sample units in the Census of Agriculture. These models include factors that are unique to establishment surveys, such as features of the establishment (size, type of commodities, location, etc.), their history of response in NASS surveys (number of surveys they had been selected for, number of survey refusals, total burden, etc.) and other available information. These models were developed based on prior year's survey response/nonresponse and then validated on subsequent Crops APS samples not used in development of the models. Using a classification tree model, any sample unit can be assigned to a group based on a set of simple rules. The classification tree predicting survey refusals classified sample units into four mutually exclusive groups with increasing proportions of survey refusals (1 = high probability of refusal, 4 = low probability of refusal). A separate classification tree predicting survey inaccessibles classified sample units into five mutually exclusive groups with differing probabilities of being survey noncontacts (1 = high probability of being not contacted/inaccessible, 5 = low probability of being not contacted/inaccessible). While over 50 variables were used as inputs to the models, the only variables providing predictive power in the models are those related to the sample units' previous response history. The classification tree models are currently being used to assign a rank order score to Crops APS sample units. These rank order scores are currently provided to field operations for use in planning data collection. However, no specific actions for their use are required.

We focus on two states – Tennessee and Iowa – because these two states differ in the overall amount of agricultural production, the average size of farm and the distribution of commodities produced. In the 2012 Census of Agriculture, Iowa had 28,797,873 acres of cropland and an average farm size of 345 acres. Tennessee had 9,082,099 acres of cropland and an average farm size of 160 acres. This illustrates the common phenomenon of widely differing characteristics of units in establishment surveys.

Because all known agricultural operations are included in the Census of Agriculture (COA), and response to the COA is mandatory, most NASS survey samples have COA data available. Thus, we used COA data as a proxy for similar survey items in the Crops APS. In our case, we selected two key outcome variables: acres of soybean production and acres of alfalfa production. The COA has data for both survey respondents and nonrespondents. Therefore, the response indicator (e.g., Crops APS respondent and nonrespondent) and sampling frame information for the Crops APS is combined with the proxy survey outcome variables taken from the COA. Once combined COA data were successfully merged to the Crops APS, data were available for approximately 70% of the sampled cases, both respondents and nonrespondents. We can compare these match rates across responders and nonresponders, since we have these data for a proportion of both.

In Iowa, 71.7% of Crops APS respondents and 64.0% of Crops APS nonrespondents had COA data matched. The match rates for Tennessee were 72.9% for Crops APS respondents and 64.9% for Crops APS nonrespondents in 2010.

A similar approach was taken by Weber and Clay (2013). A strength of this approach is that we have complete data for a majority of the sample, both respondents and nonrespondents. The data from the COA about agricultural production are highly correlated with measures from the Crops APS (i.e., acres of corn harvested reported in the COA is highly correlated with acres of corn reported in subsequent Crops APS). A weakness is that we are using 2012 COA data as a proxy for 2010 Crops APS data. These data do not reflect the dynamic growth and death of agricultural production and operations. An additional weakness is that the COA data are not available for every case in the Crops APS sample. The cases for whom the data are available may differ from those for whom it is not. Thus, we looked into the county level characteristics such as the USDA's rural-urban continuum code which describes the degree of urbanicity of a county, and measures available from the Census Bureau such as the percent of the county in farmland, and the 2000 county population density. We found that there were no major significant differences in the variables of cases where there are COA data and for those without COA data. When there were significant differences, the direction is what we expected. For example, the average percentage of farmland in the county was higher for those with COA data in both Iowa and Tennessee and the average county population density was lower for those with COA data in Tennessee. These differences were as expected because the COA surveys producers who tend to have more farmland and live in less populated areas. Larger operations are more likely to survive. This differential matching rate might seem like a potential source of bias. However, the smaller farms, for whom the data are less often available, are also less influential on final estimates of production. Our analysis, by necessity, consists of only sample records with matching COA data.

