Westat

Improving Lives Through Research®

Study of the Feasibility Survey for Redesign of the National Intimate Partner and Sexual Violence Survey

July 21, 2021

Authors
David Cantor   Mina Muller
Ting Yan   Darby Steiger
Eric Jodts   Stephanie Mendoza
J. Michael Brick

Submitted to:
Centers for Disease Control and Prevention
Atlanta, GA

Submitted by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

Table of Contents (continued)

List of Tables

Table

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

Acronyms

AAPOR	American Association for Public Opinion Research
ABS	address-based sampling
ACASI	audio computer-assisted interviewing
ACS	American Community Survey
CATI	computer-assisted telephone interviewing
IPPA	intimate partner psychological aggression
IPPV	intimate partner physical violence
IPV	intimate partner violence
NCVS	National Crime Victimization Survey
NHIS	National Health Interview Survey
NISVS	National Intimate Partner and Sexual Violence Survey
NRB	nonresponse bias
NRFU	nonresponse follow-up
NSFG	National Survey of Family Growth
PTSD	post-traumatic stress disorder
RBP	Rizzo-Brick-Park probability method
RDC	Research Data Center
RDD	random digit dial
YMOF	Youngest Male Oldest Female

1. Introduction

This report summarizes findings from the Feasibility Study testing two alternative designs for the National Intimate Partner and Sexual Violence Survey (NISVS). One design uses random digit dial (RDD) as a sample frame and computer-assisted telephone interviews (CATI) as the mode of interviewing. The second design uses an address-based sample (ABS) that pushes respondents to the web and follows up with multimode alternatives. One alternative gave respondents a choice between the web and calling in to do a CATI. The second gave the choice between web and filling out an abbreviated paper survey. The goal of this report is to summarize results that are key for deciding the design for the next NISVS.

Chapter 2 provides an overview of the methodology used in the data collection, weighting, and analysis. Chapters 3-7 summarize results addressing each research question. The final chapter makes a recommendation on for the next NISVS.

Research Questions and What Is Covered in This Report

The analysis of the Feasibility Study is organized around six research questions:

1. How do response rates compare between the RDD and ABS surveys?

2. How do prevalence estimates compare between ABS and RDD surveys?

3. How do ABS and RDD compare on bias due to nonresponse?

4. How do ABS and RDD compare on other measures of data quality (item missing data) as well as measures of other outcomes (comparison to other surveys measuring sexual violence, relationship of victimization with age; measures of consequences of victimization)?

5. How do ABS and RDD compare on respondent burden, privacy, and confidentiality?

6. What are results of the experiments for ABS and RDD?

These cover the critical issues needed to decide on whether RDD, ABS, or a combination of these frames should be part of the NISVS design moving forward.

The results of the analyses (Chapters 3–8) begin with a summary of the major findings for each research question. This is followed by more specific descriptions of the results in the form of highlights. The tables are included at the end of each chapter.

2. Methodology

The Feasibility Study was designed to compare random digit dial (RDD) and address-based sample (ABS) methodologies for the National Intimate Partner and Sexual Violence Survey (NISVS). This chapter provides a brief overview of the methodology used for the study.

Methods for the Random Digit Dial Survey

The RDD portion of the Feasibility Study is a dual-frame national telephone survey with approximately 77 percent of the sample being cellphone numbers and 23 percent landline. The cellphone frame comprises the majority of the sample because of its superior coverage, as well as the tendency for young adults and males to be better represented than on a landline. The cell numbers were prescreened to take out businesses and numbers identified as not likely to be active. For the landline frame, the numbers were matched to an address list. For those that match to an address, a letter was sent alerting the household about the survey. The letter contained a $2 bill to encourage the respondent to read the letter. Respondents were offered an incentive to complete the survey, with the amount depending on the stage of the survey. One experiment was conducted that tested whether sending out a text prior to calling a cellphone had an effect on the response rate.

For the cellphone component, the person who answered the telephone was considered the eligible respondent, as long as that person was 18 years old or over. For the landline survey, a screening interview was conducted to randomly select an adult in the household using the Rizzo-Brick-Park method.[1]

Calling for the RDD survey began on May 4, 2020 and ended on July 12, 2020. There were two phases to the survey. The first phase occurred from May 4 to June 21, 2020 and consisted of calling and following up all eligible telephone numbers. Follow-up calls were placed when no one answered the phone or the person eligible for the interview was not available. Refusal conversion was attempted for those that refused but did not express hostility. Those who completed the survey were paid $10. The second phase, or nonresponse follow-up (NRFU), began June 29 and ended on July 12, 2020. For the NRFU, the nonrespondents who had not displayed hostility were sampled and called back. A $40 incentive was offered to these individuals to complete the survey.

Once the survey was completed, the data were weighted[2] by first computing the probabilities of selection. After that, an adjustment was made for nonresponse within the landline and cellphone frames. After this was done, the two frames were combined using information collected on the survey about those who could have been selected in both of the sample frames. The final weights were created by raking to national totals from the American Community Survey for sex; age; race-ethnicity; marital status; education; and telephone status (cell, landline, both).

---

[1] Rizzo, L., J. Michael Brick, and Inho Park. (2004). A minimally intrusive method for sampling persons in random digit dial surveys. *Public Opinion Quarterly*, 6, 267-274.

[2] Westat. (2020). National Intimate Partner and Sexual Violence Survey (NISVS) Redesign – Weighting Plan. Rockville, MD: Author. Delivered to CDC on November 21, 2020. See the appendix.

Methods for the Address-Based Sample

The sampling frame for the ABS was drawn from a database of addresses provided by the sample vendor MSG. The MSG database is derived from the Computerized Delivery Sequence File (CDSF), which is a list of addresses from the United States Postal Service (USPS). The CDSF is estimated to cover approximately 98 percent of all households in the country. All nonvacant residential addresses in the United States present on the MSG database, including post office (P.O.) boxes; throwbacks (i.e., street addresses for which mail is redirected by the USPS to a specified P.O. box); and seasonal addresses were considered eligible.

Data collection began on May 4 and ended October 6, 2020. As with the RDD, data collection was conducted in two phases. Phase 1 began on May 4 and NRFU began on August 17, 2020. Unlike RDD, data collection for the two phases continued concurrently and both ended on October 6, 2020.

The procedures for the ABS frame are shown in Figure 2-1. The first step was to send a letter asking an adult to complete the screening survey on the web (Figure 2-1 - Box A). The letter contained a monetary incentive of $5 cash and a promised $10 Amazon gift code upon completing the web screener. Letters included a unique PIN for each household and the URL to launch the survey. The letter also included a helpdesk toll-free number for any questions about the study. The web screener included questions about the household needed to select an individual to be the respondent for the NISVS survey. If the person selected for the extended interview was the screener respondent, then that person was instructed to proceed directly to the extended interview. If the screener respondent was not selected for the extended survey, the screener respondent was instructed to ask the selected adult to log in to the website and complete the survey.

