Health *of* Houston Survey 2017-18

HEALTH OF HOUSTON SURVEY 2017-18

Methodology Report

Institute for Health Policy
School of Public Health
The University of Texas Health Science Center at Houston

The HHS is based at the Institute for Health Policy at the University of Texas Health Science Center at Houston (UTHealth), School of Public Health.

We extend our sincere thanks to our donors. Without their generous support, this 2017-18 round of the Health of Houston Survey would have not been possible.

Houston Endowment, Inc.
Episcopal Health Foundation
Texas Children's Hospital
Memorial Hermann Health System
Community Health Choice/Harris Health System
UTHealth, President's Excellence Fund
UTHealth School of Public Health, Office of the Dean

We appreciate the experience and technical proficiency of our partners at ICF International and Dr. David Gimeno at the UTHealth School of Public Health in San Antonio. Thanks as well to the Texas Medical Center Health Policy Institute for temporary staff support. Finally, we wish to acknowledge the assistance of the many individuals and organizations who made valuable suggestions on new themes and topics for the 2017-18 cycle.

Survey design, data collection and weighting were conducted by ICF International, a nationally-recognized company in survey research, in collaboration with the Institute for Health Policy at the UTHealth School of Public Health.

Suggested citation: Health of Houston Survey. HHS 2017-18 Methodology Report. Houston, TX: Institute for Health Policy, The University of Texas School of Public Health, 2019.

The survey team is: Dr. Stephen Linder, Dr. Dritana Marko, Dr. Thomas Reynolds and Dr. Jessica Tullar.

TABLE OF CONTENTS

OVERVIEW

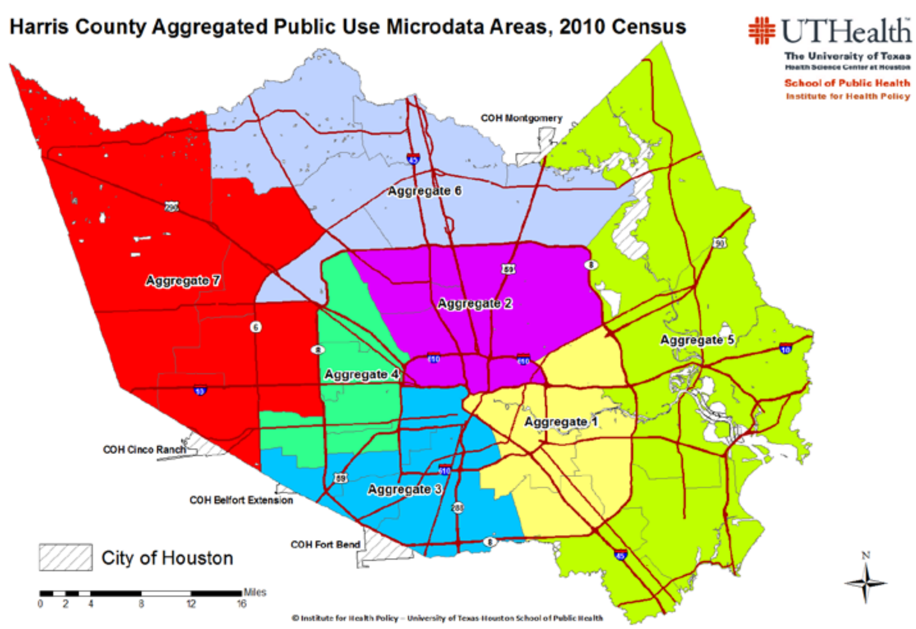
The Health of Houston Survey (HHS) is the most comprehensive household survey of public health and health care access in Harris County and the City of Houston. The survey supports communities in Houston area with their planning and evaluation efforts, assists communities and organizations in better serving their population by understanding health priorities, and enables individuals and organizations to increase funding opportunities aimed at the most serious public health needs of the population. Major public health topics covered for various segment of the population include health status and chronic conditions, health insurance and health care access, behavioral risk factors, preventive health services, mental health, neighborhood conditions, and social and economic indicators.

The HHS 2017-18 was designed to collect reliable estimates for:

1. Overall population of Harris County and City of Houston
2. Each of the seven subcounty areas created by aggregating the American Community Survey (ACS) Public Use Microdata Areas (PUMAs) in Harris County (Figure 1.)
3. Main racial and ethnic groups (Whites, Hispanics, African Americans, and Asians).
4. A range of age and income cohorts

To achieve the above objectives in the most cost-effective way, HHS employed a dual-frame Random Digit Dialing sample design, using a combination of landline phones and cellphones. HHS started data collection in June 2017 but half-way to completion, in August 2017, it had to pause due to Harvey storms making landfall in southeast Texas and devastating the area. The survey resumed again in February 2018, at which time the instrument was modified to include questions on how the Hurricane impacted the lives of Houstonians, specifically in areas related to health conditions post-Harvey, flooding and property damage, income, employment, evacuation, assistance/aid and recovery. At the same time, to accommodate for the Harvey-related questions, other questions that were not part of core questionnaire were dropped from the questionnaire.

FIGURE 1. Subcounty sampling areas



SAMPLE DESIGN

The Health of Houston Survey 2017-18 sampling goals were to produce reliable estimates for the non-institutionalized population of the Harris County and City of Houston, as well as for seven sub-county areas. In addition, a goal of the survey was to produce estimates based on sociodemographic characteristics including poverty level, age, and race/ethnicity. The sample design was a stratified, list-assisted RDD sample of landlines and cell phones, supplemented with an oversample to increase the number of Asian responses.

Original Plan: The original overall sample size was 6,000 telephone interviews, with 5,000 allocated to the base sample and 1,000 allocated to the Asian oversample. All respondents in the Asian oversample were eligible to be interviewed, even if they identified as non-Asian. By conducting 1,000 interviews in areas with high Asian populations, we increased the overall number of potential Asian respondents in the sample. In addition, an oversample was added for areas of Baytown and Pasadena to increase their sample sizes to 500 each.

Post-Harvey Plan: When interviewing resumed after Hurricane Harvey, the overall sample size was reduced to 5,500 telephone interviews, with 4,500 allocated to the base sample and 1,000 allocated to the Asian oversample. The oversamples in Baytown and Pasadena areas were eliminated. Table 1 below presents the estimated sample size and 95% confidence intervals for the aggregate areas and the target sociodemographic groups. The confidence intervals include a design effect based on oversampling of geographic areas with significant Asian populations, as well as a general inflation factor of 1.5 due to weighting. The general weighting factor includes three components:

1) Dual frame weighting effect, which we estimate at 1.05 based on the optimal allocation below
2) The weighting effect from within household sampling and multiple telephone households, estimated to be 1.20 based on 2015 Texas BRFSS
3) The weighting effect due to raking to the population, estimated to be 1.20 based on Texas BRFSS.

Combined, these components result is a design effect of roughly 1.5 (DEFF = 1.20 x 1.20 x 1.05 = 1.51).

TABLE 1: Original Sample Size and Estimated 95% Confidence Intervals

	N	n	Deff	+/-95% CI
Harris County, TX	4,269,608	6,000	1.61	1.60%
1 - East central	561,035	689	1.53	4.60%
2 - North central	561,375	677	1.52	4.60%
3 - South central	676,973	1,073	1.66	3.90%
4 - West central	548,929	873	1.66	4.30%
5 - East	547,122	733	1.58	4.60%
6 - North	699,622	947	1.58	4.00%
7 - West	674,552	1,008	1.62	3.90%
Households w/kids	572,196	2,336	1.64	2.60%
18-64	2,723,073	5,285	1.61	1.70%
65+	371,250	715	1.61	4.60%

	N	n	Deff	+/-95% CI
<100% Poverty level	778,703	1,060	1.62	3.80%
100-199% Poverty level	910,278	1,255	1.63	3.50%
≥200% Poverty level	2,580,627	3,685	1.61	2.00%
Hispanic	1,766,483	2,349	1.58	2.50%
Non-Hispanic Black	789,802	1,081	1.6	3.80%
Non-Hispanic Asian	272,171	511	1.63	5.50%
Non-Hispanic White	1,369,752	1,953	1.61	2.80%

N = population totals; n = sample totals; Deff = design effect; CI = confidence interval

During the post-Harvey hiatus period, the study's targets were revised to accommodate for impact of the hurricane on landlines availability and added questions regarding the Hurricane, bringing the total number of completes to 5,500.

