Is the Price Right? Predicting Future Survey Costs Using Previous Cycles

Kayla Varela[1], Allison Zotti[1]
[1]U.S. Census Bureau, 4600 Silver Hill Road, Suitland, MD 20746

Abstract

A recent survey climate of decreasing response rates and increasing budgetary constraints has burdened surveys in a wide range of fields. Survey sponsors are looking for ways to not only increase response, but also to most effectively allocate their limited resources. Ideally, a survey sponsor could leverage historic data from previous data collection cycles to estimate survey costs and optimize the next cycle's budget. The National Household Education Survey (NHES) is a national, cross-sectional survey of households sponsored by the National Center for Education Statistics (NCES). Using a variety of sources, we constructed cost estimates for both the 2016 and 2019 NHES data collection cycles. Using only our 2016 cost estimates, our goal is to analyze prior cost trends, and to model future cost trends for the 2019 cycle. We can then compare our predictions to the actual incurred costs in the 2019 NHES. Information about the trade off between cost and response rate are critical to the design of all surveys, from basic cost estimation to the development and implementation of alternative data collection features.

Key Words: Survey Cost, Cost Modeling

1. Introduction[1]

Cost plays a major role alongside other metrics of data quality in decision making and planning for a survey, because at the end of the day, a survey is bound by a finite budget. Discovering the cost implications of certain data collection pathways for different types of cases could help survey sponsors get the most 'bang for their buck' by optimizing how they utilize their data collection budget. In order to get a clear understanding of the relationship between the cost and the effectiveness (measured by response rate and representativeness) of a given data collection strategy, we have built case-level cost estimates for the 2016 and 2019 administrations of the National Household Education Survey (NHES). Ultimately, our goal is to use these historic cost estimates to predict future survey costs based on the different case-level treatments in the NHES. To achieve this goal we will perform exploratory analyses and modeling techniques using both 2016 and 2019 survey data. We will evaluate our models by using our 2016 data to retroactively 'predict' our 2019 data.

2. Background & Motivation

2.1 Survey Cost Background

This research project was originally motivated by a lack of current literature, but there is some notable work being done that touch on cost estimation in the survey setting that has helped guide this research. This session is also a great example of the growing library of current research on the topic of cost estimation in this field. Many of the same cost topics were covered at a 2006 workshop on survey costs organized by the National Institute of Statistical Sciences (NISS). Participants at this workshop laid out the fundamentals of the cost troubles facing federal statistical agencies, as well as the obstacles to overcoming these challenges (Karr, Last 2006). One notable literature on cost-work in the survey methodology field is James Wagner's cost work for the National Survey of Family Growth (NSFG) (Wagner 2019). Because the NSFG relies largely on field operations for

[1] Any views expressed on statistical, methodological, technical, or operational issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.
[2] All results have been approved by the Disclosure Review Board, Approval #CBDRB-FY21-283

data collection, a majority of Wagner's work focuses around the estimation of field costs. This paper, on the other hand, focuses on the estimation of mail costs and telephone costs. Additionally, Bob Groves has put together significant research surrounding the relationship between survey cost and survey error (Groves 2004). While Groves does not focus specifically on survey cost estimation using priors, he highlights many reasons why cost estimation is critical to survey success. Recently, Stephanie Coffey has completed significant cost estimation work for her dissertation focusing on estimating and predicting survey costs for the national survey of college graduates (NSCG) (Coffey 2020). She considers the use of both historic case data, and live paradata collected during data collection in her models, which we hope to implement in an extension of this work.

Because there has not been much work done at this scale in the field of survey methodology, it might also be reasonable to explore literature on cost modeling and estimation from adjacent fields. Healthcare and pharmaceutical industries have significant literature regarding cost model estimation, analysis, and evaluation. Some of these studies do implement predictive modeling for both cost and outcome, which could be useful. Ultimately there is very limited cost research done in the survey methodology field that relies on the sorts of logistic regression models that we plan to implement. While cost allocation strategies might be gleaned from existing research in this field, we might need to rely on a combination of research from other industries and our knowledge and applications of regression modeling to complete the modeling phase of this work.

2.2 National Household Education Background

The National Household Education Survey (NHES) is sponsored by the National Center for Education Statistics (NCES) and is administered by the US Census Bureau (Census). The American Institutes for Research (AIR) provided statistical planning, support, and data post-processing for NHES:2016 and NHES:2019. NHES:2019 included two Topical surveys: Parent and Family Involvement in Education (PFI) and Early Childhood Program Participation (ECPP). In addition to the NHES:2019 Topical surveys, NHES:2016 included the Adult Training and Education Survey (ATES) Topical survey as well. Data is collected nationally and provides descriptive information on the educational experiences of Americans ranging from early childhood through students enrolled in grade 12 during NHES:2019 or through age 65 during NHES:2016. Historically, NHES has been collected approximately every other year; however, following the collection in 2016, NHES will be collected roughly every three years going forward. NHES:2016 was predominately collected using paper instruments, while NHES:2019 was predominately collected using a web-first sequential mixed-mode design including two requests mailed to sample members to complete the survey by web followed by two mailings of paper instruments. In both 2016 and 2019, a Telephone Questionnaire Assistance (TQA) line was available for inbound calls. In 2016, only the Screener instruments were completed by telephone interviewers and paper Topical surveys were sent. In 2019, telephone interviewers were able to complete the entire survey with inbound callers.