3.2. Methods

By dropping those without COA data, we are treating the Crops APS sample with merged COA data as the sample for our simulation studies. This gives us a "gold standard" sample estimate without any nonresponse, and therefore no nonresponse bias (subject to the limitations described in the previous section). Using Crops APS response indicators, both respondent and nonrespondent COA reports of soybean and alfalfa production and a base sampling weight were used to get the population total estimates. These estimates were used as a benchmark against which estimates simulating different patterns of nonresponse were compared (It should be noted that these estimates do not simulate any NASS official estimates. There are several reasons why these estimates are not comparable to Crops APS estimates. First, data for the sample was taken from the census year, not the current survey year. In addition, because records without matching COA data are not included, all of our simulated estimates will underestimate the full population estimate.).

Nonresponse is then simulated using rank scores of noncontact propensity and/or refusal propensity derived from the classification trees and response indicators from the Crops APS. Response rates were derived within each cell of the cross-classification of the two ordinal variables (refusal and inaccessibility probability). The cross-classification of

the two ordinal variables, four from the refusal probability and five from the inaccessibility probability, should result in 20 cells of nonresponse rates. However, 19 cells were formed because one cell is empty (i.e., the cell that contains the second least likely to respond and the cell that contains the second least likely to be accessible). Nonresponse rates ranged from 7.5% to 55.6%. Observed response rates were then altered for each of our simulations. We selected different patterns of nonresponse either because they were alternatives we believe could be feasibly achieved (through case prioritization, switching modes from telephone to personal interview, or other design changes) or, in one case, because it was a worst case scenario for low response rates. The three patterns of nonresponse used in the simulation study are presented in Table 1. Pattern 1 diverts effort from cases that are likely to refuse and places more effort on cases that are difficult to contact. It is assumed that this re-allocation of effort would improve the response rate among the noncontact cases while reducing it among the refusal cases. Pattern 2 diverts effort from the cases most likely to refuse to cases that are somewhat less likely to refuse. This represents a scenario where the most likely to refuse cases are extremely unlikely to be persuaded to respond, but other cases with higher response propensities can be convinced to cooperate with additional resources. The assumptions about changes in response rate are listed in the table. Pattern 3 is an examination of what would happen if all cases had their probability of response reduced to half of current estimates. This is a worst-case scenario.

The nonresponse simulation was then implemented to evaluate the performance of the nonresponse adjustment on the crops estimates. Within each simulation, subgroups had their response propensity increased, decreased, or held constant as described in Table 1. We include a "baseline strategy" against which these altered protocols can be compared. This baseline is based on the estimates of response propensities under the current design. The targeted designs in Patterns 1–3 altered response probabilities based on posited design changes. For example, Pattern 1 posits that diverting effort from likely refusals to likely noncontacts will increase the response probabilities of the targeted noncontacts by ten percent and will decrease those of the likely refusal by ten percent. The cases were simulated to be either respondents or nonrespondents using a stochastic approach. This meant that a random number between 0 and 1 was drawn from a uniform distribution and

Table 1. *Four simulation designs.*

	Change to Response Probabilities (%)
Baseline Strategy	None
Pattern 1: Divert effort from likely refusals to likely noncontacts	
Refusal rank score highest	−10
Noncontact rank score in highest two categories	+10
All others	0
Pattern 2: Divert effort from most likely to refuse to less likely to refuse	
Refusal rank score highest	−10
Refusal rank score second highest	+10
All others	0
Pattern 3: Reduce response propensities by 50%	−50%

compared to the assigned response propensity for the given case. If the assigned response propensity was larger than the random draw, the case was flagged as a responder. Otherwise, the case was flagged as a nonresponder. In other words, cases whose corresponding response rate is larger than the random number drawn for that case were treated as respondents, otherwise they became nonrespondents. For the nonrespondents the COA data on soybean and alfalfa production were treated as missing.