A reminder postcard was sent approximately one week after the first mailing. If the screener was not completed, a letter was sent by express delivery asking the respondent to either fill out the screener on the web or complete a paper version of the screener (Figure 2-1 – Box C). If the web screener was still not completed and the paper screener was not returned after this mailing, the household was considered for subsampling for the nonresponse follow-up (Figure 2-1 – Box G).

Additional follow-up contacts were attempted for those households that completed the screener, but the selected respondent did not complete the extended interview. This included two groups. One group was those who completed the web screener, but there was no response for the extended survey. The second group was those who returned the paper screener. For this second group, Westat home office staff selected a respondent from the household roster (Figure 2-1 – Box F). In the case of both of these groups a follow-up letter was sent inviting the selected individual to complete the survey on the web (Figure 2-1 – Box D). The person was promised an incentive of $15 to complete the survey. If there was no response, a letter was sent express delivery to the selected respondent (Figure 2-1 – Box E), asking the person to complete the extended survey by web or by an alternative mode. Half of these were given the choice between the web and a paper version of the questionnaire. The other half were given the choice between the web or to call in and complete the survey over the phone.

To provide the incentives for those who completed the survey on the web, an Amazon gift code was provided at the end of the survey. If respondents completed the survey by paper or computer-assisted telephone interviewing (CATI), a letter thanking them for completing the survey and the cash incentive was mailed by USPS First Class.

Once all attempts were exhausted, a subsample of the nonrespondents was selected. Contacts using the same procedures as described above were attempted, but offered a larger incentive ($40 for completing the web or by CATI; $30 to do it by paper).

Fifty percent of the screener nonrespondents were subsampled for NRFU. Those who were selected received a letter encouraging them to go to the web to complete the screener or complete the enclosed paper screener. The mailing also included a $5 cash prepaid incentive, but there were no promised incentives for completing the screener whether by web or by paper.

Additionally, 50 percent of the extended survey non-respondents who completed the phase 1 protocol were selected for NRFU[3]. This group was split into two mode options: web/paper or web/CATI. They were offered $40 for completing the extended survey on the web or on the phone and offered $30 for completing it by paper.

The incentives provided at each stage for each frame are provided in tables 2-1 and 2-2.

Table 2-1. Incentives for the screener by frame and NRFU status		
	ABS	**RDD**
With first invitation letter	$5 Cash	$2* Cash
Complete by web	$10 AC	NA
Complete by paper	$5 CM	NA
NRFU with Invitation Letter	$5 Cash	NA

NA – Not aplicable
NRFU - Nonresponse follow-up.
AC - Amazon code available online after completing survey.
CM - Cash sent by mail after completing survey.
* Sent to landline telephone numbers where an address was found.

Table 2-2. Incentives for the extended interview by frame, mode and NRFU status				
	ABS			**RDD**
	Web	**CATI**	**Paper**	**CATI**
Complete	$15 AC	$15 CM	$5 CM	$10 CM
NRFU - Complete	$40 AC	$40 CM	$30 CM	$40 CM

NRFU - Nonresponse follow-up.
AC - Amazon code available online after completing survey.
CM - Cash sent by mail after completing survey.

---

[3] Paper screener returns from the initial screener nonresponse mailing were not included in phase 2 subsampling since they had not fully completed the extended phase 1 protocol at the time of phase 2 subsampling.

Figure 2-1.   Mailing sequence for ABS frame

A — Mail letter to household requesting to go to web to fill out screener

Is web screener complete?
- Yes → Is extended interview completed?
- No → B — Reminder Postcard

Is extended interview completed?
- Yes → Complete
- No → D — Follow-up letter to selected respondent to complete on the web

B — Reminder Postcard → Is web screener complete?
- Yes → Is extended interview completed?
- No → C — Express delivery of letter requesting to go to the web. Also included is a paper survey asking for household roster

Is extended interview completed? (F)
- Yes → Complete
- No → Select respondent at home office – Probability Method Only.

C — Express delivery of letter requesting to go to the web. Also included is a paper survey asking for household roster → Is web screener complete?
- Yes → Is extended interview completed?
- No → Is paper screener returned?

Is paper screener returned?
- Yes → Select respondent at home office – Probability Method Only. (F)
- No → Nonresponse

D — Follow-up letter to selected respondent to complete on the web → Is extended interview completed on the web?
- Yes → Complete
- No → E — Express Delivery: Send letter with request to go to web; include alternative mode

E — Express Delivery: Send letter with request to go to web; include alternative mode → Is extended interview completed?
- Yes → Complete
- No → Nonresponse

G — Sampled for Nonresponse Followup

Nonresponse → Sampled for Nonresponse Followup ← Nonresponse

The weights for the ABS were first computed by calculating the probability that a respondent was selected into the sample. The weights were then adjusted for nonresponse, with a final raking adjustment using the same characteristics as those for the RDD except for telephone status (sex, age, race-ethnicity, marital status, and education).[4]

Variance Estimation and Significance Test

To compute standard errors a stratified jackknife (JK2) variance estimator[5] was used. This involves creating a set of replicate weights created by applying the same adjustments made to the full sample weight. The JK2 method has good statistical properties. The JK2 method with 100 replicates is used for the analysis.

For purposes of the discussion in the report, we have selectively used the standard errors to test for significant differences. These tests were carried out using a two-sample z-test. The discussion generally makes the distinction between differences that are statistically significant at least at the 5 percent level (two–tailed test) and those that are different but do not reach statistical significance.

---

[4] See the Appendix for details.

[5] Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3), 283-310.

3. What Are Response Rates for RDD and ABS Surveys?

Summary

➤ Different definitions of a complete and partially completed survey were used for the different survey modes. For the computer-assisted telephone interviewing (CATI) and paper surveys, the definition from prior National Intimate and Sexual Violence Survey (NISVS) surveys was used. For the web survey a different definition was used to account for the larger number of respondents dropping out and skipping questions.

➤ The ABS response rate was approximately three times the rate for that of RDD.

➤ The ABS response rate is about the same as the NISVS response rates during the 2010 to 2015 period.

➤ For the ABS sample, a large portion of the surveys were completed on the web (3,306 out of 3,526 – 93.7%). There were 187 completed by paper and 33 for the call-in CATI.

Definition of a Completed Survey

The definition of a complete and partial complete survey was different across collection modes. For surveys completed by CATI and paper, a partial complete was defined as any survey that is completed through the stalking questions but did not qualify for a full complete. The stalking section is the first violence section. A full completed interview was any survey that had responses through the consequences section, the last section asking about violent victimization. Both full and partial completes the web, CATI and paper surveys were used in the final weighting and in the analyses for this report. There were a total of 3526 for ABS and 1461 RDD completes and partial completes (Table 3-4).[6]

A different definition of a complete and partial was used for the web survey. This was in recognition that web respondents tend to drop out of the survey more often than those completing by CATI or paper. There also tends to be more item-missing data on the web. For the web, if a partial survey was defined as early as completion of the stalking section, there would be a significant amount of missing data for measures that rely on later sections of the questionnaire. Of particular concern were the sections asking about rape and made to penetrate, which come relatively late in the survey. These definitions also considered a minimum level of data quality, such as providing data for the basic measures used on NISVS.