TABLE 2: Post-Harvey Sample Size and Estimated 95% Confidence Intervals

	N	n	Deff	+/-95% CI
Harris County, TX	4,269,608	5,500	1.90	1.8%
1 - East central	561,035	624	1.79	5.2%
2 - North central	561,375	611	1.77	5.3%
3 - South central	676,973	994	1.98	4.4%
4 - West central	548,929	808	1.97	4.8%
5 - East	547,122	669	1.86	5.2%
6 - North	699,622	865	1.86	4.5%
7 - West	674,552	929	1.92	4.5%
Households w/kids	572,196	2,138	1.93	2.9%
18-64	2,723,073	4,845	1.90	1.9%
65+	371,250	655	1.90	5.3%
<100% Poverty level	778,703	968	1.91	4.4%
100-199% Poverty level	910,278	1,147	1.92	4.0%
≥200% Poverty level	2,580,627	3,385	1.90	2.3%
Hispanic	1,766,483	2,142	1.86	2.9%
Non-Hispanic Black	789,802	988	1.88	4.3%
Non-Hispanic Asian	272,171	479	1.93	6.2%
Non-Hispanic White	1,369,752	1,793	1.91	3.2%

N = population totals; n = sample totals; Deff = design effect; CI = confidence interval

RDD SAMPLE ALLOCATION

The sample allocation was optimized to minimize the variance of the dual-frame composite estimator, as outlined in Lohr and Brick[1] (2014). Our allocation was based on:

1) A cell to landline cost ratio of $2:1

[1] Lohr, Sharon L, and J M Brick. 2014. "Allocation For Dual Frame Telephone Surveys with Nonresponse." *Journal of Survey Statistics and Methodology* 388-409.

2) 46% of households in Harris County (46%) were cell-only[2] in 2015
3) 60% of cell phone surveys were cell-only and 20% of landline surveys were landline-only.

The optimal sample allocation under these assumptions was 58.7% cell phone, which we rounded to 60%. With this allocation, we expected 36% of respondents to be cell-only to represent the estimated population of 46% (see Table 3).

After fielding the first three waves of the study, we observed that rate for the cell-only was higher than expected at 70% of respondents. The percentage of landline respondents reporting they were landline only was lower than expected at 15%. Further, the cost of conducting a cell phone sample relative to a landline sample was lower than expected ($1.5:1). This resulted in an optimal allocation of 45% landline and 55% cell phone. We increased the landline sample to 50% to allow for better targeting of the aggregate areas (as described below). Based on 70% of the cell phone completes reporting no landline in their household, we expected 35% of the total sample to be cell-only. This increase in landline sample results in a small increase in the design effect (3% overall).

After about one month of fielding post-Harvey, protocols were further revised as productivity on the landline sample proved to be significantly lower than expected. This was likely due to the ongoing impact of the hurricane, as Houston residents continue to rebuild—moving homes or shutting down their phone lines in the process. The cost to conduct a landline survey is now more than a cell phone survey. This change in the cost of the survey would result in an optimal allocation of 70-75% cell phone. After consideration, the study team decided to revise the sample stratification to 75% cell and 25% landline. This change should also encourage more completes among demographic groups that have heretofore been under-represented, including Hispanics, Asians and males.

TABLE 3: Expected Distribution by Phone Status

	18+ Population Benchmark	Total Sample	Cell Sample	Landline Sample
Original				
Cell-only households	46.0%	36.0%	36.0%	N/A
Landline households	54.0%	64.0%	24.0%	40.0%
Total	*100%*	*100%*	*60.0%*	*40.0%*
Revision (75% Cell)				
Cell-only households	46.0%	52.5%	52.5%	N/A
Landline households	54.0%	47.5%	22.5%	25.0%
Total	*100%*	*100%*	*75.0%*	*25.0%*

The percentage of Hispanics in the cell phone sample is 35%, compared to 13.6% for landline. Similarly 6.1% of cell phone completes are Asian (vs 2.7% for landline). In effect, the higher the cell allocation, the higher the Asian and Hispanic sample sizes. While the Asian and Hispanic populations are still

[2] MSG's April 2015 estimate for the state of Texas is 46.6%, which is close to the 2013 annual estimate of 48.4% released by the National Health Interview Survey. Given the similarity between the state estimates, we are confident in the Harris County estimate provided by MSG. The state estimates from these two sources are available at http://www.m-s-g.com/CMS/ServerGallery/MSGWebNew/Documents/GENESYS/wireless-estimates/wireless-estimates-04-15.pdf and http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless_state_201412.pdf, respectively.

underrepresented relative to the population (as is typical in RDD), the effect is reduced with higher cell allocation.

SAMPLING FRAMES

The North American Numbering Plan Administration governs the assignment of 1,000-blocks to service providers. A 1,000-block is the series of 1,000 telephone numbers defined by the last three digits of a 10-digit phone number (NPA-NXX-Z000 to NPA-NXX-Z999). The 1,000-blocks dedicated to cell service or landline service were identified by codes from the Telcordia® Local Exchange Routing Guide. Those dedicated to landline service comprised the landline frame, while those dedicated to cellular service comprised the cell phone frame.

Selecting the Landline Sample

We selected the landline sample using RDD with equal probabilities of selection (EPSEM) from working banks associated with Harris County. A "working" bank is a 100-block (NPA-NXX-ZZ00 to NPA-NXX-ZZ99) in which at least one telephone number is assigned to residential service. Note that this frame definition is an improvement over traditional list-assisted frames, which only include blocks with one or more "listed" telephone numbers. By excluding zero-blocks, the traditional list-assisted frame typically excludes about 5% of residential households.[3] The assignment-based frame included these households that would have otherwise been excluded.

Telephone lines are not restricted by geographic borders, but are generally associated with particular geographic areas. Each 1,000-block of telephone numbers is associated with a single geographic area by tallying the number of geocoded landline households in each geographic area. The 1,000-block is assigned to the geographic area with the greatest number of geocoded telephones (the rule of plurality). The landline frame included 4,379 1000 blocks associated with Houston/Harris County and the number of 1+ working banks was 35,143. Known business listings were removed from the sample. The landline frame was stratified based on the aggregate area to ensure the sample was allocated proportionately across the County. Landline 1000-blocks were assigned to an aggregate area based on the census tract definitions for the aggregates. The estimated sample sizes are presented in Table 4a. The frame changed slightly for the post-Harvey fielding (Table 4b.).

After purging landline numbers for known businesses, they were matched against the Neustar database to determine whether they have been ported to cell phone. If they had, they were included as part of the cell sample for the interview. Non-working landline numbers were removed at the time of fielding by our automated dialing system.

TABLE 4a: Estimated Landline Sample Size by Aggregate – Pre-Harvey

Aggregate Sample	RDD Frame	Landline Assignments	Working Number Density	Expected Sample Size
Total	3,514,300	674,274	19.2%	307224
1 - East central	348,900	53,255	15.3%	31915

[3] Boyle, J., Bucuvalas, M., Piekarski, L., & Weiss, A. (2009). Zero Banks: Coverage Error and Bias in RDD Samples Based on Hundred Banks with Listed Numbers. *Public Opinion Quarterly*, 673: 729–750.

Aggregate Sample	RDD Frame	Landline Assignments	Working Number Density	Expected Sample Size
2 - North central	505,300	92,672	18.3%	43787
3 - South central	829,100	155,055	18.7%	52516
4 - West central	499,200	58,528	11.7%	37188
5 - East	416,300	91,580	22.0%	36461
6 - North	483,700	115,772	23.9%	34647
7 - West	431,800	107,412	24.9%	70710

TABLE 4b: Estimated Landline Sample Size by Aggregate – Post-Harvey

Aggregate Sample	RDD Frame	Landline Assignments	Working Number Density	Expected Sample Size
Total	3,531,900	655,672	18.6%	307224
1 - East central	348,300	49,664	14.3%	31915
2 - North central	507,500	90,069	17.8%	43787
3 - South central	832,300	150,433	18.1%	52516
4 - West central	500,700	55,283	11.0%	37188
5 - East	418,000	90,542	21.7%	36461
6 - North	488,000	114,492	23.5%	34647
7 - West	437,100	105,189	24.1%	70710

Selecting the Cell Phone Sample

We selected the cell phone sample using RDD with EPSEM. All telephone numbers from the cellular frame were manually dialed in accordance with laws that prohibit cell numbers from being called by an automated dialer, including the Telephone Consumer Protection Act (TCPA). We identified rate centers (midpoint of the rate area, generally a town or a group of towns, covered by a bank of telephone numbers) associated with Harris County (Figure 2). The rate center represents the geographic location where the cell number was originally assigned. While cell phones are portable to other geographic locations, the location of the rate center is an indicator of the location of the cell phone. Using the rate centers in Harris County, the cell frame included 8,208 1000-blocks.