After simulating nonresponse, the estimates based on the simulated set of responders were adjusted using the nonresponse methodology used in operational Crops APS. The Crops APS is stratified for nonresponse adjustment. The strata are defined by the type of commodity raised and the size of the operation. Table 2 shows the nonresponse strata, including the proportion of total alfalfa and soybean production from our sample and their observed response rates to the Crops APS survey. Imputation groups were formed within the same agricultural statistical district within the same stratum. Agricultural statistical districts are standard groupings of counties within a state. Imputation groups with fewer than five responding operations were merged together. The first merging of imputation group occurred for records within the same stratum regardless of agricultural statistics district. If this merging of cells did not produce a group with at least five records, then cells were collapsed into three "super-strata" namely the largest operation records that were sampled with 100% probability, the large records not sampled with 100% probability, and all other records. The missing reports of crop production are then filled-in through a ratio-adjusted mean imputation procedure carried out separately within each imputation group. First, the ratio of current acreage with the crop (alfalfa or soybeans) to total acreage is calculated for the cases with observed data. Second, this ratio was multiplied by the COA report of total acreage for the nonresponders. Once imputed, the sample total was obtained by summing the weighted crop estimate from each unit using the sampling weight.

Variance estimates are performed using a jackknife variance estimator (Kott 2001a; Kott 2001b). Since Crops APS samples without COA data were removed from the simulation, the jackknife replicate weights for the original Crops APS sample may be unbalanced. Thus, this study recreated 15 jackknife replicate weights. Starting with the fully imputed data set, the sample units from the self-representing (SR) strata were excluded. The remaining Non-Self-Representing (NSR) units were randomly divided into 15 groups of approximately equal size. For each JRR replication, one group of NSR units was deleted and the remaining NSR sample units were inflated by a factor of 15/14. The replicate weight associated with each SR unit equaled the original survey weight, since SR units are always in the sample so their overall contribution to the sample would not change. Using this method, 15 sets of replicate weights were created. There was no finite population correction included. Thus, the variance estimator for the total acreage was given by

$$var(\hat{y}) = \left(\frac{14}{15}\right)\sum_{i=1}^{15}(\hat{y}_{(i)} - \hat{y})^2$$

where $\hat{y}_{(i)}$ is the estimated total crop acreage computed using the ith set of replicate weights, and $\hat{y}$ is the full sample estimate of total crop acreage using the original survey weight. This variance should capture the effects of both original survey design and imputation.

Table 2. Production of soybeans and alfalfa and response rates by nonresponse adjustment stratum for establishments included in the simulation.

| | Stratum | Response rate | Percentage of production | |
			Soybeans	Alfalfa
Iowa	62: Cropland Acres 200–599	0.77	0.06	0.10
	64: Grain Storage Capacity 1–14,999 bushels	0.77	0.11	0.32
	68: Cropland Acres 600–1,599	0.74	0.04	0.06
	69: Grain Storage Capacity 15K–49,999 bushels	0.67	0.29	0.29
	73: Grain Storage Capacity 50K–299,999 bushels	0.68	0.29	0.16
	77: Cropland Acres 1,600–4,999	0.61	0.15	0.06
	78: Grain Storage Capacity 300K–999,999 bushels	0.71	0.04	0.00
	95: Cropland Acres 5000+	0.74	0.02	0.00
	97: Grain Storage Capacity 1,000K+ bushels	0.65	0.01	0.00
	Total		1.00	1.00
Tennessee	60: Cropland Acres 50–7,499	0.79	–	0.02
	61: All Hay Acres 20+	0.80	–	0.42
	62: Grain Storage Capacity 1–249,999 bushels	0.84	0.01	0.09
	69: All Tobacco Acres 1–5	0.83	–	0.04
	70: All Tobacco Acres 5–14.9	0.79	–	0.03
	71: Selected Cropland Acres 5–499	0.84	0.11	0.08
	72: All Tobacco Acres 15+	0.77	0.00	0.01
	74: Selected Cropland Acres 500–1,749	0.75	0.43	0.02
	77: Cotton Acres100+	0.65	0.17	0.00
	78: Alfalfa Hay Acres 25+	0.86	0.02	0.29
	79: Sorghum Acres 100+	0.71	0.06	0.01
	95: Cropland Acres 7,500+	0.65	0.09	0.00
	97: Grain Storage Capacity 250K+ bushels	0.79	0.12	0.00
	Total		1.00	1.00

The procedure of simulating nonresponse, imputing for nonresponse, recalculating the jackknife replicate weights, and using the jackknife replicate weights for variance estimation was implemented 500 times. In this way, the average of the 500 sample totals can be compared to the gold standard estimate for bias. The simulation should be able to provide an expected value of the bias and the variance.