Each NHES data collection since 2012 has included a series of randomized experiments with the goal of testing innovative methods to increase overall response, while maintaining data quality and not exceeding budget constraints. Historically, it has been difficult to truly measure how cost and budget are affected across the different experimental treatments. For NHES:2019, one of the goals was to attempt to capture the cost-quality trade off for each of the experimental treatments in an effort to provide additional information for making future design decisions. We first conducted a

retrospective cost analysis of the NHES:2016 experiments and then applied the same methodology to the NHES:2019 experiments.

For each experiment, there are a variety of potential costs associated with each treatment condition that would affect a cost-per-case or cost-per-complete analysis. NHES:2016 included two experiments applied to nearly 40% of the total sample size of 206,000 households. Some conditions included variable Screener incentive amounts (ranging from $0 to $10) and an experiment to test data collection by web with a sequential multi-mode contact procedure. NHES:2019 included a series of experiments applied to roughly 80% of the total sample size of 205,000 households. Some conditions included an additional incentive (offered conditionally upon completing the survey), targeted mail materials to reach Spanish-speaking respondents, or a variety of advance mailing materials.

While the first part of this paper focuses on the sources and methodologies used to build, the ultimate goal of this research is to start exploring the feasibility of using prior rounds of data collection to predict future data collection costs. For this, it is necessary to consider some of the differences between data collection in 2016 and 2019. Below, Table 1 displays the case counts for the most comparable data collection groups between the years. These are the only cases that will be included in our initial modeling work.

Table 1. Case Counts by Year

	2016	*2019*
Paper Only	126,000	4,000
Multimode	35,000	40,000

3. Cost Estimates

3.1 Data Sources

For both years, cost information was gathered for each piece of the mail-out process. This information was gathered from a number of different sources. Each source is listed below with a description for what information each source provided.

Financial Management Reports (FMRs)

FMRs are monthly reports that are internal to Census. These reports detail the amount of money that gets charged to the NHES project code on a monthly basis during the course of data collection. The reports are broken-out by division, center, and branch, which allows costs to be determined for specific operational areas. In some cases, the FMRs further provide detail by assigning costs within area to specific tasks. This extra detail sometimes allows costs to be determined by an area's specific functions, such as the check-in of paper questionnaires or development of the web instrument. The FMRs include labor charges and select direct costs.

CenDocs

CenDocs is an internal online repository of mail-piece orders that are contracted to an external printer. In many cases, National Processing Center (NPC) prints the letters and mail-pieces internally; however, paper questionnaires or other larger mail-pieces are contracted to an external printer to save time and money. CenDocs includes information about the type and length of a mail-piece, how many were ordered, and the final cost of the order. For any externally printed mail-piece, the cost of that mail-piece can be

determined by dividing the total cost of the order by the number of mail-pieces ordered. For NHES:2016 and NHES:2019, all paper Screener questionnaires, questionnaire envelopes (both Screener and Topical), and all return envelopes were printed commercially by an external printer. All other mail materials were printed on-site at NPC.

NHES:2016 and NHES:2019 master files

The master file was used to identify which treatment condition was assigned to each case, as well as available demographic characteristics for each case. The source of demographic information was MSG, the vendor for the sample frame. The master file also identified which mail-outs each case received, when the case responded, the mode of response, intermittent and final status codes, and whether or not the case called the Telephone Questionnaire Assistance (TQA) line.

Data collection overview documentation

The data collection overview documentation, prepared by NHES survey managers at Census, was used to identify information about each of the 2016 or 2019 experiments. The document covered each mail-out, including what mail piece type (letter, postcard, pressure-sealer, etc.), envelope, questionnaire, and incentive was sent in each package based on which treatment condition was assigned to the case. Additionally, this documentation provides the schedule for data collection operations for the standard data collection pathway and each of the experiments, so costs from FMRs and other sources can be assigned to the correct operations.

Subject Matter Experts (SMEs)

In addition to available cost-related documentation, important pieces of information for cost analysis came from the subject matter experts for specific data collection operations. Personnel at NPC provided the average cost-per-sheet of printed material for each type of mail-piece printed at NPC, the average postage amounts for each package-type mailed from NPC, including estimates for average FedEx postage amounts, and information about how to break down the costs for any mail service that NPC provides to the survey. Personnel within the call-center area provided information about how to allocate costs for anything related to the TQA service that NHES utilizes.