We pre-screened the sample using MSG's CellWINS, a non-intrusive process to identify whether the cell phone is active or inactive. Inactive numbers were excluded from the sample. The cell phone sample is a 2-phase sample. The first phase is a sample of cell phone numbers from the cell RDD frame. These numbers were matched to a database containing geographic information for the billing address associated with the cell phone number and stratified as matching a geographic location in Harris County ("match-in"), matching a geographic location outside Harris County ("match-out"), or did not have a matching record in database ("unmatched"). In phase 2, we selected a subsample of the match-out cases to increase the efficiency of reaching residents of Harris County. Table 5 includes the expected cell phone sample sizes. The expected sample sizes assumed a match rate of 60% (40% unmatched), with 85% of the numbers matching a Houston zip code (15% match-out.) The results from the first three waves were 50% match, with 60% matching inside the county. The expected cell phone sizes have been updated to reflect the results of the first three waves.

A final cell phone stratum included the out-of-area numbers. These are the cell phone numbers that have billing geography in Harris County, but do not originate from a Harris County rate center. This sample was selected from Survey Sampling International's SmartCell.[4] The total number of out-of-area cell phone numbers was 327,732. The selected cell phone numbers were matched against the Neustar database to determine whether they had been ported to landline. If they had, they were included as part of the landline sample for the interview.

FIGURE 2. Rate Centers in Harris County, Texas

TABLE 5: Expected Cell Phone Sample Sizes

Stratum	**Frame size**	**Phase 1**		**Phase 2**		
		Sample	*Active*		*Match*	*Sample*
Harris County Rate Centers	8,107,000	137,500	112,500	Match-in	33,750	33,750
				Unmatched	56,250	7,800
				Match-out	22,500	0
Outside Harris County Rate Centers	327,732	1500	1500		1500	1500

Asian Oversample

Approximately 6.3% of the Harris County population is Asian. This is considerably smaller than the other population groups, Hispanic (41.4%), black (18.5%) and white (32.1%). To increase the sample size of Asian people, we oversampled geographic areas with a high percentage of Asians, based on data from the 2010-2014 American Community Survey. Block groups where Asians comprise 10% or more of the population were included in the oversample. There were a total of 443 block groups in the county that met this criteria. These block groups represent 68.6% of the total Asian population.

[4] https://www.surveysampling.com/about/news/2016/ssi-launches-smart-cell-sample-increasing-incidence-rates/

Once the block groups were identified, the geographic areas were translated to landline and cell phone sampling frames. For the landline, the 1,000 blocks associated with one of the high Asian block groups constituted the oversampling frame. There were 1,738 1000-blocks associated with the high Asian block groups. The cell phone frame was based on cell phone numbers where the billing zip code is geocoded to one of the high Asian block groups. The cell phone frame was based on SSI's SmartCell.[5] The Asian oversample frame count total was 145,952.

As the fielding period drew to a close, the number of completes with self-identified Asian respondents was below target. To remedy this, we selected a sample of cell phone numbers that SmartCell flagged as likely to lead to an Asian respondent. The Asian sample flag is based on an algorithm that used first name, last name, and geographic location. We completed 171 additional interviews, 109 of these were with Asian respondents.

Pasadena and Baytown Oversample

Original Plan: The Pasadena and Baytown oversamples were selected for both landline and cell phone. The landline oversample was based on RDD sample of 1,000 blocks (NPA-NXX-Z000 to NPA-NXX-Z999) associated with the zip codes that make up Pasadena (77034, 77075, 77089, 77502, 77503, 77504, 77505, 77506, 77587) and Baytown (77520, 77521, 77523, 77562). Each 1,000-block of telephone numbers is associated with a single geographic area (e.g. a zip code) by tallying the number of geocoded landline households in each geographic area. The 1,000-block is assigned to the geographic area with the greatest number of geocoded telephones (the rule of plurality). The geographic association of 1000-blocks to geography was performed by our sampling vendor (Marketing Systems Group). There were 208 1000-blocks making up the Pasadena frame and 135 1000-blocks for the Baytown frame.

For the cell phone, we selected an oversample of the two areas from Survey Sampling International's (SSI) SmartCell frame. The sample is based on numbers whose cell phone billing address is located in either the Baytown or Pasadena zip codes. Both the cell phone and landline oversamples were unduplicated against the base sample, so that no number is sampled twice. The completed interviews from the base sample with Pasadena zip codes would be combined with the completed interviews from the Pasadena oversample for a total of 500. Similarly, the completed interviews from the base sample with Baytown zip codes would be combined with the completed interviews from the Baytown oversample for a total of 500.

Note: *The original plan for oversampling Pasadena and Baytown was cancelled when the study resumed after Hurricane Harvey.*

Within Household Selection

For the landline sample, one adult within the household was randomly selected for the survey. The selection was based on the Rizzo, Brick, and Park (RBP) selection method:[6]

- If there was one adult, that person was automatically selected.
- If there were two adults, the screener adult was selected one-half of the time. If the screener adult was selected, the interview continued with that adult about him or herself. If the screener

[5] https://www.surveysampling.com/about/news/2016/ssi-launches-smart-cell-sample-increasing-incidence-rates/

[6] Rizzo, L., J.M. Brick, and I. Park, A minimally intrusive method for sampling persons in random digit dial surveys. Public Opinion quarterly, 2004. 68(2): p. 267-274.

adult was not selected, we asked to speak to the other adult in the household and complete the interview if possible, or schedule a callback.

- If there were three or more adults, we asked to speak with the adult in the household who had the most recent birthday and complete the interview, if possible, or schedule a callback.

To randomly select a reference child for the children questions, we asked the number of children in the household, C. We randomly selected a child 1, 2, …, C, where 1 represents the oldest child and C represents the youngest child after asking the respondent to think about the children in order of their birth, from oldest to youngest.

For cell phone, we did not randomly select an adult since cell phones are generally personal devices. We selected a reference child as above. A $5 incentive was offered to each participant for completing the survey. An additional $5 incentive was offered to those that conduct the survey on a prepaid cellphone.

SET-UP

Testing the program

ICF conducted testing of the Computer-Assisted Telephone Interviewing (CATI) program by going through the programmed instrument multiple times, with several different testers, and following a number of different scenarios. Through this process, testers verified that the programmed text, logic, page breaks and formatting completely match the questionnaire as it was laid out in the programmer's document. Content changes to questions were only be made if deemed absolutely necessary. Testers paid careful attention to:

1. Skip logic;
2. Advancement between pages and questions (i.e. question validations);
3. Dispositioning; and
4. Differences in programming between landline and cell samples

The ICF Internal testing plan included a two-week testing period with three rounds of changes. IHP staff tested a web-based "demo" of the survey and communicated any necessary changes to the vendor. Following Hurricane Harvey, the survey instrument underwent changes to measure the storm's impact and issues related to Houston's ongoing recovery. ICF and UT tested changes to the instrument before resuming fielding.

Interviewer Training

All Health of Houston Survey interviewers received a two-hour training specific to the project, followed by 2-3 hours of practice and observation before fielding. An interviewer training was conducted prior to the pretest and updates to the training materials were made for the full-scale launch. The training manual included the following topics.

- Purpose and scope of the survey
- Detailed review of questions
- Probing
- Dealing with uncooperative respondents, including avoiding refusals and unnecessary break-offs.
- Role-Playing/Practice Interviews
- Proper dispositioning
- Reading verbatim: Interviewers are trained to read all the text on their screen verbatim. Interviewer instructions and text in parentheses are optional and can be read if the respondent is confused or needs additional information.

ICF Staff conducted a third interviewer training to cover new and revised questionnaire items following Hurricane Harvey. The IHP staff attended the pre-test and/or full-scale interviewer trainings via teleconference.

DATA COLLECTION

Pretest

After the first training, interviewers, under close supervision, began dialing the sample and obtained 34 English and four Spanish completed interviews over a four-night dialing period. All interviews were recorded.

We gathered feedback on the pretest in three ways:

1. Examining paradata and survey item data for high numbers of "don't know", refusals and breakoffs because these may be indicative of a problem area within the question that should be revised. If the team found that a question was frequently skipped, refused, or answered in a way deemed inconsistent, we flagged it for review.
2. Collecting interviewer and call center feedback after each interview documenting details of note about the call including areas of confusion or other trouble spots regarding screening, securing cooperation, or conducting the interview: At the end of each evening, interviewers discussed the calls with the QA supervisor during a debrief session. Additionally, ICF and IHP explored the possibility of having clarifying questions at the end of a specified module, as well as at the end of the questionnaire to obtain information from respondents directly.
3. Document PM feedback from recordings: Project staff listened to a number of interview recordings to note any apparent issues with comprehension or flow.