Several things should be noted from Table 2. First, response rates differ across strata, although not dramatically. Second, some strata contribute much more to the population estimate than others, and this differs both by state and by commodity. This is typical for establishment surveys which produce numerous estimates.

4.   Results

We present the results from the two estimates separately. First, we will look at soybean production and then alfalfa production. Simulation results for estimates of the acreage of soybean production are presented in Table 3. The table presents the bias of the estimates, averaged over 500 simulations, and the percent of the simulations that include the gold standard population parameter (total acreage in soybean production) that is within the estimated 95% confidence interval, that is, the coverage rate. The estimated bias is presented for each state separately.

From this table, it appears that none of the strategies have a significant impact on the bias. The possible exception is for Pattern 3 – "Lower Overall Response Rate" simulation, which reduces the overall response rate by 50%. This substantial reduction in the response rate leads to a slightly larger bias of the estimate in Tennessee and this also leads to over-confidence in the estimate as the actual coverage is lower than the stated 95% nominal coverage. The targeted strategies – "increase contact" and "soft refusals" – do not reduce the bias.

Table 4 shows similar results for the acreage of alfalfa production. As a reminder, this is a smaller, specialty crop, reported by fewer establishments. In this case, the bias is actually increased under each of the designs. It may be the case that changes in response by smaller establishments can have a larger influence on estimates for a crop such as this. An examination of Table 2 shows that a large proportion of alfalfa is produced by smaller farms in both Tennessee (stratum 62) and Iowa (stratum 64 and 69). In fact, we found negative correlations between response propensities and alfalfa production within several of these strata. For example, in Tennessee there are statistically significant correlations between response propensity and alfalfa production in stratum 60 (-0.436, $p < 0.01$) and stratum 61 (-0.187, $p < 0.05$). In Iowa, there are similar negative correlations between response propensity and alfalfa production in stratum 64 (-0.080, $p < 0.10$), stratum 69, (-0.153, $p < 0.001$), and stratum 73 (-0.135, $p < 0.01$). These correlations provide a strong basis for bias. Soybeans, on the other hand, have a larger proportion of the production from larger establishments. There are fewer significant correlations between response propensity and soybean production. In Iowa, in stratum 95 (-0.252, $p < 0.05$)

Table 3.   *Simulated nonresponse bias under three different targeted data collection strategies for acreage in soybean production (major crop).*

Patterns of Nonresponse	Iowa		Tennessee	
	% Bias	Coverage rate (%)	% Bias	Coverage rate (%)
Baseline Estimates	0.16	100	0.32	100
Pattern 1 – Increase Refusals and Decrease Non-Contacts	0.14	100	0.35	99.8
Pattern 2 – Increase response for Soft Refusals	0.11	100	0.40	99.8
Pattern 3 – Lower Overall Response Rate Substantially	-0.18	99	1.32	91

Table 4. Simulated nonresponse bias under three different targeted data collection strategies for acreage in alfalfa production (small, specialty crop).

Patterns of nonresponse	Iowa		Tennessee	
	% Bias	Coverage rate (%)	% Bias	Coverage rate (%)
Baseline Estimates	− 4.26	90	6.75	97
Pattern 1 – Increase Refusals and Lower Non-Contacts	− 4.18	90	5.79	97
Pattern 2 – Increase Response from Soft Refusals	− 3.79	94	6.64	98
Pattern 3 – Lower Overall Response Rate Substantially	− 2.60	80	23.7	77

and in Tennessee, stratum 78 (-0.490, $p < 0.001$) have significant correlations. Neither stratum is very important in terms of the proportion of total soybean production in each state.