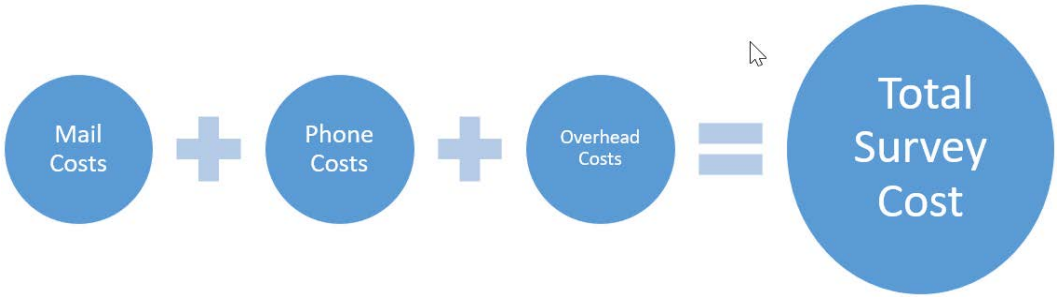
TQA output files

The TQA output files provide a call log for every call-in made by the survey participants during the course of data collection. While the master file identifies which respondents responded by calling TQA, the TQA output files captures every call, including calls where the caller only sought information about the survey or to ask questions. In addition to apportioning the cost of TQA to the cases that responded via TQA, the TQA output files were used to allocate a portion of the cost of TQA to the cases that simply called with questions, but did not respond via TQA.

3.2 Cost Allocations

For the NHES, cost allocation falls into three main categories: mail costs, Telephone Questionnaire Assistance (TQA) costs, and web development costs. Overhead costs are not included in the cost-per-case estimates developed for these analyses. Other surveys might consider additional categories of cost, such as field operation costs, but Figure 1 displays all considered categories of cost for the NHES.

Figure 1



Mail Costs

Mail costs are any costs that a survey accrues as part of any mail operations during data collection. These mail costs can be broken down into three main categories: outgoing material and postage costs, outgoing-mail labor costs, and return-mail costs.

Material and Postage Costs

Material costs cover any physical supplies that are mailed out during data collection, as well as the forms design for each of these materials. This includes letters, questionnaires, data products, specialty envelopes, and postcards. The data collection documentation for each cycle of NHES lists every piece of material included in each separate mailing, which makes it possible to add each material cost for each mailing into one cost-per-mailing. The costs of these materials can be determined using CenDocs and/or SMEs. A breakdown for which mail materials and postage costs are included in the cost analysis is shown in Table 2, as well as which costs are not included in the cost analysis.

Table 2. Breakdown of Which Mail Material and Postage Costs are Included in the Cost Analysis

Cost Included	Cost Not Included
Material and Postage	
Postage	Cognitive testing of letters
Commercially printed letters	Cognitive testing of questionnaires
Paper and ink for NPC-printed letters, pressure-sealers, and postcards	OMB package development and approval
Commercially printed Screener questionnaires	
Paper and ink for NPC-printed Topical Questionnaires	
Commercially printed envelopes	
Ink for NPC-printed stock envelopes	
FedEx envelopes	
Address labels	
Letter, envelope, and postcard design	
Questionnaire design	
Incentives	
Face value of cash incentive[1]	Acquisition
	Storage and Security

[1] The cash value of the incentive was not included in the cost of undeliverable as addressed (UAA) packages that included an incentive, as the incentive value itself is recovered by NPC; however, the cost of the labor for recovering the incentive from the UAA package is included, as described in the Return-mail Costs section.

All mail-outs are sent from NPC during the data collection cycle. The NPC cost estimate provides an average estimate of the postage cost for each mail-out based on the size and weight of the mail-out. The cost of return postage is accounted for in the return-mail costs section, because that postage is only charged if a piece of mail is returned using the return postage.

Incentives

Incentives are included at both the Screener level and Topical level, as specified in the data collection documentation. The obvious cost of the incentives are the cash value of the incentives themselves; however, there are additional costs related to acquisition, storage, and security that currently are not captured in the FMRs or cost estimate documentation from NPC and therefore, cannot be included. If a mail-out containing an incentive is a UAA, the incentive is recovered by NPC. The incentive recovery process requires time and labor, meaning the full cash-value of the incentive is not recovered. More about how this cost is determined is discussed below in the return-mail costs section.

In 2019, a $10 or $20 contingent incentive was rewarded to any respondent within the two choice-plus treatment groups if the respondent completed all required questionnaires via the web instrument or by phone after calling into the Telephone Questionnaire Assistance (TQA) toll-free phone number. If the respondent was not sampled for a Topical, they were only required to complete the Screener questionnaire via the web instrument or telephone to qualify for the contingent incentive. If the respondent was sampled for a Topical, they were required to complete both the Screener and the Topical questionnaires via the web instrument or telephone to qualify for the contingent incentive. For the sake of uniformity across all cases, the amount of the contingent incentive was included as a Screener cost for this cost analysis.

Outgoing-mail Labor Costs

Outgoing-mail labor costs account for all of the activities associated with assembling and mailing out the necessary mail pieces to each sample case. Costs associated with these activities came from the FMRs within the Document Services Branch, the Support Services Branch, and the Project Coordination Office. Table 3 shows a description for each activity and how the cost was applied to each case.