A web survey was considered a full complete if there were responses through the consequences section and met the following criteria:

- There had to be substantive answers to at least half of the items that screen for lifetime victimization.

- The respondent had to correctly answer at least one of the two questions designed to catch those who were not paying attention. One of these items was placed at the beginning of the questionnaire, in the Health section, which asked the respondent how carefully they were reading the survey:

---

[6] This the total uweighted count, not including the subsampling completed for the NRFU.

[SAT1]   **I am reading this survey carefully.**

Yes...   1
No   2

The second item is in the last section of the survey. This instructed the respondent to select a particular response category.

[SAT2]   **We just want to see if you are still awake. Please select ''Neutral'' to this question.**

Strongly Disagree..   1
Disagree...   2
Neutral...   3
Agree..   4
Strongly Agree..   5

A partial complete were those:

- that provided answers through the rape/made to penetrate section, but did not get through the consequences section

- had substantive answers to at least half of the items that screen for lifetime victimization.

- The respondent had to correctly answer at least one of the two questions designed to catch those who were not paying attention.

There were relatively few people who failed two questions designed to check on attention. In total, there were six individuals who marked not reading the survey carefully on SAT1. There were 20 who did not pick the neutral category in SAT2. Only one respondent picked both "not reading carefully" and a non-neutral category. This one individual was taken out of the analytic dataset (n=3526, Table 3-4).

Survey Results

1. For the RDD sample, the total number of phone numbers purchased was 58,838 (Table 3-1). Approximately half of these were taken out after pre-screening. A total of 27,476 were dialed.

2. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 4 (RR4) (Tables 3-2 and 3-3). They include both the Phase 1 and the NRFU stages of the surveys. The response rate is weighted to account for the sampling at the NRFU stage.

2. The overall ABS response rate was 33.1 percent (Table 3-3). The screener response rate was 50.3 percent and the extended response rate was 65.3 percent.

3. The RDD response rate was 10.8 percent. The landline sample had a response rate of 18.4 percent compared to the cellphone sample with a rate of 8.5 percent.

4. For RDD, there is a large difference in the response rate between landline and cell at the screener level (28.5% for landline vs. 14.9% for cellphone). A large percentage of the RDD

nonresponse is in the unknown category (UH), primarily from no one answering the phone (58%). The second biggest category are refusals (21%).

5. For the ABS, many of those who dropped out did so at the beginning of the survey (Table 3-4). There were also increases in dropouts at the sexual harassment and first section on rape.

6. For the ABS sample, a large portion of the surveys were completed on the web (3,306 out of 3,526 – 93.7%; data not shown). There were 187 completed by paper and 33 for the call-in CATI.

Table 3-1.	Disposition of sampled telephone numbers			
		Frame		
		Landline	Cell	Combined
Total numbers purchased		**29,158**	**29,680**	**58,838**
Ineligible pre-screened out		**23,015**	**8,347**	**31,362**
Nonresidential		2,014		
Nonworking		20,663		
Inactive and Unknown			8,347	
LL ported to CP		338		
Total numbers released		**6,143**	**21,333**	**27,476**
Phone numbers released in predictor and main		4,143	13,333	17,476
Phone numbers in reserve		2,000	8,000	10,000
Total dialed		**6,143**	**21,333**	**27,476**

Table 3-2.   ABS and RDD survey results by final disposition codes*

Classification	ABS Screener	ABS Extended	ABS Overall	RDD Overall Screener	RDD Overall Extended	RDD Overall Overall	RDD Landline Screener	RDD Landline Extended	RDD Landline Overall	RDD Cell Screener	RDD Cell Extended	RDD Cell Overall
I: Complete	5,765	3,630	3,630	3,014	1,606	1,606	872	500	500	2,141	1,106	1,106
P: Partial	0	131	131	0	132	132	0	42	42	0	90	90
R: Refusal	50	183	233	4,465	500	4,965	762	146	908	3,703	354	4,057
O: Other Nonresponse	1	0	1	1,227	247	1,474	164	70	234	1,063	176	1,240
NC: Noncontact	0	1,814	1,814	0	33	33	0	6	6	0	27	27
UO: Unknown Other	5,996	0	5,996	0	0	0	0	0	0	0	0	0
UH: No answer/call blocking	0	0	0	13,593	458	14,051	2,612	93	2,705	10,982	365	11,346
SO: Ineligible	755	7	762	229	4	233	3	0	3	226	4	230

- Data weighted to account for the subsampling at the nonresponse follow-up. RDD counts exclude those numbers taken out in the pre-screening.

Table 3-3. Response, cooperation, refusal, contact, and yield rates for ABS and RDD samples

| Preliminary measures | ABS | | | RDD | | | | | | | | |
| | | | | Overall | | | Landline | | | Cell | | |
	Screener	Extended	Overall	Screener	Extended	Overall	Screener	Extended	Overall	Screener	Extended	Overall
Response Rate AAPOR RR4	50.3%	65.3%	33.1%	18.1%	58.5%	10.8%	28.5%	63.3%	18.4%	14.9%	56.6%	8.5%
Cooperation Rate: (I+P)/(I+P+R+O)	99.1%	95.3%	94.1%	40.3%	77.7%	25.9%	53.4%	78.8%	37.4%	36.6%	77.2%	22.8%
Refusal Rate: R/((I+P)+ (R+NC+O)+ UH + UO)	0.4%	3.2%	2.0%	25.6%	16.8%	29.4%	24.9%	17.0%	30.7%	25.8%	16.8%	29.0%
Contact Rate: (I+P) +R+O / (I+P)+R+ O+NC+(UH + UO)	49.2%	68.5%	33.8%	50.6%	83.7%	48.8%	58.7%	88.6%	57.0%	48.2%	81.7%	46.4%
Yield Rate: I/Released Cases	45.9%	63.0%	28.9%	11.0%	53.3%	5.8%	13.8%	57.3%	7.9%	10.1%	51.6%	5.2%

For ABS:
- Response Rate 4 = (I+P)/((I+P) + (R+NC+O) + e(UH+UO)]
- e calculated using methods in DeMatteis, J. (2019). Computing "e" in self-administered addressee-based sampling studies. *Survey Practice*, 12(1). Available at: https://www.surveypractice.org/api/v1/articles/8282-computing-e-in-self-administered-address-based-sampling-studies.pdf

For RDD:
- Response Rate 4 = (I+P)/(I+P+R+NC+O)+(e1xUO)+(e1xe2xUH)
- e1is the calculated the proportion of eligibles divided by the number of eligibles and known ineligibles for the household screener. E2 is the same proportion but for the extended interview.
- Overall response rate calculated by combining the cell and landline samples in proportion to thenumbers that were dialed for each frame (23% for landline; 77% for cell).