From the results in the table, it appears that the targeted strategies produce biases similar to the current practice, with the exception of the strategy that dramatically lowers response rates in Tennessee. Reducing response rates by 50% in Tennessee did lead to a large bias in estimates of alfalfa production. The targeted strategies, on the other hand, led to little or no reduction in bias.

5. Discussion

Our simulation results demonstrate the potential impact of changing response rates for specific subgroups of an establishment survey sample. By comparing two different adjusted crop acreage estimates and two different states, we evaluated the impact of these changes on estimates from a single survey with different characteristics. Iowa has considerably more agricultural operations and acres than Tennessee. Further, more producers are likely to raise soybeans than alfalfa in both states. For these reasons, interventions to increase response rates for subsets of the sample may have very different impacts on the different resulting population estimates.

Our simulations show that for large crops reported by more agricultural establishments (soybeans in Iowa), targeting subgroups has no substantial impact on nonresponse bias. Even drastic reductions (50%) in response do not introduce bias into adjusted estimates. For practical purposes, this is good news and demonstrates that our nonresponse adjustments are effective.

Perhaps surprisingly, for smaller commodities reported by fewer respondents (alfalfa in Tennessee), these hypothetical proposed subgroup interventions also did not demonstrate much change in the bias of the adjusted estimates. Again, this suggests that the nonresponse adjustments are appropriate. Although as expected, bias does greatly increase if response rates are decreased dramatically, as in our final simulations. One strategy suggested by the simulations is to break the correlation between response and size within

the adjustment strata for alfalfa production. This could be accomplished by focusing on relatively larger establishments within the adjustment strata, since many of the correlations were negative.

Our results are likely quite different from similar simulations had they been conducted in a household survey. This is due to the unique characteristics common to many establishment surveys. For example, the inclusion of self-representing units that are responsible for a large proportion of estimates of totals and for which there is no nonresponse may substantially reduce the potential impact of changes in data collection. Because of the importance of these units, the recruitment protocol may already be highly differentiated and tailored for them. Similarly, the skewed nature of populations with greater or lesser numbers (and sizes) of sample units reporting the estimated commodity means that the impact of strategies to change response may be quite different for different estimates within the same survey. The example in our simulations was the difference in results for alfalfa in our two different states. We did not eliminate the self-representing units in our simulations because of the key role they will always play in operational data collection. Again, this is a key difference in how adaptive designs may impact estimates in establishment and household surveys.

Given our results, survey organizations may want to consider the goals of targeting specific subgroups in similar establishment surveys. Rather than trying to reduce nonresponse bias, subgroups might be selected in order to reduce variance or costs. The potential impact of targeting subgroups for alternative data collection protocols could be evaluated with simulation studies such as ours and should include multiple key estimates.

There are several limitations to our study. Our results may be very specific to agricultural establishments, or even the two states or two crops that we examined. However, we chose these states and commodities to show the impact across key estimates of an establishment survey which may have only minimal correlation. Similar analyses could be conducted on other establishment surveys in order to build a broader understanding of the consequences of targeted data collection.

The present study has focused on bias. The changes that we proposed might also influence estimated variance. Although we tried to account for this by including coverage rates, we do not explore the implications for sampling error of the simulated design changes.

It might be the case that the subgroups we targeted are less than ideal for the purpose of reducing bias. The axis of our interventions was based on estimated response propensities and has been an approach used by NASS. This approach was used since it involved the relatively inexpensive task of reallocating existing recruitment effort. It may be that the response propensities under this design are not correlated with the outcomes of interest (i.e., total acreage in the production of soybeans and alfalfa). This would be a good situation since when response propensities are not related to the outcome variables, bias is unlikely to occur. However, we may be able to define more powerful interventions that work differently across subgroups that differ with respect to the survey outcome variables. In our case, this might be, for instance, interventions that work differently across operations with different sizes and past history of response. Further, since adjustment strategies also rely upon these characteristics, it might be better to look within groups defined by these characteristics for establishments that are likely to have changed since the

last COA. For instance, farming establishments with younger operators tend to change more quickly (Zech and Pederson 2003; El-Osta and Johnson 1998; Villatoro and Langemeier 2006). Future work will examine how to identify establishments that might have an impact on adjusted estimates rather than focusing on targeting with respect to predicted response propensities.