Table 3. Breakdown for How Costs are Allocated for each Outgoing-mail

Branch	Labor Activity	How Does the FMR Charge Get Applied?
Project Coordination Office	Management/Administrative Procedure Writing and Processing	Divide FMR total by the total number of sampled Screener cases to get a one-time cost-per-case
Support Services Branch	Transportation and Warehousing	
Document Services Branch	Document Programming	
	General Management	Divide FMR total by the total number of mailed packages to get a cost-per-package
	Distribution Finishing[1] Printing	Using NPC's cost estimate for each labor task, sum the applicable labor costs associated with each package type into a labor-cost-per-package
	Quality Assurance (QA)	Distribute FMR total proportionally based on NPC's cost estimate for each QA task, then divide the total QA-cost-per-task by the number of applicable packages to get a QA-task-cost-per-package

[1] Incentive handling and insertion are labor tasks that are included within the Finishing labor activity.

The NPC cost estimate includes sub-tasks for each activity listed in the table that can be applied to the different types of mail pieces (questionnaire, letter, postcard, etc.), as well as differential costs for tasks associated with the Screener operation versus the Topical operation. The distribution activity includes sub-tasks such as preparation of the questionnaires, letters, and postcards. The finishing activity includes sub-tasks such as folding, inserting, manual assembly of the incentive, cutting postcards, applying labels, sealing, or making the Topical booklets. The printing activity includes sub-tasks such as overprinting the envelopes and addressing the Screener questionnaires.

Return-mail Costs

In addition to the outgoing mail-out materials, postage, and labor, any mail piece that is returned to Census would also accrue return-mail costs. These incoming costs include the return postage and the cost of processing the returned questionnaire. All returned questionnaires are processed through NPC, within the Data Capture Branch (DCB). The FMRs for each NPC area detail how much money the DCB charged to NHES by month. Ideally, the monthly totals could be summed together into an overall cost-per-operation for the months in which the operation occurs; however, because the Screener and Topical operations overlap by at least two months, the calculation for cost-per-operation is not available directly from the FMRs. Instead, processing costs were allocated for both returned Screener and Topical forms based on the length of the form returned, in order to fairly assign costs to different cases across operations. While this method helps to allocate differential costs based on the length of the returned questionnaire so conclusions can be drawn for the Screener and Topical phases separately, the overall cost comparisons are more reliable because the overall measures include both phases without the need to differentiate between the two.

The return-mail costs can be separated into three different categories: programming costs, check-in costs, and processing costs. Programming costs, which include linking the case-level identifiers and USPS barcodes to the mail pieces for data capture and mail piece

tracking, occur prior to each mail-out and are not dependent on the number or type of returned mail pieces. Check-in costs, which include opening, sorting, and designating a check-in status for each returned mail piece, account for any activity that occurs for any returned mail piece, regardless of status (respondent versus UAA versus refusal). Processing costs account for any activity that would only occur for respondent mail pieces, which includes scanning and keying. Table 4 shows how each of these DCB charges were separated and which types of cases would accrue each of the charges. Returned packages include responses, UAAs, and refusals; returned-response packages include only responses.

Table 4. Breakdown for How Costs are Allocated to each Return-mail Activity

Activity	Category of Activity	For Which Type of Case Does the Activity Occur	How Does the FMR Charge Get Applied?
Programming (for document tracking)	Programming	All cases[1]	Divide FMR total by the total number of sampled Screener cases to generate a cost-per-case
Postage	Check-in	Responses, UAAs, Refusals	Known, documented postage cost-per-package
Management Mail Sort General Clerical Check-in Open/sort	Check-in	Responses, UAAs, Refusals	Divide FMR total by the total number of returned packages to get a cost-per-returned-package
Batching Guillotining	Processing Preparation	Responses[2]	Divide FMR total by the total number of returned-response packages to get a cost-per-returned-response-package
Scanning Imaging Key from Image Quality Assurance	Processing	Responses	Divide FMR total by the total number of returned questionnaires pages, then for each questionnaire multiply the cost-per-page by number of pages in the returned questionnaire

[1] The cost of data capture programming is a sunk cost that occurs before the start of data collection and therefore is applied equally to all sample cases, regardless of whether or not the case ever returns a mail piece.

[2] For NHES:2019, Screener returned-response questionnaires that answered that they had no children (and therefore were ineligible for a Topical) did not accrue any of the processing costs – these questionnaires accrued only the check-in costs, as they are handled in the same way as a returned-refusal questionnaire.