		RDD			ABS	
Section	**Total**	**Complete or Partial**	**Not complete**	**Total**	**Complete or Partial**	**Not complete+**
Health characteristics	23	0	23	188	0	188
Stalking	27	0	27	72	0	72
Sexual harassment, unwanted touching	50	50	0	138	0	138
Completed alcohol/drug-facilitated rape and made to penetrate	24	24	0	87	0	87
Unwanted sex due to threats of harm/physical force	3	3	0	67	1	66
Attempted physically forced sex	12	12	0	26	1	25
Sexual violence: outcomes of rape and made to penetrate	0	0	0	2	2	0
Psychological aggression	4	4	0	48	48	0
Physical violence	7	7	0	32	31	1
Consequences and follow-up	1	1	0	16	14	2
Debriefing questions	5	4	0	103	76	27
Completed all applicable questions	1,356	1,356	0	3,409	3,353	56
Total	1,512	1,461	50	4,188	3,526	662

Table 3-4.   Last section completed for RDD and ABS by completion status X

[X] Unweighted counts

[+] Surveys that are not complete but dropped out after attempted forced sex are those that did not meet the other criteria related to completed survey (e.g., too much missing data; did not answer trap questions correctly).

4. How Do Prevalence Estimates Compare Between ABS and RDD Surveys?

Summary

➢ There are both statistical and substantive differences in prevalence estimates between random digit dial (RDD) and address-based sample (ABS) surveys.

➢ Lifetime prevalence estimates are much lower for ABS when compared to RDD. This holds for both males and females. The differences are somewhat smaller for intimate partner violence and intimate partner psychological aggression (IPPA).

➢ For 12-month estimates, the differences between ABS and RDD are much smaller, or nonexistent. The direction of the differences depends on the specific type of victimization.

➢ Proportionally fewer RDD lifetime victims are reporting 12-month victimization than ABS lifetime victims. There are several possible reasons for this pattern:

  - The question sequences each survey used to ask about 12-month victimizations are different. RDD asks respondents about 12-month victimization using a single question that summarizes across all perpetrators. The ABS asks about 12-month victimization for each perpetrator. The change for the ABS was made to facilitate recall and dating accuracy.

  - The 12-month questions are more sensitive than lifetime questions. Respondents are generally less willing to disclose sensitive information for events that occur in the recent past. The self-administration of the web and paper surveys may increase the willingness of respondents to report this information when compared to the RDD.

  - The difference is due to differential nonresponse. There may be proportionately more 12- month victims among ABS respondents.

➢ The relationship between sex and victimization differs between ABS and RDD, especially for the 12-month estimates. For ABS, females are generally higher estimates than males for both lifetime and 12-month estimates. For RDD, the relationship is weaker. For the ABS lifetime estimates, two types of male victimization estimates are not statistically different from females, with one estimate (IPPV) being equal to females. For 12-month estimates, males have nominally higher estimates for IPPV and IPPA. Of those where females are higher than males, several are not statistically different.

Lifetime Prevalence Estimates

1. Estimates for total population for all types of victimization are higher for RDD than ABS. All but one of the differences are statistically significant at least at the 5 percent level (Table 4-1). The one exception is for stalking by an intimate partner.

2. For the total population, most of the ABS estimates range between 58 percent to 75 percent of the RDD estimate. For example, the ratio between the ABS to RDD estimates for unwanted touching, sexual coercion, and rape are .58, .66, and .65, respectively. The ratio gets somewhat larger for victimizations involving intimate partner violence, ranging from .73 to .80.

3. The pattern is similar for the sex-specific estimates (Table 4-2). All ABS estimates are nominally lower than RDD estimates. Many of the differences are statistically significant at least at the 5 percent level.

4. For females, all but one of the comparisons between ABS and RDD are significant at least at the 5 percent level. Using the ratio of ABS to RDD, the biggest difference is for unwanted touching where the ABS estimate is about 60 percent of the RDD estimate (35.0% vs. 58.5%). The ratio of the ABS estimate to RDD for other types of Contact Sexual Violence are .65 (sexual coercion) and .62 (rape) of the RDD estimate.

5. For males, while all ABS estimates are lower than RDD, not all differences are statistically significant. Part of this may be the lower estimates for males, which require larger sample sizes to have the power to detect statistical significance.

6. For females, the difference between ABS and RDD are larger for contact sexual violence/stalking relative to intimate partner victimizations of the same type. Ratios of ABS to RDD for contact sexual violence/stalking are between .60 and .66, while it is between .74 to .85 for intimate partner victimizations. It is more mixed for males, with no clear pattern.

7. For both RDD and ABS, the estimates for females are generally higher than for males. Many differences between males and females are statistically different; most well beyond the 5 percent level (test not shown). For RDD there are two exceptions to this. The male RDD estimate for intimate partner physical violence (IPPV) is about the same as for females (33.7% vs. 33.8%). The estimate for male IPPA is less than for females, but it is not statistically different (44.4% vs. 49.6%; p>.10).

12-Month Prevalence Estimates

8. For the total population 12-month prevalence estimates, there is not a consistent difference between ABS and RDD (Table 4-3). RDD is larger for some of the estimates, but only one reaches statistical significance (unwanted touching - 4.0 vs. 2.3; p<.05). For several types of victimizations, the ABS is larger, but do not reach statistical significance. The 12-month estimates are generally less than 5 percent, which reduces the statistical power when testing for differences, relative to lifetime estimates. However, with a few exceptions, the nominal differences between the estimates are not large.

9.   The estimates for males are generally larger for the RDD relative to ABS (Table 4-4), with a few reaching, or approaching, statistical significance, including unwanted touching (3.2 vs. 1.1; p<.06) and IPPA (6.6% vs. 3.6%; p<.05).

10.   For females, none of the differences between ABS and RDD are statistically significant. The differences are not consistently in one direction.

11.   The relationship between sex and victimization differs between ABS and RDD. There is a strong relationship for ABS. All ABS estimates for females are higher than for males. All differences are statistically significant, except for IPPV (3.2% vs. 2.0%; p<.10; significance test not shown). For RDD, males have higher estimates than females for IPPV and IPPA. All the other differences for RDD are higher for females, although not all statistically significant.