ICF reviewed pretest data and summarized these into a short report that contained general findings, areas of a concern, and an appended questionnaire with track changes. One concern that emerged during the pretest was survey length. Following questionnaire revisions a second short pretest was conducted to evaluate length within true field conditions. Pretest #2 took place in early June 2017 and achieved 10 completes.

Dialing Protocols

The Dialing protocols for the Health of Houston Survey 2017-18 followed a suggested monthly interviewing schedule; all calls for a given survey month should be completed in the same sample month if possible. In some cases samples begun in one month could be completed in the first 7-10 days of the next month. It was possible to make up to 15 calling attempts for each landline phone number and up to 8 for each cell phone number in the sample. Calling attempts are described below:

- Cell protocol was eight attempts => 2 day, 3 night and 3 weekend.
- Landline protocols was 15 attempts => 3 day, 3 night, 3 weekend and 6 additional attempts which could be anytime night or weekend.
- Calling hours were Mon-Fri 9am-9pm and Sat. and Sun. 10am-9pm.

The call centers also changed schedules to accommodate holidays and special events, made weeknight calls after 5:00 PM CST and adhered to respondents' requests for specific callback/appointment times whenever possible.

Attempt protocols were adjusted for the post-Harvey relaunch:

- Cell protocol included five maximum attempts.
    - 1 daytime; 2 evening; 2 weekend.
    - 1 refusal to terminal disposition.

- 1 voicemail message left on first attempt.

- Landline protocol consisted in ten maximum attempts.
  - 1 daytime; 3 evening; 2 weekend; 4 anytime.
  - 2 refusals to terminal disposition regardless of respondent selection.
  - 2 voicemail messages left on first and fifth attempt.

The landline protocol was further reduced to eight attempts as part of our transition to a cell-only protocol in the final two months of fielding.

Dispositioning

Each telephone number in the sample was assigned a final disposition code to describe the result of the call:

- A completed or partially completed interview (determined at variable DPA_8) or
- A determination that:
  - A household was eligible to be included but an interview was not completed or
  - A telephone number was ineligible or could not have its eligibility determined.

A list of standard dispositions can be found in Table 6. If a record received a terminal disp, that record was deemed resolved and was removed from dialing. Temporary dispositions (callbacks, answering machines) kept the record in the active sample to move forward in the protocol.

Disposition counters tracked the types of dispositions, to allow certain combinations of temporary dispositions to create a terminal disposition. The best example of this is two soft refusals on a record: When the interviewer dispositioned the record as a soft refusal for the second time, it was automatically dispositioned as a hard refusal in real time. Any record that did not have a final disposition was redialed through protocol.

TABLE 6: Standard Temporary Dispositions of the Health of Houston Study

Type	Temporary Disposition	AAPOR Final Disposition Code[7]	AAPOR Disposition Category Assignment
Callback	Scheduled - Selected Respondent	2.21	Eligible, Non-Interview
	Scheduled - Non Selected Respondent	2.21/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Unscheduled - Selected Respondent	2.21/2.3	Eligible, Non-Interview
	Unscheduled - Non Selected Respondent	2.21/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Dead Air	3.16	Unknown Eligibility, Non-Interview
	Busy	3.12	Unknown Eligibility, Non-Interview
	Ring No Answer * *Terminate after 10 consecutive*	3.13	Unknown Eligibility, Non-Interview
	Answering Machine - Household	2.22/3.14	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Answering Machine - Not A Residence	4.5	Not Eligible
	Answering Machine – Unknown **Left on 1st, 4th and 9th attempt*	2.22/3.14	Eligible, Non-Interview/Unknown Eligibility, Non-Interview

[7] When two codes are shown, the first corresponds to known eligibility, the second to unknown eligibility.

Type	Temporary Disposition	AAPOR Final Disposition Code[7]	AAPOR Disposition Category Assignment
	Dialer/TCPA Unclassified code	4.4	Not Eligible
	Dialer/TCPA hung up	3.16	Unknown Eligibility, Non-Interview
	Time out	3.16	Unknown Eligibility, Non-Interview
	Disconnected by supervisor	3.16	Unknown Eligibility, Non-Interview
Refusal	Hang Up ** If the respondent hangs up before the interviewer has finished reading the client name, it is coded as a hang up. Otherwise, it was coded as a refusal.*	2.12/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Refused - Selected Respondent ** A soft refusal by the selected respondent was re-attempted, with a 3 day cool-off period. Two soft refusals are coded as a hard refusal*	2.112/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Refused - Non Selected Respondent	2.111/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Hard Refusal - Selected Respondent ** A hard refusal on the first attempt terminates the record*	2.112/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Hard Refusal - Non Selected Respondent	2.111/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
Communications Barrier	Lang Barrier - second attempt needed ** First attempt gets a Language Barrier disp (non terminal)*	2.33/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Lang Barrier – terminal ** Second attempt carried out by Spanish speaking interview (the subcontractor). A 2nd Lang Barrier attempt is terminal*	2.33/3.211	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Physical/Mental Impairment - Selected Respondent ** Terminate*	2.32	Eligible, Non-Interview
	Physical/Mental Impairment - Non Selected Respondent	3.211	Unknown Eligibility, Non-Interview
	Bad Audio Connection	3.16	Unknown Eligibility, Non-Interview
Technical Barrier	Nonworking ** Terminate*	4.3	Not Eligible
	Fax/Modem	4.2	Not Eligible
	Temporarily Disconnected	4.33	Not Eligible
	Privacy Manager - Household	2.22/3.14	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
	Privacy Manager - Not A Residence	4.5	
	Privacy Manager - Unknown	2.22/3.14	Eligible, Non-Interview/Unknown Eligibility, Non-Interview
Ineligible Sample	Not a Residence	4.5	Not Eligible
	Not a Land Line / Cell Phone	4.42	Not Eligible
	Household Unavailable	3.2	Unknown Eligibility, Non-Interview
	No Adults Associated w/Line	4.7	Not Eligible
Project Specific Dispositions	Not a resident of Houston or Harris County • *If City = 03* *If ZIP = Not allowable*	4.1	Not Eligible
	Could not determine if within study area	3.2	Unknown Eligibility, Non-Interview

Type	Temporary Disposition	AAPOR Final Disposition Code[7]	AAPOR Disposition Category Assignment
	• *If HC/COH BRDR or COU or CITY = 98, 99*		
	Could not determine number of adults	3.2	Unknown Eligibility, Non-Interview
Complete	Complete	1.1	Interview
	Partial Complete • *Through DPA_8*	1.2	Interview

The final disposition codes were then used to calculate response rates (AAPOR Response Rate 4), cooperation rates (AAPOR Cooperation Rate 2), and refusal rates (AAPOR Refusal Rate 2). Data collectors were required to follow the rules for assigning disposition codes, and train and monitor interviewers in the use of specific dispositions.

Completes vs Partials

An interview was considered complete if data was collected for all questions. Partially completed interviews are defined as those where three quarters of the interview were completed, including major portions of sociodemographics. If the respondent did not provide responses for weighting variables, imputed values for these variables were generated. A partial complete interview was defined as having completed questions through variable DPA_8 (barriers to physical activity).

Eligibility

An eligible household was defined as a housing unit that has a separate entrance, where occupants eat separately from other persons on the property, and that is occupied by its members as their principal or secondary place of residence. The following were non-eligible households: group homes, institutions, and (in the landline telephone sample) households outside of the Houston area or Harris County, Texas. The Health of Houston was a self-reported survey; if respondents reported that they live in private residences, the interviewers did not question them about their residence.

Eligible household members included all related adults (aged 18 years or older), unrelated adults, boarders/roomers, live-in au pairs or students and domestic workers who consider the household their home, even though they may not be home at the time of the call. College housing residents are treated as single adult households. Household members did not include adult family members who were currently living elsewhere.

Respondent Selection

The interviewer asked the respondent to report the number of adults living in the household:

1. If there were no adults in the household, the interview terminated.
2. If there was 1 adult, that person was selected to complete the interview.
3. If there were 2 adults, one of the two adults was randomly selected.
4. If there were 3 or more adults, the interviewer asked which of the adults has the next birthday coming up. That person was the selected respondent.

If the selected respondent was not available, the interviewer moved to scheduling a call-back interview. Cell phone respondents are asked if it was a safe time to talk (whether the cell respondent was driving or in a place where speaking could jeopardize their confidentiality).