Due to the differing lengths of the questionnaires, the differential cost between processing the Screener versus one of the Topical questionnaires also needed to be accounted for. Opening, sorting, batching, and guillotining of the returned mail packages are processing operations that occur at the package-level. The costs of these activities then, are at the package-level, regardless of the length of the returned questionnaire. However, scanning, imaging, key from image, and quality assurance costs can be different based on the length of the returned questionnaire. To account for this, the total number of pages returned was determined by multiplying the number of pages in the

questionnaire by the number of questionnaires returned for each questionnaire type, then summing across all questionnaire types. Then, the total cost-per-activity was divided by the total number-of-pages-returned for all questionnaires to get a cost-per-returned-page. From there, the cost-per-return was simply the number of pages of the returned questionnaire multiplied by the cost-per-returned-page. This allowed more of the processing charges to be attributed to the lengthier Topical questionnaires. If a case returned multiple questionnaires, the cost-per-return would be included multiple times within the case's overall cost-per-case (once for every returned questionnaire).

For sample cases that respond to the Screener via the paper questionnaire, there is an additional sampling process that occurs to create the Topical universe and select Screener respondents to receive a Topical. This cost was not included in this analysis, which was limited to the operational costs of collecting survey data. However, this cost should be noted as a potential factor when comparing across modes of data collection, as the web respondents would not incur this cost because this sampling process occurs automatically within the web instrument.

An additional cost that was included for UAA packages that contained an incentive was the cost of recovering the incentive upon return to NPC. Using the NPC's projected number of hours handling the UAA incentives, as well as the salary information for the employees that complete this work, a cost-per-incentive-recovery-operation was calculated, then divided by the number of UAA packages containing an incentive that were returned. This resulted in an incentive-recovery-cost-per-package, which is the net cost of sending an incentive to a case that ends up being UAA, given that the cash value of the incentive is returned.

TQA Costs

The total cost of the call-center areas for NHES can be found in the FMRs. The following table shows the specific tasks that are included in the call-centers FMRs, as well as which cases would accrue each charge. All sampled cases had the option to call the TQA line to respond to the Screener questionnaire or the Topical questionnaire (in 2019 only) or to ask questions. The telephone center also provided a reminder call operation in both cycles. In 2016, Screener non-respondents received a reminder call six weeks after the fourth Screener mail-out and Topical non-respondents received the reminder call on the same day that the fourth mail-out was mailed. In 2019, the reminder call occurred on the mail-out date of the third mail-out for both Screener and Topical non-respondents. Table 5 details each TQA activity that is included on the FMRs and how each activity's cost was allocated for the cost analysis.

Table 5. Breakdown for How Cost is Allocated to each TQA Activity

Activity	Category of Activity	For Which Type of Case Does the Activity Occur	How Does the FMR Charge Get Applied?
Programming (for document tracking)	Programming	All cases[1]	Divide FMR total by the total number of sampled Screener cases to generate a cost-per-case
Postage	Check-in	Responses, UAAs, Refusals	Known, documented postage cost-per-package
Management Mail Sort General Clerical Check-in Open/sort	Check-in	Responses, UAAs, Refusals	Divide FMR total by the total number of returned packages to get a cost-per-returned-package
Batching Guillotining	Processing Preparation	Responses[2]	Divide FMR total by the total number of returned-response packages to get a cost-per-returned-response-package
Scanning Imaging Key from Image Quality Assurance	Processing	Responses	Divide FMR total by the total number of returned questionnaires pages, then for each questionnaire multiply the cost-per-page by number of pages in the returned questionnaire

[1] For NHES:2016, cases could only respond to the Screener questionnaire via TQA

Web Instrument Development

For NHES:2016, AIR was contracted to develop the web instrument for web self-responders to access during data collection. The web instrument captured both the Screener and Topical questionnaire. If a Screener respondent was eligible for a Topical questionnaire, the respondent was pushed directly from the Screener into the Topical questionnaire during the same web session. Due to confidentiality associated with contracts, Census staff do not know the cost of the web instrument developed by AIR. Census developed a separate web instrument for the Screener questionnaire that was only accessible to the TQA staff for handling Screener respondents that called the TQA line. A respondent could not complete the Topical by calling the TQA line.

In preparation for the NHES:2017 Web Test, Census developed a new web instrument for web self-responders to access during the web test, which was used in place of the web instrument developed by AIR in 2016. Similar to the 2016 web instrument, the Census-developed web instrument also captured both the Screener and Topical questionnaires. Then, for NHES:2019, extensive updates were made to the Census-developed web instrument, in keeping with the evolving changes to the survey.

To estimate the web instrument development costs for NHES:2016, the cost of the web instrument development for the 2017 NHES Web Test was used as a placeholder, assuming the cost of developing the web instrument for the 2016 cycle would have been the same had it been done by Census. This cost was determined by summing the FMR costs across the applicable months for when web development occurred during fiscal year 20016 (FY16) and fiscal year 2017 (FY17). If we did not consider this cost in our analysis, cases that were part of the mixed mode treatment group would have artificially

reduced costs as this cost component would not have been included in their costs. To determine the cost-per-case, the total FMR cost for web instrument development was divided across all web-eligible cases. In 2016, this included only the cases within the mixed mode treatment group.