Table 4-1. Lifetime prevalence estimates for selected measures of sexual, physical, and emotional abuse*

	RDD		ABS		Sig. testing p-value
	Number of victims	**Prevalence estimate**	**Number of victims**	**Prevalence estimate**	
Contact sexual violence	113,677,147	47.1	73,337,937	30.4	0.000
Unwanted touching	100,827,179	41.8	58,266,012	24.4	0.000
Sexual coercion	44,361,636	18.4	28,932,456	12.1	0.000
Rape	43,781,903	18.1	28,281,935	11.7	0.000
Made to penetrate+	n/a	n/a	n/a	n/a	n/a
Stalking	33,892,785	14.0	22,136,378	9.2	0.000
Intimate partner violence	94,033,173	38.9	69,125,448	28.6	0.000
Contact sexual violence by intimate partner	35,373,223	14.6	28,514,862	11.8	0.050
Stalking by intimate partner	16,253,714	6.7	12,417,282	5.2	0.114
Intimate partner physical violence	81,443,393	33.7	54,232,032	22.7	0.000
Intimate partner psychological aggression	113,621,584	47.0	79,080,933	33.1	0.000

n/a – Not applicable. Not asked of female respondents
* Combines estimates for males and females
+ Made to penetrate only asked of males.

Table 4-2. Lifetime prevalence estimates for selected measures of sexual, physical, and emotional abuse by sex+

	Males					Females				
	RDD		ABS			RDD		ABS		
	Number of victims	Prevalence estimate	Number of victims	Prevalence estimate	p-value	Number of victims	Prevalence estimate	Number of victims	Prevalence estimate	p-value
Contact sexual violence	36,028,444	30.3	20,215,501	17.5	0.000	77,648,703	63.3	53,122,435	42.2	0.000
Unwanted touching	29,097,986	24.5	14,723,766	12.9	0.000	71,729,193	58.5	43,542,246	35.0	0.000
Sexual coercion	10,832,375	9.1	6,609,717	5.8	0.066	33,529,261	27.3	22,322,739	17.9	0.000
Rape	5,025,974	4.2	3,638,372	3.1	0.261	38,755,929	31.6	24,643,563	19.7	0.000
Made to penetrate	8,802,790	7.4	5,831,723	5.0	0.130	n/a	n/a	n/a	n/a	n/a
Stalking	8,427,334	7.1	5,615,161	4.9	0.104	25,465,451	20.8	16,521,217	13.2	0.001
Intimate partner violence	42,182,620	35.5	23,853,762	20.6	0.000	51,850,551	42.3	45,271,686	36.0	0.032
Contact sexual violence by intimate partner	7,030,991	5.9	5,968,605	5.2	0.343	28,342,232	23.1	22,546,256	17.9	0.051
Stalking by intimate partner	4,214,824	3.5	2,493,065	2.2	0.168	12,038,890	9.8	9,924,217	7.9	0.203
Intimate partner physical violence	40,021,937	33.7	20,896,232	18.3	0.000	41,421,457	33.8	33,335,799	26.8	0.009
Intimate partner psychological aggression	52,770,530	44.4	33,525,608	29.2	0.000	60,851,053	49.6	45,555,325	36.7	0.000

n/a – Not applicable. Not asked of female respondents.
+ p-value is the result from a two-sample difference of means z-test.

Table 4-3.	12-month prevalence estimates for selected measures of sexual, physical, and emotional abuse+#				
	RDD		ABS		
	Number of victims	**Prevalence estimate**	**Number of victims**	**Prevalence estimate**	**p-value+**
Contact sexual violence	12,725,988	5.3	11,672,415	4.8	0.353
Unwanted touching	9,570,647	4.0	5,601,152	2.3	0.042
Sexual coercion	5,065,328	2.1	6,025,207	2.5	0.318
Rape	2,621,355	1.1*	2,824,323	1.2	0.391
Made to penetratex	535,783	0.5*	705,739	0.6*	0.379
Stalking	8,044,867	3.3*	7,052,414	2.9	0.347
Intimate partner violence	12,183,906	5.0	11,863,637	4.9	0.395
Contact sexual violence by intimate partner	4,332,759	1.8*	5,696,394	2.4	0.245
Stalking by intimate partner	3,272,433	1.4*	2,561,998	1.1	0.339
Intimate partner physical violence	8,153,303	3.4	6,298,749	2.6	0.256
Intimate partner psychological aggression	13,907,769	5.8	11,707,172	4.9	0.259

Estimates combine male and females
* CV>30%
+ p-value is the result from a two-sample difference of means z-test.
x Made to penetrate only asked of males.

Table 4-4.   12-month prevalence estimates for selected measures of sexual, physical, and emotional abuse by sex+

	Males					Females				
	RDD		ABS			RDD		ABS		
	Number of victims	Prevalence estimate	Number of victims	Prevalence estimate	p-value	Number of victims	Prevalence estimate	Number of victims	Prevalence estimate	p-value
Contact sexual violence	4,275,526	3.6	2,409,239	2.1	0.159	8,450,462	6.6	9,263,074	7.4	0.381
Unwanted touching	3,813,523	3.3*	1,314,381	1.1	0.058	5,757,084	4.0	4,286,734	3.4	0.246
Sexual coercion	1,151,793	1.0*	656,891	0.6*	0.315	3,913,285	3.5	5,368,316	4.3	0.251
Rape	189,379	0.2*	452,275	0.4*	0.208	2,431,975	2.0	2,372,049	1.9	0.397
Made to penetrate	535,783	0.5*	705,739	0.6*	0.375	n/a	n/a	n/a	n/a	n/a
Stalking	2,532,928	2.1	2,132,029	1.9	0.380	5,511,940	4.5	4,920,389	3.9	0.353
Intimate partner violence	5,495,404	4.6	3,265,812	2.8	0.131	6,688,503	5.5	8,598,328	6.8	0.249
Contact sexual violence by intimate partner	1,658,280	1.4	796,062	0.7	0.228	2,674,529	2.2	4,900,332	3.9	0.089
Stalking by intimate partner	798,304	0.7*	392,260	0.3*	0.318	2,474,129	2.0	2,169,738	1.7	0.379
Intimate partner physical violence	4,854,894	4.1	2,332,904	2.0	0.083	3,298,409	2.7	3,965,842	3.1	0.349
Intimate partner psychological aggression	7,828,648	6.4	4,108,231	3.6	0.035	6,079,320	5.0	7,598,941	6.1	0.265

n/a – Not applicable. Not asked of female respondents.
* CV>30%
+ p-value is the result from a two-sample difference of means z-test.

5. How Do ABS and RDD Compare with Respect to Nonresponse Bias?

Summary

➢ When comparing the demographic profile, before the weights are applied, both the ABS and RDD surveys represented the population closely for sex and income. They both underrepresented those with low education and overrepresented non-Hispanic Whites and those with an internet connection. ABS was significantly closer to the national distribution for those between 18 to 44 years old.

➢ Selected health measures from the Feasibility survey were compared to the National Health Interview Survey (NHIS). In several cases, both ABS and RDD estimates were very close to the NHIS. The RDD estimates tended to be closer to several NHIS measures, but not by a large amount. The largest differences were for the percent who had been told they were depressed (27.7% RDD, 24.2% ABS, 15.7% NHIS).

➢ Lifetime prevalence estimates were compared for the initial Phase 1 and NRFU. The patterns differed by sex and sample frame.