Refusal conversion

With the exception of hard refusals, eligible people who initially refuse to be interviewed were contacted at least one additional time and given the opportunity to be interviewed. Preferably, this second contact was made by a supervisor or a different interviewer. Generally, a period of two days between the initial refusal and second attempt was standard protocol. Interviews were coded as "**Hang-up"** when respondent hung-up without hearing the survey introduction, and therefore did not know who we were or why we were calling. We use a "**Refusal"** disposition if the respondent hung up after the interviewer identified themselves as calling on behalf of the University of Texas School of Public Health. Interviewers left a message for the next interviewer when the screen prompts them to leave one. We employed a standard of three refusals from the non-selected respondent, and two from the selected respondent.

Cell records and Cell-phone Flag

Interviewers manually dialed all cell phone numbers. Interviewers disconnected their phones from computer systems that have autodialing capability while calling cell samples, as hand dialing alone is not sufficient to comply with TCPA regulations. The sample included a flag on any cell records where the respondent has a pre-paid cell phone. For these records, the CATI script was modified to offer the participant a $10 Amazon gift code at the end of the survey to reimburse them for their cell phone minutes.

Answering machine and privacy manager

Messages left on answering devices/voicemail devices were left by interviewers and not by any automated voice devices. In the voice message, the interviewer described the reasons for the call and when respondents might expect a return call. Messages were left after any attempt, however, messages were not left after every attempt, so that respondents would not be burdened by repeated messages. The survey was programmed to display the answering machine message on the 1st, 4th and 9th attempt (up to 3 times for LL and up to 2 times for cell).

IVR and callbacks

ICF set up a toll free 1-800 Interactive Voice Response (IVR) Respondent Helpline and created a menu of options in both Spanish and English. Respondents could hit redial on the local Houston number on their caller ID and it connected to the IVR. Voicemails were checked multiple times throughout a shift and respondents were called back immediately if a callback was requested. If respondents called after calling hours, they heard the IVR prompts and could leave a message if an agent was not available. The IVR script was shared with UT for approval before recording the prompts.

Incentive distribution

Incentives were read out to respondents at the end of the survey. The incentive was a $5 gift code, which is a unique set of 14 numbers and letters which can be used by the respondent to purchase items online at amazon.com. The codes never expire, and the interviewer provided instructions on how to store and use the code. Furthermore, the IVR menu included a path for respondents to report incentive problems or ask additional questions after the interview is over. Respondents who had previously refused to participate ("refusal conversions") as well as respondents with a pre-paid cell phone flag received a $10 code. A total of 1,082 x $10 codes were distributed during the fielding period.

SUBCONTRACTOR ACCESS AND LOGISTICS

For the Health of Houston Survey, our survey vendor, ICF, subcontracted with an experienced historically underutilized business (HUB) with extensive CATI facilities called Customer Research International (CRI). CRI has 90 CATI stations located in Houston. ICF maintained the CATI program and sample on their servers, and CRI used remote call center VoIP communication services to access ICF's systems via a secure cloud system. In this way, all dialing and interviewing was integrated. ICF trained their interviewing counterparts at CRI and monitored them alongside CRI supervisors. ICF's CATI systems enabled case tracking and the ability to determine the progress of all sample numbers assigned to CRI. If any numbers seemed to be problematic, project managers could pull them back to the ICF system for follow-up.

ICF's calling team was assigned to complete 580 interviews, while CRI's team was assigned to complete 4,920, including all the Spanish-language interviews. ICF created a special disposition to enable timely call-backs from CRI in Spanish. ICF and Recon call center supervisors monitored 10% of all interviews (include those conducted by the subcontractor). All respondents were notified at the beginning of the call that the call may be recorded.

- QA/on-site minimum monitoring targets, 10% using live recordings
- PM supervisions/monitoring

During the interviewer training, we discussed a list of frequently asked questions. Interviewers could access these at any point during the interview.

OUTCOME RATES CALCULATION

Response rates

The overall response rate for the Health of Houston Survey is a composite of the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed) and the extended interview completion rate (i.e., success in getting the selected person to complete the extended interview). The Response rate formula is AAPOR Response Rate 4 (RR4), defined below:

$$RR4 = \frac{(I + P)}{(I + P) + (R + NC + O) + e(UH + UO)}$$

Cooperation rates

$$\text{COOP2} = \frac{(\text{I} + \text{P})}{(\text{I} + \text{P}) + \text{R} + \text{O}}$$

Refusal rates

$$\text{REF2} = \frac{\text{R}}{(\text{I} + \text{P}) + (\text{R} + \text{NC} + \text{O}) + \text{e}(\text{UH} + \text{UO})}$$

RR = Response rate
COOP= Cooperation rate
REF = Refusal rate
CON = Contact rate
I = Complete interview (1.1)
P = Partial interview (1.2)
R = Refusal and break-off (2.10)
NC = Non-contact (2.20)
O = Other (2.30)
UH = Unknown if household/occupied HU (3.10)
UO = Unknown, other (3.20)
e = Estimated proportion of cases of unknown eligibility that are eligible

Overall and by frame type response rates are depicted in Table 7.

TABLE 7: Outcome Rates of the Health of Houston Study

		Resolution Rate	Interview Completion Rate	Cooperation Rate	Refusal Rate	Response Rate
Cellphone Sample	Billing location in Harris County	16.7%	90.9%	70.6%	1.2%	11.8%
	Billing location not in Harris County	10.0%	87.3%	68.9%	1.0%	6.9%
	No billing match	26.2%	90.9%	66.8%	1.8%	17.5%

		Resolution Rate	Interview Completion Rate	Cooperation Rate	Refusal Rate	Response Rate
	Oversample	13.3%	91.8%	70.3%	0.8%	9.3%
Landline Sample	East central	90.6%	86.5%	57.6%	8.1%	52.1%
	North central	88.1%	87.2%	65.3%	8.4%	57.1%
	South central	87.5%	80.5%	56.3%	11.9%	49.3%
	West central	91.8%	82.0%	58.7%	11.7%	53.7%
	East	86.6%	81.0%	54.8%	11.2%	47.5%
	North	85.1%	86.4%	57.3%	7.7%	48.8%
	West	82.4%	84.3%	57.1%	8.8%	47.0%
	Oversample	77.5%	82.1%	59.1%	10.0%	45.8%
Combined	Landline	86.7%	83.0%	57.1%	10.2%	49.5%
	Cell	39.0%	91.0%	69.6%	2.7%	27.1%
	Overall					**32.8%**

WEIGHTING

Survey weights were computed to correct for disproportionate sampling probabilities introduced by the sampling design, including unequal probabilities due to the dual-frame sample and Asian oversample; and to correct for differences in demographic characteristics of the sample versus the population, reducing the risk of nonresponse and coverage biases in substantive estimates that may be associated with those demographics. The weighted dataset included six weights:

TABLE 8: Weights included in final dataset

FINAL_WT_ADULT	Final weight adult
FINAL_WT_ADULT_PRE	Final weight adult pre-Harvey
FINAL_WT_ADULT_POST	Final weight adult post-Harvey
FINAL_WT_CHILD	Final weight child
FINAL_WT_CHILD_PRE	Final weight child pre-Harvey
FINAL_WT_CHILD_POST	Final weight child post-Harvey

We calculated the weights in three steps: 1) calculating cell and landline design weights, 2) combining the cell phone and landline samples, and 3) population calibration (i.e. post-stratification and raking).

Design Weights

The design weights were computed as the inverse of the probability of selection of the phone number from the sampling frame (landline and cell phone). The selection probability has two components, the base probability of selection and the Asian oversample probability of selection. We combined the two selection probabilities into a joint probability of selection.

Landline Weight

The base landline phone sample is selected by drawing n_L landline phone numbers from N_L numbers on the frame for each of the aggregate strata. The sample selection probability is calculated as:

$$Pr_B(L)=(n_L/N_L).$$

The landline sample for the Asian oversample were selected from telephone exchanges associated with block groups that have 10% Asian population or more. The landline probability of selection is calculated based on the telephone numbers selected divided by the frame size, $Pr_A(L)=(n_{LA}/N_{LA})$. Further, although the Baytown and Pasadena samples were eliminated post-Harvey, there was pre-Harvey sample target to these areas. These oversamples are also adjusted for through the probability of being selected in the town oversample, $Pr_T(L) = (n_{LT}/N_{LT})$. The joint probability of selection for the landline sample is:

$$\begin{aligned} Pr(L) &= Pr_B(L) + Pr_A(L) + Pr_T(L) \\ &- Pr_B(L) * Pr_A(L) - Pr_B(L) * Pr_T(L) – Pr_A(L) * Pr_T(L) \\ &+ Pr_B(L)*Pr_A(L)*Pr_T(L). \end{aligned}$$

This probability represents the probability that a landline number is selected in the base sample or the Asian oversample. For telephone numbers selected from banks that are not in the Asian oversample frame, $Pr_A(L) = 0$ and the landline probability is equal to the probability of selection for the base sample. The sampling weight for the landline sample is the inverse of the selection probability, W1 = 1/ $Pr(L)$. For the landline sample, we made two adjustments to the weights to compute probability of selecting an adult.