6. References

Atrostic, B.K., N. Bates, G. Burt, and A. Silberstein. 2001. "Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics* 17(2): 209–226.

Brick, J.M. and D. Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *The ANNALS of the American Academy of Political and Social Science* 645(1): 36–59.

Calinescu, M., S. Bhulai, and B. Schouten. 2013. "Optimal Resource Allocation in Survey Designs." *European Journal of Operational Research* 226(1): 115–121. Doi: http://dx.doi.org/10.1016/j.ejor.2012.10.046.

Couper, M.P. 1998. "Measuring Survey Quality in a CASIC Environment." In Proceedings of the Survey Research Methods Section of the American Statistical Association, August 1998. 41–49. Dallas, TX.

Couper, M. and L. Lyberg. 2005. "The Use of Paradata in Survey Research." In Proceedings of the International Statistical Institute Meetings. April 2005. Sydney, Australia.

Couper, M.P. and J. Wagner. 2011. "Using Paradata and Responsive Design to Manage Survey Nonresponse." ISI World Statistics Congress, Dublin, Ireland.

Curtin, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69(1): 87–98. Doi: https://doi.org/10.1093/poq/nfi002.

de Leeuw, E. and W. de Heer. 2002. "*Trends in Household Survey Nonresponse: A Longitudinal and International Comparison.*" In *Survey Nonresponse*, edited by R.M. Groves, 41–54. New York: John Wiley & Sons.

Dillman, D.A., E. Singer, J.R. Clark, and J.B. Treat. 1996. "Effects of Benefits Appeals, Mandatory Appeals, and Variations in Statements of Confidentiality on Completion Rates for Census Questionnaires." *Public Opinion Quarterly* 60(3): 376–389. Doi: https://doi.org/10.1086/297759.

Earp, M.S. and J.S. McCarthy. 2009. *Using Respondent Prediction Models to Improve Efficiency of Incentive Allocation*. United States Department of Agriculture, National Agricultural Statistics Service. Available at: http://purl.umn.edu/235087 (accessed April 2017).

Earp, M. and J. McCarthy. 2011. "Using Nonresponse Propensity Scores to Improve Data Collection Methods and Reduce Nonresponse Bias." In Proceedings, American Association of Public Opinion Research. May 2011. Joint statistical meeting, Phoenix, AZ. Available at: http://ww2.amstat.org/sections/srms/Proceedings/y2011/Files/400192.pdf (accessed April 2017).

El-Osta, H. and J.D. Johnson. 1998. *Determinants of Financial Performance of Commercial Dairy Farms*. Resource Economics Division, Economic Research Service, U.S. Department of Agriculture. Technical Bulletin No. 1859. http://purl.umn.edu/33561 (accessed April 2017).

Groves, R.M., D. Cantor, M. Couper, K. Levin, K. McGonagle, E. Singer, and J. Van Hoewyk. 1997. "Research Investigations in Gaining Participation from Sample Firms in the Current Employment Statistics Program." In Proceedings of the Section on Survey Research Methods of the American Statistical Association: 289–294. August 1997. Anaheim, CA. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/papers/1997_047.pdf (accessed April 2017).

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: http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x.

Interagency Group on Establishment Nonresponse (IGEN). 1998. *Establishment Nonresponse: Revisiting the Issues and Looking to the Future*. Council of Professional Associations on Federal Statistics, November 1998 Conference. Available at: https://fcsm.sites.usa.gov/files/2014/05/IGEN-COPAFS.pdf (accessed April 2017).

Kott, P. 2001a. "The Delete-a-group Jackknife." *Journal of Official Statistics* 17(4): 521–526.