For NHES:2019, only the marginal costs associated with updating the previously developed web instrument were included in the cost-per-case analysis. Continuing to include the baseline cost of developing a web instrument from scratch would lead to an over-estimation of the cost-per-case for web-eligible cases. To capture this cost, the FMRs costs from FY18 and FY19 were summed across the months in which updates were made to the web instrument. Similar to 2016, this cost was then divided across all web-eligible cases, which included all cases except those in the random paper only treatment group or the cases modeled to receive only paper in the modeled mode treatment group.

All web instrument development costs were allocated to the Screener costs, not the Topical costs.

Overhead or Other Fixed Costs

Overhead costs are considered all charges on the FMRs that do not apply to direct labor or materials associated with the physical collection of the NHES data. Unlike most mail-out costs, TQA costs, and web development costs which can be reasonably broken out and applied at the individual case-level, overhead costs are typically fixed. This means that, regardless of sample or treatment group size, these total costs would not change based on experimental treatments or adaptive interventions. Additionally, overhead costs are proportionally very large compared to most mail-out costs, telephone costs, and web development costs, which would make it difficult to determine cost differentials between different treatment groups. For these reasons, overhead costs are not included in the NHES cost analysis.

Some examples of overhead costs are:

- The cost of the program management staff, who are responsible for scheduling, budget, documentation, and management of all NHES activities
- The cost of the survey programming staff, who are responsible for all NHES data files
- The cost of frame development, sampling, weighting, estimation, or data analysis
- The cost of questionnaire, letter, or envelope cognitive testing

3.3 Cost Estimates

To estimate a total cost for each case, we can sum up the outgoing and incoming mail costs accrued by a case, as well as any phone costs from TQA line/reminder call operations. When applicable, we also include the web development cost for household that were provided the option to respond via the web. The average cost-per-case is determined by summing the cost for all cases in a particular treatment group and dividing by the number of sampled cases within that treatment group. The overall average cost-per-case includes all costs that were accrued during both the Screener and Topical portions of the survey. At the Screener level, the average cost-per-case includes only the costs that were accrued during the Screener portion of the survey. At the Topical level, the average cost-per-case includes only costs that were accrued during the Topical portion of the survey and is divided by the total number of cases sampled for a Topical

within a treatment group. Once we can assign each individual case a cost, we can start comparing the costs of different data collection modes, as well as costs differentials between years. Below, Table 6 displays the average cost-per-case at both the Screener and Topical levels for both groups of interest to our cost modeling.

Table 6. Average Cost per Case by Year

	Screener Cost		**Topical Cost**	
	2016	*2019*	*2016*	*2019*
Paper Only	$ 21.44	$ 19.97[1]	$ 22.20	$ 36.56[1]
Multimode	$ 24.24	$ 18.55[1]	$ 10.50	$ 15.53[1]

[1] Denotes statistically significant difference from the 2016 estimate at the 0.10 level

We expected the multimode data collection group to be less expensive than the paper only group across the board because we know it is more costly for a case to return a paper questionnaire to our mail processing center then to simply respond online. The only exception here is the 2016 Screener cost. The web group appears to be slightly more expensive – this is likely because of the increased web development cost in 2016 (which we discussed above in Section 3.2) and also because the web development cost is applied solely as a Screener cost. This is because as soon as a household is sampled into the survey, they are accruing that web development cost. We see that even in 2016 that slight increase in cost at the Screener phase is more than made up for in savings during the Topical phase.

Another key difference we notice between the years is the increased Topical cost for the paper only group in 2019. We found this was largely due to the increased cost for the return and intake of paper questionnaires at the mail processing center. During the time that NHES 2019 was being conducted, a government shutdown affected much of the work and charging processes at our mail processing center. We cannot say with certainty whether this is a break in the series and a one-time occurrence, or if the cost of returning and keying these questionnaires will continue to increase, but continuing this cost estimation in future cycles will help us determine that.

3.4 Limitations

While every effort was made to accurately analyze true survey costs, there are many instances where lack of information or lack of details about each cost caused limitations to this work. Using the FMRs, it was impossible to separate the costs spent on the Screener portion of the survey versus costs spent on the Topical portion survey in most instances. Further, it was impossible to separate the FMRs costs into detailed tasks, such as the cost of inserting the incentive into the mail piece or the cost of quality assurance on each different mail piece. If the FMRs could be detailed in this way, this would eliminate many of the assumptions made when dividing the costs between the detailed Screener tasks and the detailed Topical tasks.

Additionally, in some instances, necessary operational decisions get made that affect the reporting of costs inconsistently across the survey cycles. Most notable, the first month of the 2019 data collection cycle was greatly affected by the Federal Government shutdown,

which caused changes in how NPC recorded their spending on data collection tasks on the FMRs. Due to this, the 2019 costs for returning a questionnaire may be artificially inflated. Changes and improvements in how the cost estimates are calculated can also lead to measurement inconsistency between survey cycles. Unforeseen circumstances such as these can make cost comparisons between cycles more difficult.