ABS. The most significant differences were for females in the ABS sample, which found the NRFU estimates to be higher than the Phase 1 sample. Interpreted as a measure of nonresponse bias, this indicates a negative bias (estimates are too low) for all types of victimization. There was no indication of bias for males.

RDD. For males, the Phase 1 estimates were less than that of the NRFU for selected estimates. This is indicative of a negative bias for unwanted touching, sexual coercion, rape, IPPV and IPPA. For females, the Phase 1 estimates were higher than NRFU for selected estimates. This is indicative of positive bias for contact sexual violence by intimate partner and stalking by intimate partner.

➢ The ABS and RDD lifetime estimates were compared to the 2015 and 2016-17 NISVS estimates. The lifetime estimates for the ABS are closest to 2015, while the RDD estimates are closest to 2016-17. If one assumes the 2015 estimates are less subject to nonresponse bias than 2016/2017 given the higher response rate, this indicates that the RDD estimates in the Feasibility Study are biased in the positive direction (too high)

The analyses in this chapter analyzes nonresponse bias (NRB) using four different methods: 1) Comparing demographic estimates to national benchmarks, 2) Comparing health indicators to the NHIS, 3) Comparing key NISVS outcome measures by the amount of effort to complete the survey, and 4) Comparing estimates to prior NISVS data.

Comparing Demographic and Socioeconomic Characteristics to National Benchmarks

One method to assess NRB is to compare the distribution of key demographic and socioeconomic indicators from the survey to national benchmarks before the full population weights are applied. The extent these characteristics are different than the national population is an indication of possible NRB.

Table 5-1 provides this comparison for age, sex, marital status, race-ethnicity, education, income, born in the U.S., access to internet, and homeownership. For these comparisons, the ABS and RDD surveys were weighted to reflect their probabilities of selection. For ABS this consists of weighting for selection within the household and selection for the NRFU. For RDD this adjustment had three components: 1) for the landline sample an adjustment was made for selection within the household; 2) for both frames an adjustment for selection for the NRFU; and 3) for both frames, an adjustment for the probability of selection into the landline or cellphone frame.

Many of the above demographic characteristics were used when raking to control totals when computing the final population weights. Comparisons using the above selection weights provides a profile of respondents before forcing the totals to be equal to the control totals used in the raking while still controlling for differential probabilities of selection built into the different survey designs.

1. Both the ABS and RDD surveys were close to the American Community Survey (ACS) distribution by sex and income (Table 5-1). The RDD survey had slightly more females than the national benchmark (50.0% vs. 48.7%; p=.288), while the ABS is slightly lower (47.1% vs. 48.7%; p=.076). With respect to income, both samples are close to the national distribution as measured by the ACS.

2. Both surveys underrepresent individuals who have low education. Both the RDD and ABS surveys are well below the national benchmark for those that have less than a high school degree (RDD - 24.6%, ABS - 21.1%, 36.4% - ACS). The two surveys differ somewhat for the other two education groups. The RDD has significantly more persons in the highest level of education (Bachelor's degree or higher) than benchmark (46.6% vs. 28.9%) and fewer among those with some college (28.8% vs. 34.3%). The ABS survey also over-represents those in the highest education group (44.6% vs. 28.9%) but is very close for those who have some college (34.3% vs. 34.8%).

3. Both surveys over-represent non-Hispanic Whites (RDD 68.7%, ABS 71.3%, and ACS 64.1%).

4. Both RDD and ABS underrepresent Hispanics. RDD underrepresents by a larger amount than ABS (8.8% RDD, 12.2% ABS, and 15.7% ACS). RDD has good representation of non-Hispanic Blacks (12.5% RDD, and12.0% ACS), while ABS underrepresents this group (6.6% ABS).[7]

5. Both surveys over-represent married individuals and underrepresent those who have never been married. The ABS had the highest proportion of married individuals (57.3% ABS, 53.5% RDD, and 50.4% ACS) and least never married (20.0% ABS, 25.6% RDD, and 30.0% ACS).

---

[7] The survey was administered in English. No Spanish version was administered.

6.   Both RDD and ABS over-represent those who have access to the internet by a considerable amount. With respect to those with any type of access, almost all of the RDD (96.4%) and ABS (98.6%) have access. This is about 10 percentage points higher than the ACS estimate (87.4%).

7.   Table 5-2 provides the distributions for the same characteristics once applying the final population weights. The final population weights were created by raking to ACS distributions for age, race-ethnicity, marital status, and education. Given this, it is not surprising that both RDD and ACS are very close to the ACS distributions for these characteristics. The distributions for the other characteristics shift a bit. For example, there is now an overrepresentation in the lower income groups.

Comparing Selected Health Measures to the National Health Interview Survey

One shortcoming of using demographics to assess NRB is the analysis does not account for the fact that the population weights compensate for differences in the national distribution for many of the characteristics of interest. It only provides an indication of differences before the final weights are applied. A second method to assess the representation of the sample is to examine the alignment for characteristics that might be related to nonresponse but are not used to develop the weights. To apply this method the analysis compares several different measures of health and injuries to the NHIS. These comparisons use the final NISVS weights which adjust to align the population distributions to national totals. The assumption is that since the NHIS has a high response rate it can be used as a standard against which to assess potential NRB of the Feasibility Study surveys. To do this, several questions from the NHIS were placed on the Feasibility Study surveys.

There are differences between the NISVS surveys and NHIS that may confound the comparisons. One is that not all questions were worded exactly as on the NHIS. A second difference is the mode of interviewing. The NHIS is a mix of both in-person and telephone interviewing. A third difference is who is selected as the respondent. For some questions the NHIS respondent is a household respondent, while it is a randomly selected person for the Feasibility Study surveys.

Nonetheless, it is still useful to compare the ABS and RDD surveys to the NHIS when assessing the extent of NRB.

8.   The surveys included a question on whether any household member had a physical, mental or emotional problem preventing them from working (Table 5-3). The RDD estimate is virtually the same as for the NHIS (12.9% RDD and 12.8% NHIS). The estimate for the ABS is higher (14.8%) than both, but only by two percentage points (p<.20).

9.   The surveys included a question on whether an adult in the household had been hospitalized overnight in the last 12 months. The RDD estimate (19.4%) is above the NHIS estimate (15.6%). The ABS estimate is below (12.3%).

10.   All surveys included a question on whether the respondent had been told he or she had asthma. The RDD estimate (13.2%) is virtually the same as the NHIS (13.4%). The ABS estimate is slightly above the NHIS estimate (14.1% vs. 13.4%). The difference between the RDD and ABS is not statistically significant (p=.325).

11.	The NISVS estimates for those depressed are somewhat higher than the NHIS. The RDD estimate is 27.7 percent compared to 24.2 percent for the ABS. Both of which are significantly higher than the estimate for the NHIS of 15.7 percent. The RDD is significantly higher than the ABS (p<.05).