1. Within household selection. We randomly select one adult within each household to complete the survey. Therefore, the within household sampling weight is equal to the number of adults eligible for the survey.
2. Adjustment for multiple phone lines. Since households are selected with probability proportional to their number of telephone numbers, we adjust for multiple phone lines.

Using these two adjustments, the design weight is:

DESIGN_WT = W1 × ADULTS/PHONES.

Cell Phone Weight

The cell phone sample were selected in two phases (double sampling for stratification). The first phase sample is a selection of n_C^* cell phone numbers from N_C numbers on the frame. The n_C^* numbers are matched to a database to check for phone activity and to obtain the block group associated with the cell phone number. The numbers are then classified into the aggregate strata, plus a stratum for numbers that could not successfully be matched to geography based on billing information. In the second phase sample, the cell phone numbers are subsampled with "match-in" numbers oversampled relative to the "unmatched" numbers. The two phase sample selection probability for matching stratum *s* is calculated as:

$$\Pr_B(C)=(n_C^*/N_C)\times(n_{Cs}/n_{Cs}^*)$$

The cell phone sample for the Asian oversample is selected from a frame of cell phone numbers with a billing location in one of the block groups identified for the Asian oversample or selected in the Asian flagged sample, as $\Pr_A(C)=n_{CA}/N_{CA}$. The Baytown and Pasadena oversamples are adjusted for through the probability of being selected in the town oversample, $\Pr_T(C) = (n_{CT}/N_{CT})$.
The joint probability of selection for the cell sample is:

$$\begin{aligned}\Pr(C) &= \Pr_B(C) + \Pr_A(C) + \Pr_T(C)\\ &- \Pr_B(C) * \Pr_A(C) - \Pr_B(C) * \Pr_T(C) - \Pr_A(C) * \Pr_T(C)\\ &+ \Pr_B(C)*\Pr_A(C)*\Pr_T(C).\end{aligned}$$

This probability represents the probability that a cell phone number is selected in the base sample or either of the Asian oversamples. For cell phone numbers that are not eligible for the Asian oversample, $\Pr_A(C) = 0$. For those who are not on the Pasadena or Baytown frame, $\Pr_T(C) = 0$. For those not on either oversampling frame, the cell phone probability is equal to the probability of selection for the base sample. The sampling weight for the cell phone sample is the inverse of the selection probability, W1 = 1/ Pr(C). There are no household adjustments required for the cell phone sample so the design weight is equal to the sampling weight, DESIGN_WT = W1.

Frame Integration

The sample design is a fully overlapping landline and cell phone dual frame, meaning those who have a landline and cell phone are eligible to be selected via either sample. To account for the overlap of dual-users selected in the cell sample and the dual-users selected in the landline sample, we use a composite weight.

First, the design weighted total of dual-users from the landline sample and the design weighted total of dual-users from cell phone sample are averaged based on a composite weight designed to optimize the variances of weighted estimates. The composite weight is a ratio of the effective sample sizes, c = $neff_1$/

($neff_1$+ $neff_2$), where $neff = n/deff$ is the effective sample size; $deff = n \times \sum w_i^2 \times \left[\sum w_i\right]^{-2}$ is a measure of variability of respondent level weights (w_i) and n is the sample size for the survey. The landline design weight is multiplied by c, where 0 < c < 1 and the cell phone design weight by 1-c. Before averaging the landline and cell samples, we adjust each individually to match the estimated number of cell-only and landline population based on the estimated cell-only percentage (46%) from Marketing Systems Group (MSG). The MSG cell-only estimate is calculated by subtracting the estimated landline households from the estimated telephone households. Table 9 includes the cell-only and dual-user percentages from the cell phone and the number of landline completes for the entire fielding period.

TABLE 9: Cell-only and Landline Completed Interviews by Sample Type

	Landline Sample	Cell sample	Total (%)	Population Estimate
Cell-only	0	3,163	3,163 (55.5%)	46%
Has landline	1,455	1,076	2,531 (44.5%)	54%

Child Design Weights

Since children are selected via an adult in the household, the child design weights are calculated from the adult dual-frame adjusted weight. However, since there is only one randomly selected child from the household, we divide the weights based on the number of children in the household. This represents the within household probability of selecting a child. Further, since the child could be selected through any adult in the household, we remove the within household adult selection in the landline sample. The child dual-frame weight is calculated as:

DUAL_WT_CHILD = DUAL_WT_ADULT/ADULTS*CHILDREN.

Raking Ratio Adjustment

Finally, we used an iterative ratio adjustment, called raking, to adjust for nonresponse and noncoverage (of the non-telephone population). This process aligned the weighted survey sample with benchmark demographic distributions for the target population. The targets were based on the age, gender, race/ethnicity, marital status, and educational attainment for each of the aggregates as well as the entire city. These targets were obtained from the most currently available data from the American Community Survey conducted by the U.S. Census (currently 2011-2015 5 year estimates). The child and adult surveys are weighted separately. The weighting targets are:

Adults

1) Age (18-24; 25-34; 35-44; 45-54; 55-64; 65-74; 75+) by gender
2) Race/ethnicity (Hispanic; non-Hisp white; non-Hisp black; non-Hisp Asian, non-Hisp other/multi)
3) Gender by race/ethnicity (Hispanic; non-Hisp white; non-Hisp black; non-Hisp other/multi)
4) Education (Less than high school; HS grad; some college; Bachelor's degree)
5) Gender by Marital Status (Married; widowed, divorced or separated; never married)
6) Aggregate by housing tenure (Own or rent)
7) Aggregate by race/ethnicity (Hispanic; non-Hisp black; non-Hisp white/other/multi)
8) Aggregate by age (18-34; 35-54; 55+)
9) Aggregate by gender

Children

1) Age (0-4; 5-9; 10-14; 15-17) by gender

2) Race/ethnicity (Hispanic; non-Hisp white; non-Hisp black; non-Hisp other/multi)
3) Gender by race/ethnicity (Hispanic; non-Hisp black; non-Hisp white/other/multi)
4) Aggregate by race/ethnicity (Hispanic; non-Hisp black; non-Hisp white/other/multi)
5) Aggregate by housing tenure (Own or rent)
6) Aggregate by age (18-34; 35-54; 55+) by gender

Weight Trimming

The weight trimming were integrated with the raking process using Izrael et al.'s (2009) rake and trim algorithm.[8] Weights were trimmed using the global high cap value (GHCV) method. That method reduces large weights and increases small weights when they exceed the global lower or upper bounds (on the basis of factors of the average weight). The weights are constrained from increasing or decreasing beyond the individual lower or upper bounds. For example, a weight cannot increase more than four times its input weight.

[8] Izrael, D., Battaglia, M. P., & Frankel, M. R. (2009). Extreme survey weight adjustment as a component of sample balancing (a.k.a. Raking). SAS Global Forum.

IMPUTATIONS

Missing values are imputed for all weighting variables as well as variables required for calculating poverty status. We use four different imputation strategies as described in Table 10. The imputation is performed within imputation classes (e.g. males and females imputed separately). The imputation is sequential, meaning each step is based on the results of the following steps.