While the cost estimate documentation was helpful for determining costs at a more detailed task level, there were a number of discrepancies between the cost estimate for a particular branch or area and the actual FMR amount that got spent in that branch or area. Therefore, instead of using the cost estimate amount, the percentage of total money spent on each task as described by the cost estimate was applied to the final FMR amount to get an estimate of true cost at the more detailed task level, such as quality assurance of the mail-outs.

4. Cost Modeling

Ultimately, our goal is to use these historic cost estimates to predict future survey costs in the NHES. To achieve this goal we have performed exploratory analyses and modeling techniques using both 2016 and 2019 survey data. We can evaluate our models by using our 2016 data to retroactively 'predict' our 2019 data. We built linear regression models using 2016 frame and response data to retroactively predict 2019 costs. We also built linear regression models with the same variables using 2019 frame and response data to compare parameter estimates between the two years. We would eventually like to refine these models, and maybe explore non-linear regression, which will be discussed further in our next steps section.

Below, Table 7 lists all of the variables we kept in our model using 2016 costs to retroactively predict 2019 costs. All of these variables were available on the frame and sample files for both years, so we know this information is available for both cycles before data collection begins. The experimental group defines whether a household was in the paper only data collection group or the multimode (web-first) data collection group. Income level defines the household's income, binned in ascending order, own/rent status defines whether the household owns or rents their home, education level defines the highest educational degree held in the household. Dwelling type defines whether the home is a single-family or multi-family home. Stratum defines a series of race/ethnicity categories. The response propensity group defines the response propensity score calculated for each household before the start of data collection. A higher response propensity score indicates that we believe that household is more likely to respond. We binned households into four response propensity groups (low, medium low, medium high, and high) at the 15th, 75th, and 95th percentiles. The child in home indicator (also from the frame) defines whether or not that household has any children. We also looked for significant interactions between these variables. For simplicity, we kept in the two most interesting/significant interactions – the interaction between the experimental group and response propensity group, and the interaction between experimental group and the child in the home indicator.

Table 7. Initial Predictive Modeling Variables

Variable Category	Number of Levels
Experimental Group	2
Income Level	14
Own/Rent Status	3

Education Level	6
Dwelling Type	3
Stratum	4
Response Propensity Group (RPG)	4
Child in Home Indicator (Child)	2
Experimental Group*RPG	8
Experimental Group*Child	4

All variables that we retained in these models contained at least one category that was statistically significant at predicting cost. While some of these categories were significant, many did not have a consistent direction across levels. The most interesting predictors were experimental group, the child in the home indicator, and the response propensity group (RPG). Below, Table 8 displays a list of select model covariates for both Screener and Topical phases in 2016.

Table 8. Select 2016 Cost Model Covariates

Variable	Screener Level		Topical Level	
	Parameter Estimate	Significance	Parameter Estimate	Significance
Intercept	21.48	<.0001	22.67	<.0001
Web Group	3.254	<.0001	-10.62	<.0001
Child in the Home	0.214	<.0001	0.790	<.0001
RPG_2 – Medium-Low	-0.387	<.0001	-0.471	0.000
RPG_3 – Medium-High	-1.353	<.0001	-1.254	<.0001
RPG_4 – High	-2.116	<.0001	-1.865	<.0001
Web Group * RPG_2	-0.327	<.0001	-0.890	0.000
Web Group * RPG_3	-0.617	<.0001	-0.822	0.003
Web Group * RPG_4	-0.907	<.0001	-0.058	0.895
Web Group * Child	-0.394	<.0001	-1.073	<.0001

Cases that were in the multimode group (the web-first indicator) are predicted to have a higher Screener cost when compared to cases within the paper group, which was expected because the web development cost is included as a Screener cost (as soon as the household becomes an active case). Because that cost is not present in the Topical phase, we then see that the cases within the multimode group ultimately have a lower predicted Topical cost when compared to the cases within the paper group. We also see that the frame variable indicating the presence of a child in the household (child) increases cost slightly in both the Screener, and a little more in the Topical phase. This makes sense, particularly in the Topical phase, because a household would be more likely to be sampled into the Topical phase after indicating the presence of a child.

We also see that in both the Screener and Topical phases the more likely a case is to respond (higher response propensity group) the less costly their total cost is predicted to be. We hypothesize that this could be from various factors. Possibly these predicted likely responders responded earlier in data collection, which could cut the need for additional mailout costs. Possibly they were also more likely to respond via the web

(cutting incoming questionnaire costs). We plan to explore these possibilities further in future research.

Using the models we built with 2016 frame variables and response data, we were able to retroactively predict 2019 costs. Below, Table 9 compares the actual 2019 costs of different data collection modes to the predicted 2019 costs using our initial 2016 prediction model. We used a simple t-test to determine statistical significance, but in future work we would like to consider implementing a bootstrap analysis to get estimates of standard deviation. Again, this initial stage of models is meant to simply give direction to our prediction models.