Kott, P.S. 2001b. *Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys*. United States Department of Agriculture, National Agricultural Statistics Service. Available at: https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/allreports/Using_the_Delete-a-Group_Jackknife_Estimator_in_NASS_Surveys.pdf (accessed December 2016).

Luiten, A. and B. Schouten. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: an Experiment in the Survey of Consumer Satisfaction." *Journal of the Royal Statistical Society:* Series A (Statistics in Society) 176(1): 169–189. Doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01080.x.

McCarthy, J.S., T. Jacob, and A. McCracken. 2010. *Modeling Non-response in National Agricultural Statistics Service (NASS) Surveys Using Classification Trees*. Research and Development Division Research Report RDD-10-05, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC.

Mitchell, M. and J.S. McCarthy. 2012. "Voted "least likely to respond": Using Classification Trees to Identify Likely Non-respondents and Proactively Manage Data Collection in NASS's Quarterly Agricultural Survey." Presented at the Fourth Annual International Conference on Establishment Surveys, Montreal, Quebec, Canada.

Petroni, R., R. Sigman, D. Willimack, S. Cohen, and C. Tucker. 2004. *Response Rates and Nonresponse in Establishment Surveys BLS and Census Bureau*. Presented to the Federal Economic Statistics Advisory Committee: 1–50.

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(1): 21–29. Doi: http://dx.doi.org/10.18148/srm/2010.v4i1.3037.

Phipps, P.A., S.J. Butani, and Y.I. Chun. 1995. "Research on Establishment-Survey Questionnaire Design." *Journal of Business & Economic Statistics* 13(3): 337–346. Doi: http://dx.doi.org/10.2307/1392193.

Rivière, P. 2002. "What Makes Business Statistics Special?" *International Statistical Review* 70(1): 145–159.

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

Schouten, B., F. Cobben, P. Lundquist, and J. Wagner. 2016. "Does More Balanced Survey Response Imply Less Non-Response Bias?" *Journal of the Royal Statistical Society,* Series A (Statistics in Society) 179(3): 727–748. Doi: http://dx.doi.org/10.1111/rssa.12152.

Tomaskovic-Devey, D., J. Leiter, and S. Thompson. 1994. "Organizational Survey Nonresponse." *Administrative Science Quarterly*: 439–457. Doi: http://dx.doi.org/10.2307/2393298.

Tulp, D.R., C.E. Hoy, G.L. Kusch, S.J. Cole. 1991. "Nonresponse under mandatory vs. voluntary reporting in the 1989 survey of pollution abatement costs and expenditures (PACE)." In Proceedings of the Survey Research Methods Section of the American Statistical Association, August 1991, 272–277. Alexandria, VA.

Villatoro, M. and M. Langemeier. 2006. "Factors Impacting Farm Growth." *Journal of the American Society of Farm Managers and Rural Appraisers* 69: 74–80.

Wagner, J.R. 2008. "Adaptive Survey Design to Reduce Nonresponse Bias." Program in Survey Methodology. Ann Arbor, University of Michigan. PhD: 158.

Wagner, J., B.T. West, N. Kirgis, J.M. Lepkowski, W.G. Axinn, and S.K. Ndiaye. 2012. "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." *Journal of Official Statistics* 28(4): 477–499.

Weber, J.G. and D.M. Clay. 2013. "Who Does not Respond to the Agricultural Resource Management Survey and Does it Matter?" *American Journal of Agricultural Economics* 95(3): 755–771. Doi: https://doi.org/10.1093/ajae/aas171.

Willimack, D.K., E. Nichols, and S. Sudman. "Understanding Unit and Item Nonresponse in Business Surveys." In *Survey nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J. A. Little, 213–227. New York: Wiley, 2002.

Zech, L. and G. Pederson. 2003. "Predictors of Farm Performance and Repayment Ability as Factors for Use in Risk-Rating Models." *Agricultural Finance Review* 63(1): 41–54. Doi: http://dx.doi.org/10.1108/00214990380001140.