TABLE 10: Imputation Plan

Data Item	Imputation Method	Imputation Classes
Age	Mean—assign the mean value to all missing cases	Call type (landline or cell), gender, race <u>or</u> Call type, gender (if race missing)
Race/ethnicity	Mode—assign the modal value to all missing cases. If race is missing for child, assign race of adult. When the adult race is missing, impute with the child race (if available). These will be coded as iracecell4=2.	Call type, aggregate, gender
Educational attainment, marital status, tenure number of adults, number of children	Random Hot deck-- missing values are substituted from a respondent with non-missing values (limited to a single donor.) These values are simultaneously imputed to preserve variable relationships.	Call type, gender, race, age group
Income	Regression-- a model predicting income based on other variables in the survey	Modeled separately by call type Variables: number of children, marital status, race/ethnicity, education, tenure, employment, age group , gender and aggregate

The motivation for imputation of the HHS2017-18 is to create a public-use dataset for general users with as little as possible missing values for any possible future analysis. Missingness (i.e., nonresponse) in a survey, such as the HHS2017-18, can be categorized as follows:

(1) Unit nonresponse: cases sampled for the survey but not participating in an interview, such as noncontacts and refusers. The weighting procedure applied by the vendor takes into account this non-response.
(2) Survey block nonresponse ("Not In Survey"): set of questions that were not asked to some participants by design such as asking about hurricane Harvey related damage, which could only be asked to the part of the survey sample collected after Harvey.
(3) Item nonresponse: This includes questions not asked as a result of skip patterns ("Not In Universe"), or questions related to uncertainty ("don't know", "DK") or unwillingness (responses "refused", or "REF") or were not responded due to either participant's drop-out from the survey ("SPND", for "suspended"). Missing for other reasons ("MOR") are possible in some cases, for instance, for derived or composite variables (e.g., body mass index, which depends on the values of height and weight).

Most variables had a small amount of missing data (e.g. <5%).

<u>Reference</u>

He Y, Zaslavsky AM, Landrun MB, Harrington DP, Catalano P. Multiple imputation in a large-scale complex survey: a practical guide. Stat Methods Med Res 2010;19(6):653-70.

Implementation of Imputation

A series of a priori imputations were performed before the multiple imputation described in this document was implemented. Previous imputations were related to key variables (e.g., gender, age, income) needed to create the sampling weights. These imputations were implemented and documented by the survey vendor. Also, the study team imputed observations across the dataset based on logical relationships with other responses or by re-coding responses in the “other” option in some questions.

Imputation was only performed on missingness related to “don’t know”, “refused” or “missing due to other reasons”. Nonresponse missing values related to participation drop-out were not imputed since these group of participants did not have a chance to be asked those questions and the nonresponse is likely not missing at random. Most drop-out seemed to be related to lack of time to complete the survey, which cast doubts about missing-at-random assumptions for the imputations.

Further, when skip patterns were present, questions with missing values were imputed following the appropriately chain of questions as indicated in the questionnaire. That is, missing responses within a skip pattern branch were conditionally imputed based on the skip pattern questions. We created 25 imputations for each missing value and then selected the median value out of the 25 predicted values.

Specifying Imputation Models

Selection of predictors for the imputation

Predictive variables

To carry out MI, the inclusion of several variables is often recommended in an attempt to keep the assumption of missing-at-random (i.e., missing value (y) depends on x, but not y) plausible. Still, some literature exists indicating that MI performs well even under missing-not-at-random conditions (i.e., probability of a missing value depends on the variable that is missing such as respondents with high income are less likely to report income) and estimates are unbiased. This is possibly due to another measured variable (e.g., education) that indirectly predicts the probability of missingness in that variable (e.g., the number of years of education is associated).

We aimed to included variables that may be associated with the variable to be imputed (there’s no agreement on the recommendations about the size of the correlation, with some authors suggesting values >0.4 and other much lower ones) and that may predict the missingness pattern. These variables may not be of relevance for any analytic model, but adding them to the imputation model helps to make the assumption of MAR plausible. Given that for the HHS2017/18 we are not building a specific analytical model (e.g., a model containing a dependent variable and one or more independent variables) but we are aiming to obtain a dataset for all possible analyses on this given dataset, we were generous in the selected variables to consider for the imputation models. Further, due to the large number of variables included in the HHS2017-18 dataset, we decided to use the same variables for all the predictions (note the vendor used a similar approach for the imputation they did).

The challenge is then to select a number of variables that would help to minimize potential bias due to non-MAR and increase the precision of the imputation without introducing an excessive number of variables in the model. Not all the “good” predictors of missingness may be included in the dataset (i.e., since they have not been asked). But even if it was possible, the statistical models might be overfitted and, likely, would run into convergence problems and generate unreliable imputed values. On the other hand, if only a very small number of auxiliary variables were to be selected, the models might be underspecified. Briefly, while "some" variables may help, either "too many" or "too few" may be harmful.

Ideally, imputation models would be tailored according to an intended specific analysis. Given the general purpose of our imputation (i.e., to create a public data file which is as much complete as possible), we took the following approach. A set of sociodemographic variables were always included in all the imputation models to account for the potential variability one would expect based on age (grouped as the vendor did), gender, general socio-economic indicators (i.e., education, using the same grouping done by the vendor in their imputations) as well as area and the type of sample source.

Weights were also included. When dealing with missing data in complex survey datasets such as the HHS2017-18, the consensus in the literature leans towards recommending that including the survey weights as covariate in the imputation models would improve the multiple imputation results. Continuous weights would need to be grouped (e.g., quintiles) and modelled as a nominal variable. Missing values for adults and children as variables as well as for variables only in the pre-Harvey sample or only in the post-Harvey sample were imputed using the corresponding weights.

None of the selected predictive variables had missing data. The possibility exists that the introduction of variables in the predicted models may generate bias, although it may only be of importance in the case when correlations between the predictive variables and other variables in the model are high (that was our case). Otherwise, their potential for bias is small.

References

Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Method 2002;7(2):147-77.

Faris PD, Ghali WA, Brant R, Norris CM, Galbraith PD, Knudtson ML; APPROACH (Alberta Provincial Program for Outcome Assessment in Coronary Heart Disease) Investigators. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. J Clin Epidemiol 2002;55(2):184-91.

Hardt J, Herke M, Leohart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Medical Research Methodology 2012; 12:184.

Thoemmes F, Rose N. A cautious note on auxiliary variables that can increase bias in missing data problems. Multivariate Behavioral Research 2004; 49:443-459.

Dealing with perfect prediction

Perfect prediction may occur during estimation of categorical data, particularly when some of the categories are of small size. To minimize the occurrence of this problem, an "augmentation" strategy, which adds a few extra observations with small weights, is commonly recommended in the literature. Thus, in the predictive models of categorical variables we used the augmentation option available in Stata.

Reference

White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. Computational Statistics and Data Analysis 2010; 54: 2267-2275.

Derived variables

Derived variables were imputed directly as if they were additional variables, not re-constructed from imputed values. For instance, we imputed BMI directly rather than imputing weight and height and then re-compute BMI based on these imputed values, which increases the possibility of introducing bias.

<u>*Reference*</u>

White IR, Royston P, Wood A. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine 2011; 377-399.

Imputation modelling according to type of variable

Modelling strategies varied according to the type of variable under consideration. The variables, their categorization, and other relevant notes regarding recodification of values are detailed in the accompanying Excel sheet.

Continuous variables

For these variables we used Predictive Mean Matching (PMM) instead of linear regression since these variables were highly skewed, hence, not normally distributed. PMM is an alternative tor linear regression that relaxes the assumptions on normality of distribution and linearity for continuous variables. PMM identifies "neighbor values" who have complete data and have predicted values of the variable of interest close to the predicted value for the missing observation. One of these neighbors is randomly chosen as a "donor", and the donor's observed value on the variable replaces the recipient's missing value.

Count outcomes

Variables that reflect the number of times an event has occurred (e.g., times per week eating some kind of food, or number of years living in one place), in whole numbers equal or greater than zero but not expressed in grouped categories (e.g., 1-3, 4-7, etc., which would be consider an ordinal variable), are considered count variables. These type of variables were modeled using Poisson. In few cases (e.g., number of land lines), the small cell size in some of the categories created convergence problems. To solve this issue, the categories were grouped and, sometimes, simply two categories were created. If so, these variables were imputed via logistic regression models. In addition, in some very few instances, the imputation produced a few values outside of the expected range of the variable (e.g., 8 days when the question was about how many days something happening in the last week). In these cases, we reassigned the out-of-range value the maximum value of the expected range (e.g., 8 days in the last 7 days was coded as 7 days).

Ordinal and multinomial variables

These are variables with three or more response categories that are ordered and the most appropriate model to model them would be an ordinal model (i.e., ordinal logistic regression). A key assumption in ordinal regression is the proportional odds or parallel regression assumption, which assumes that relationship between the first response category and all the subsequent categories is the same than for the next category versus the subsequent categories, etc., and so one set of coefficients will be enough to describe the relationship. If the assumption is violated, different sets of coefficients will be needed to describe the relationship between categories. Whether or not the assumption holds should be tested for each and all the variables included as predictors in the regression model. However, the assumption will, in practice, often be violated by at least one variable. Assumption violations may or not be substantively trivial and while there are alternative ordinal models that relax the proportionally assumption (i.e., stereotype logistic regression or generalized ordered logit regression), these are not implemented in the multiple imputation routine in Stata. Therefore, to reduce potential estimation problems during the imputations, we decided to treat all these ordinal variables as nominal using polytomous logistic regression, assuming no intrinsic order among the response categories. The only drawback, if the proportionality assumption holds, and the imputed variable was truly ordinal, our models would be less parsimonious than if ordinal models were applied.