Table 9. Initial 2019 Predicted Costs

	Screener Cost		**Topical Cost**	
	Actual	*Predicted*	*Actual*	*Predicted*
Paper Only	$ 19.97	$ 21.41 [1]	$ 36.56	$ 22.28 [1]
Multimode	$ 18.55	$ 24.22 [1]	$ 15.53	$ 10.45 [1]

[1] Denotes statistically significant difference from the actual 2019 estimate at the 0.10 level

We see that across the board, our predictions are statistically significantly different than the 2019 cost actuals. Because we have already identified some of the cost differences between 2016 and 2019 data collection, we had a few thoughts on what could be driving some of these larger differences. We first considered the differences in the web development cost between the years. We also considered the differences between the paper intake costs between the years.

While we are unable to say with certainty whether or not the drastically increased cost of returning paper questionnaires to the mail center was a one-time occurrence, or if we will continue to see this increase, we can be fairly certain that the web development cost, moving forward, will be distributed to most cases, as NHES has made the shift to a web-first, multimode baseline. To capture that adjustment, we re-ran our model replacing the 2016 web development cost with the 2019 web development cost, which we expect to be typical moving forward. Below, Table 10 reflects the difference made when changing the 2016 web development cost to a more standard cost. This would be the 2019 cost prediction using the updated 2016 web development cost, which we see results in a much more accurate cost prediction, even though it is still statistically different.

Table 10. Updated 2019 Predicted Costs

	Screener Cost		Topical Cost	
	Actual	*Predicted*	*Actual*	*Predicted*
Paper Only	$ 19.97	$ 21.41 [1]	$ 36.56	$ 22.28 [1]
Multimode	$ 18.55	$ 17.84 [1]	$ 15.53	$ 10.45 [1]

[1] Denotes statistically significant difference from the actual 2019 estimate at the 0.10 level

In addition to retroactively predicting 2019 costs, we also built linear regression models with the same variables using 2019 frame and response data to compare parameter estimates between the two years. We wanted to see if any of these demographic domains had significantly different effects between the two years. Below, Table 11 shows some of the more interesting covariates for predicting Screener cost in 2016 on the left, and on the right those same covariates for predicting Screener costs in 2019. Again, the most interesting predictors were experimental group, the child indicator, and the response propensity group.

Table 11. Select Screener Cost Model Covariates

Variable	2016		2019	
	Parameter Estimate	Significance	Parameter Estimate	Significance
Intercept	21.48	<.0001	20.95	<.0001
Web Group	3.254	<.0001	-1.505	<.0001
Child in the Home	0.214	<.0001	0.478	0.031
RPG_2 – Medium-Low	-0.387	<.0001	-0.383	0.133
RPG_3 – Medium-High	-1.353	<.0001	-1.884	<.0001
RPG_4 – High	-2.116	<.0001	-2.511	<.0001
Web Group * RPG_2	-0.327	<.0001	0.055	0.834
Web Group * RPG_3	-0.617	<.0001	0.752	0.015
Web Group * RPG_4	-0.907	<.0001	0.675	0.149
Web Group * Child	-0.394	<.0001	-0.624	0.007

Knowing the cost differences between the data collection years, specifically the web development cost, we expected to see that the parameter estimate for being in the multimode group has switched directions when predicting Screener cost. The significantly lower web development cost in 2019 starts to show as early as the Screener phase, while in 2016 that cost savings did not show until the Topical phase.

5. Conclusions and Next Steps

The results of our initial predictive models show that operational differences from year to year can skew model performance. By continuing this cost estimation in future cycles of the NHES will help us determine whether larger cost differences between years are breaks in the series and one-time occurrences, or continual increases/decreases.

While our process for case-level cost estimation has been developed, we have a lot of ideas for how to expand our cost modeling work for the NHES. One next step in continuing this work is to refine the current prediction models. These initial models assume a normal distribution of data, but upon further inspection the data has more of a right skew. We would also like to include additional sensitivity analyses to assess our models. We might also consider weighting adjustments to ensure a parametric distribution of cases across years.

Additionally, further exploration of variable interactions could help improve model performance. Further expansions of these models might also make use of paradata. Using paradata in these predictive cost models could be beneficial and help track costs more accurately in 'live time' during data collection. We would also like to expand these models to predict for other data collection methods. Expanding these models to account for additional treatment groups would both expand the predictive power of our models and inform the survey sponsor of the cost effectiveness of various data collection instruments. Eventually, we would like to expand this work to predict data collection costs for the next cycle of NHES in 2023.

References:

[1] Coffey, S. (2020). "Bayesian Methods for Prediction of Survey data Collection Parameters in Adaptive and Responsive Survey Designs (Doctoral Dissertation)". University of Maryland.

[2] Groves, Robert. Survey Errors and Survey Costs. Hoboken, NJ. Wiley and Sons. 2004.

[3] Karr, A & Last, M. (2006). "Survey Costs (Workshop Report and White Paper)". National Institute of Statistical Sciences.

[4] Wagner, J. (2019). "Estimation of Survey Cost Parameters Using Paradata." Survey Practice 12(1): 1-10.