12.	Both surveys included a question, originally taken from the NHIS, on whether any adult in the household had been injured in the last 3 months. Since fielding the surveys we have learned that this particular question has been taken off the public use file. To date, we have been unable to process the data to retrieve an equivalent estimate from the NHIS. To derive the estimate, it will be necessary to get access through the NCHS restricted use enclave.

	The ABS estimate is higher than the RDD estimate (27.2% ABS vs 20.7% RDD; p<.000).

Comparing by Level of Effort to Complete the Survey

Using demographics and health characteristics to assess NRB does not directly test whether the outcomes of interest are biased. There will only be bias if the characteristic is correlated with both nonresponse and the particular outcome. For example, both the ABS and RDD over-represent those with an internet connection. For this to lead to NRB there would need to be a correlation between internet access and a particular type of victimization. Another method to test for NRB that does not rely on this assumption is to compare the outcomes of interest by how much effort the survey operations take to complete the survey. This type of analysis, called a continuum of resistance model, assumes those respondents who require more effort to complete the survey resemble those who do not respond at all. For the Feasibility survey, this can be accomplished by comparing those who responded after the initial attempts were made to get a completed survey to those who responded during the NRFU. To motivate nonrespondents to participate as part of the NRFU, respondents were offered more money to complete the survey.

The logic of this analysis is that if the NRFU estimates are higher than Phase 1 then this is indicative that the estimates discussed in the prior chapter are too low or a negative bias. That is if the NRFU respondents represent the non-respondents, then getting 100% response would entail adding individuals who have higher estimates. The opposite is the case if the NRFU estimates are lower than the Phase 1 estimates.

One disadvantage of the continuum of resistance methodology is the assumption that those who responded to the NRFU resemble the non-responders. This assumption can be particularly problematic for surveys with response rates as low as 10 to 30 percent. In this case the assumption is that those responding to the NRFU adequately represent the 70 to 90 percent who did not respond. Nonetheless, it is still useful to assess how those who did exhibit some reluctance to survey and eventually cooperated differ for the main outcomes of interest to the NISVS.

13.	The ABS prevalence estimates for the total population before NRFU are largely less than the NRFU estimates, indicating a negative bias (rates are too low) (Table 5-4). Many of the differences are statistically significant at least at the 5 percent level. However, this pattern is not consistent across males and females (Tables 5-5, 5-6).

	-	Males (Table 5-5). The direction of the difference between before NRFU and the NRFU are both negative and positive. None of these differences are statistically significant.

- <u>Females</u> (Table 5-6). All of the differences are negative, indicating a negative bias in the estimates (i.e., survey estimates are too low). All of these are statistically significant.

14. The differences for RDD are a mixture of both negative and positive signs (Table 5-4). The differences for rape and IPPV are negative and statistically significant. The differences for contact sexual violence by intimate partner and stalking by an intimate partner are positive and statistically significant. The patterns differ by sex.

    - <u>Males</u> (Table 5-5). The direction of the difference between before NRFU and the NRFU are primarily negative. Many of the significant effects are for the components of contact sexual violence including unwanted touching, sexual coercion and rape. In addition, for IPPV and IPPA the differences are negative and statistically significant.

    - <u>Females</u> (Table 5-6). None of the differences for contact sexual violence or stalking are statistically significant. The differences for Contact Sexual Violence by Intimate Partner and Stalking by Intimate Partner is positive and significant.

Comparing to Prior NISVS with Different Response Rates

Like virtually all RDD studies, the NISVS response rate has declined between 2010 (33.1%) and 2016/2017 (8.5%). At the same time, the lifetime victimization estimates have increased. Between the 2015 and 2016/2017 administrations, the response rate dropped from 26.4 percent to 8.5 percent. At the same time, the lifetime prevalence estimates increased (Table 5-7). For example, the female estimate of contact sexual violence increased from 43.6 percent to 54.3 percent. One would not expect the lifetime estimates to change this much in a 1- or 2-year time-period. The age-cohorts are not shifting dramatically. Theoretically, women age 18+ in 2015 would be between ages 19 and 20+ in 2016/2017. Lifetime experiences may change a bit, but it would be unusual that it would change as much as indicated.

A second explanation is NRB. Prior studies have not found a strong correlation between measures of NRB and the survey's response rate (Groves and Peytcheva, 2008).[8] Nonetheless, theoretically there are reasons to believe that as response rates go down, the chances of NRB goes up. In a re-analysis of data from Groves and Peytcheva (2008), Brick and Tourangeau (2017)[9] found study-level correlations of between -0.4 and -0.5 between the response rate and the average NRB for a study. The co-occurrence between the drop in response rate and spike in prevalence rate nominally suggests this explanation may apply.

An extensive analysis of NRB was conducted by CDC when assessing the quality of the 2016/2017 data set.[10] This included an assessment of changes in the demographic distributions of survey respondents across the three time periods the NISVS has been collected. In addition, the analysis compared measures of medical conditions (e.g., asthma) collected on NISVS to the same measures

---

[8] Groves, R.M., and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly,* 72(2), 167-189. <u>Impact of Nonresponse Rates on Nonresponse Bias | Public Opinion Quarterly | Oxford Academic (oup.com)</u>

[9] Brick, J.M., and Tourangeau, R. (2017). Responsive survey designs for reducing nonresponse bias. *Journal of Official Statistics,* 33(3), 735-752. <u>Responsive Survey Designs for Reducing Nonresponse Bias (sciendo.com)</u>

[10] Kresnow, M.J, Holland, K., Peytchev, A, Chen, J., Smith, S.G. and T. Simon (2021) NISVS 2016/2017 Assessment Report: An Examination of data representativeness and factors contributing to observed increases in estimates of violence victimization in the presence of low response rates. Unpublished report, CDC.

from surveys with higher response rates. The consistency of these same health measures was also measured over the NISVS time series. The theory being that if there was significant NRB, then it may also have affected these health measures. None of these analyses suggest significant NRB for the 2016/2017 data. As noted above, however, these types of analyses do not directly test whether there is NRB for the measures of victimization.[11]

Another possible explanation discussed in the CDC report is an increased willingness to report victimization on the part of the general public. As indicated by social media searches,[12] a number of highly publicized incidents boosted consciousness of sexual violence during this time period. Google trends show high interest on topics like rape or sexual assault during the 2012 – 2019 period. This is a logical explanation, but still not a direct link to the trends. Social media does not provide a clear measure of the views of the general public, so it is not clear how to generalize the observed trends to how respondents would answer questions on a survey. The trends of the NISVS during this time period run counter to several other surveys that have been conducted over the same time period and had relatively stable response rates. For example, the estimates of forced intercourse for females and males measured by the National Survey of Family Growth (NSFG) did not change for the time periods 2011-15 and 2015-17. A survey of 21 colleges in 2015 and 2019 did not find significant increases in estimates for the large majority of the schools.[13] The rape and sexual assault estimate from the National Crime Victimization Survey (NCVS) has remained stable over the 2015-17 time period.