Reference

Long JS, Freese J. Regression Models for Categorical Dependent Variables Using Stata. 2nd Edition, Stata Press, Texas, 2006.

Dichotomous variables

Prior to model-based imputation we coded all dichotomous variables as 0/1 (many were coded 1/2 in the original dataset). These are the most frequent type of variables in the dataset, and were modelled via logistic regression.

Assessment of Imputation

The results of the imputation for all variables in the HHS2017-18 dataset was assessed by examining differences between the distribution of the imputed values and the originally nonmissing values. We did not observe any consistent pattern (e.g., increase number of outliers) to arise any suspicion on the success of the imputation. Given that the amount of missing data in most variables was small, hardly any imputation will lead to severe bias.

VARIABLE CODING

Geo-coding

ICF staff assigned geocodes in the form of latitude and longitude coordinates for all records based on respondents' self-reported street name and nearest cross-street. These geocodes were determined using ArcGIS software. The project management team used their best judgement on whether an open-end should actually have been originally coded as a close-ended option by taking the following steps:

- Read/review the question and answer options.
- Read/review the open-ended response.
- Determine if the open-ended response should actually have been coded as one of the available close-ended responses.
  - If yes, insert the back code
  - If no, leave as is.
- The team met to discuss some of the more ambiguous cases and to ensure consistent coding.

Employment Coding

We assigned North American Industry Classification System (NAICS) and Standard Occupational Classification (SOC) codes for each respondent who reported that they were working at a job or business in the previous week, or had been taking temporary time off from their job. Respondents reported their occupation (EMP_5) and their usual activities or duties at that job (EMP_6). Both the NAICS and the SOC use a schema whereby specific codes are nested hierarchically under more general headings. As described in the technical proposal, we coded the industry and occupation to the third digit. For NAICS codes, the first two digits indicate industry, and the third digit signifies the subsector. For SOC codes, the first two digits signify the major occupation group, and the third digit signifies the minor group. NAICS and SOC codes were assigned in almost all cases. In absence of additional information, the industry code matches the job function itself as closely as possible. We were consistent across the spectrum in all generic cases – in the absence of specificity, we assigned a "catch all" code based on the information at hand. Below are some rules/caveats we used during coding:

- Healthcare profession: Assigned NAICS code 621 if no additional detail available
- Education profession: Assigned SOC code code 25-2 if no additional detail available
- Retail profession: Assigned NAICS codes 44-45 if no additional detail available
- Manufacturing profession: Assigned NAICS codes 31-33 if no additional detail available

Key members of the project management team employed and led a team of coders in performing this task. After a training session which laid out coding rules and processes, coders were assigned a number of records to work on independently. Coders checked in frequently with each other and the lead coder to discuss ambiguities or clarify coding rules. Once independent coding was completed, we conducted three layers of quality control:

- The lead coder reviewed every record and resolved any inconsistencies between the different coders
- The lead coder and project director met to discuss how best to code general categories such as "manager" and "administration."

- A separate coder performed spot checks and developed additional rules for particularly “vague” occupations such as analyst, provider, caregiver, etc.

Records were left uncoded if:

- They were unemployed, retired, students or homemakers.
- The description provided was completely unclear.

Table 11 below shows the NAICS and SOC coding assignments for a random group of records all working within the Healthcare field.

TABLE 11: Examples of NAICS and SOC Coding Assignments

EMP_5 - Occupation	EMP_6 - Duties	NAICS Code Assigned	SOC Code Assigned
HEALTH CARE ASSISTANT	COVER THE NEEDS OF THE PATIENTS	621 - Ambulatory Health Care Services	31-9 - Medical Assistants
HEALTH CARE CONSULTANT	MAKING SURE ALL OF THE THINGS ASKED GET DONE FOR CUSTOMERS.	621	13-1 - Business Operations Specialists
HEALTH CARE PROVIDER	PREPARE MEALS FOR PATIENTS, ASSISTING WITH THEIR HYGIENE AND TAKE THEM TO WALK AND ETC.	621	31-1 - Nursing Assistant
HEALTH CARE PROVIDER	I TAKE CARE OF MY MOTHER I COOK CLEAN AND TAKE HER TO DOCTOR NURSE	621	31-1 - Home Health Aide
HEALTH CARE QUALITY MANAGER	WELLNESS EDUCATION	621	11-9- Medical and Health Services Managers
HEALTH CARE SOCIAL WORKER	COUNSELING ON PROCEDURES AND MAKING APPOINTMENTS	621	21-1- Social Workers
HEALTH CARE WORKER	ADMINISTRATIVE/CUSTOMER SERVICE	621	43-4 Customer Service Representatives Miscellaneous Information and Record Clerks
HEALTH EDUCATOR	PROMOTE HEALTH AND TEACH ABOUT BREAST FEEDING	611 Educational Services	25-3- Miscellaneous Teachers and Instructors
HEALTH INSPECTOR FOR THE CITY OF HOUSTON	GO INSPECT FOOD PLACES	923 Administration of Public Health Programs	45-2 - Agricultural Inspectors

PROJECT TIMELINE

Table 12 below is a consolidated Master Timeline of the actual project tasks and subtasks.

TABLE 12: Project Timeline

Task	Start	End
English Programming & Testing	**1/9/2017**	**3/8/2017**
Programming (Cell and Landline)	1/9/2017	1/27/2017
Internal Testing	1/30/2017	2/13/2017
Demo Due to UT		2/14/2017
External Testing	2/14/2017	2/22/2017
Update Survey Instrument based on UT feedback	2/22/2017	3/8/2017
English Survey Finalized		3/8/2017
Spanish Programming & Testing	**1/5/2017**	**3/9/2017**
Survey Translation (UT)	1/5/2017	1/20/2017
Spanish Survey Programming	1/23/2017	2/1/2017
Internal Spanish Testing	2/3/2017	2/17/2017
Spanish Demo Due to UT		2/21/2017
External QC of Spanish Materials	2/22/2017	3/1/2017
Update Spanish Survey Instrument based on UT feedback	3/2/2017	3/9/2017
Spanish Survey Finalized		3/9/2017
Pretest & Training Activities	**3/14/2017**	**6/5/2017**
Prepare Training Manuals		3/14/2017
1st Interviewer Training (Pretest)		3/15/2017
Pretest #1	3/16/2017	3/30/2017
Pretest Datafile sent to UT		4/4/2017
Pretest Report sent to UT		4/11/2017
Pretest Report Reviewed by UT	4/11/2017	4/18/2017
Team call to discuss changes		4/17/2017
Pretest programming revisions	4/19/2017	5/31/2017
2nd Interviewer Training (Full-fielding)		5/26/2017
Pretest #2	6/1/2017	6/5/2017
Full Scale Operations	**6/8/2017**	**8/27/2017 (pause)**

Sample Load #1		6/8/2017
Begin Dialing		6/8/2017
Sample Load #2		7/11/17
Sample Load #3		8/9/2017
Raw dataset delivered		8/16/2017
Paused Data Collection following Hurricane Harvey landfall		***8/27/2018***
Interim cleaned SPSS data file delivered to UT		10/5/2017
Post-Harvey Full Fielding	**2/7/2018**	**5/6/2018**
Programming	1/30/2018	2/1/2018
Testing	2/1/2018	2/7/2018
Interviewer Re-training		2/6/2018
Sample Load #4		2/7/2018
Begin Dialing		2/7/2018
Sample Load #5a		3/22/2018
Sample Load #5b – Asian-flagged sample		
Finish Data Collection		5/6/2018
Data Processing and Reports	**5/7/2018**	**7/18/2018**
Open-end cleaning and coding	3/1/2018	5/21/2018
Data Processing	5/7/2018	5/30/2018
Raw dataset and codebook sent to UT		5/31/2018
Draft Implementation Plan sent to UT		6/1/2018
Weighting and geo-coding	6/1/2018	6/27/2018
Imputed Data File Deliverable to UT Health		7/11/2018
Draft #2 of Implementation Plan sent to UT		7/18/2018
Discussions regarding updates to datafile	7/19/2018	9/7/2018
Delivery of revised dataset to UT		9/11/2018
Delivery of revised Implementation Plan to UT		9/21/2018