A third explanation that may account for the jump in the lifetime estimates for the 2016/2017 data are changes in the questionnaire between 2015 and 2016/2017. The change streamlined the questionnaire:

- by doing less follow-up in the 2016/2017 instrument. Specifically, rather than asking about perpetrators after each 'yes' to a victimization screening item, the follow-ups were done after each section (e.g., all stalking screening items; all physically forced rape screening items, etc.).

- by changing the follow-up perpetrator questions. In 2015 the initials of perpetrators were collected and these were used to check overlap between perpetrators across different types of victimizations. This was not done in 2016/2017.

The more detailed follow-up in 2015 may have led some respondents to learn that more follow-up questions were being asked once they endorse an item. This could depress victimization estimates relative to 2016/2017 if respondents do not endorse a screening item to avoid the follow-ups.[14]

For purposes of discussion below, we assume this change was not primarily responsible for changes in estimates between 2015 and 2016/2017. Neither the 2015 nor the 2016/2017 are

---

[11] The analysis also compares selected estimates from NISVS to the NSFG. The results of this analysis differ from the results we report in Chapter 6. The CDC analysis indicates close correspondence to the NSFG estimate, while the analysis in Chapter 6 found the NISVS estimate from the Feasibility survey is significantly higher than the NSFG.

[12] Kresnow, M., Chen, J., Peytchev, A., Holland, K., Smith, S., and Simon, T. (2020). Do low response rates automatically preclude use of data? An evaluation of possible influences on survey estimates to determine data usability. Poster presented at the 2020 Annual Meeting of the American Association for Public Opinion Research.

[13] Cantor, D., Fisher, B., Chibnall, S., Harps, S., Townsend, R., Thomas, G., Lee, H., Kranz, V., Herbison, R., and Madden, K. (2020). *Report on the AAU Campus Climate Survey on Sexual Assault and Misconduct*. Westat, Rockville, MD. Available at: https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Campus-Safety/Revised%20Aggregate%20report%20%20and%20appendices%201-7_(01-16-2020_FINAL).pdf

[14] CDC is currently working with NCHS to test whether this was occurring.

'grouped' designs. A grouped design holds all follow-up questions until after all screening items are completed. Grouped designs have been shown to lead to higher reports of events relative to interleafed designs. [15] Both 2015 and 2016/2017 are interleafed designs. Both have follow-up questions in between screening items. The difference between 2015 and 2016/2017, from this perspective, is a matter of degree.

With the above caveats in mind, this section compares the 2015 and 2016/2017 results to the ABS and RDD surveys. The analysis makes the assumption that the 2015 survey was subject to less NRB than the 2016/2017 survey. If this is true, then significant deviations from 2015 is indicative of NRB. Similarly, estimates close to the 2016/2017 survey are interpreted as a sign of NRB. As discussed above, attributing the change in 2016/2017 to NRB is subject to debate and may only be part of the explanation for the increase.

To make the comparison, the NISVS estimates for 2015 were taken from the published report for that year. The CDC provided the 2016/2017 estimates.[16]

16. The difference between ABS and 2015 is consistently smaller than the difference between ABS and 2016/2017 (Table 5-8). For females, the gap is very large for most types of victimizations. For example, for contact sexual violence the absolute difference with 2015 is 1.4 percentage points. This compares to 12.1 percentage points for the difference with 2016-17. For males the differences are mostly in the same direction, but not as large.

17. The RDD estimates for females are generally closer to 2016/2017. This is most evident for sexual coercion, and intimate partner psychological aggression (IPPA). The results are more mixed for males where the RDD and ABS estimates are not consistently above or below the 2016/2017 estimates.

18. The RDD estimates for males are uniformly closer to the 2016/2017, by a larger amount than for females.

---

[15] Kreuter, F., S. McCulloch, S. Presser, and R. Tourangeau (2011), "The Effects of Asking Filter Questions in Interleafed Versus Grouped Format," Sociological Methods and Research, 40, 88–104.

[16] The questions used to determine stalking changed between 2016/2017 and the feasibility surveys. The change tended to reduce the number of stalking victims identified on the feasibility surveys. For this reason, this is not included in this analysis. There were other changes in the NISVS questionnaire between 2015 and 2016/2017. These are discussed after presenting the results for this subsection.

	Estimates (percent)			p-value		
	RDD	**ABS**	**ACS**	**RDD-ACS**	**ABS-ACS**	**ABS-RDD**
Age						
18-29	13.9	19.5	21.5	0.000	0.020	0.000
30-44	19.7	25.7	25.1	0.000	0.302	0.000
45-64	25.9	24.4	25.7	0.396	0.116	0.262
65+	40.4	30.4	27.7	0.000	0.001	0.000
Sex						
Male	50.0	47.1	48.7	0.288	0.076	0.118
Female	50.0	52.9	51.3	0.288	0.076	0.118
Marital status						
Married	53.5	57.3	50.4	0.074	0.000	0.054
Never married	25.6	20.0	30.0	0.000	0.000	0.000
Other	20.8	22.8	19.5	0.277	0.000	0.205
Race						
Hispanic	8.8	12.2	15.7	0.000	0.000	0.003
NH-White	68.7	71.3	64.1	0.003	0.000	0.120
NH-Black	12.5	6.6	12.0	0.351	0.000	0.000
NH-Multiracial	3.5	2.6	1.7	0.006	0.003	0.176
NH-Other	6.5	7.3	6.6	0.398	0.125	0.292
Education						
High school or less	24.6	21.1	36.4	0.000	0.000	0.121
Some college	28.8	34.3	34.8	0.012	0.374	0.044
Bachelor's or higher	46.6	44.6	28.9	0.000	0.000	0.293
Income						
Less than $25,000	18.1	17.3	15.7	0.127	0.030	0.359
$25,000 - $49,999	18.4	18.6	20.0	0.254	0.063	0.398
$50,000 - $74,999	16.5	15.9	18.3	0.095	0.000	0.356
$75,000+	47.0	48.3	46.0	0.343	0.022	0.332
Born in United States						
Yes	90.1	87.5	81.6	0.000	0.000	0.029
No	9.9	12.5	18.4	0.000	0.000	0.029
Access to internet not including through cellphone						
Yes	90.4	94.2	75.6	0.000	0.000	0.000
No	9.6	5.8	24.4	0.000	0.000	0.000
Any access to internet						
Yes	96.4	98.6	87.4	0.000	0.000	0.000
No	3.6	1.4	12.5	0.000	0.000	0.000
Home ownership						
Owned	72.8	73.5	66.8	0.000	0.000	0.370
Rented	25.4	25.6	31.6	0.000	0.000	0.395
Other	1.9	0.9	1.6	0.337	0.000	0.047

Table 5-1. Comparison of demographic distributions for RDD, ABS, and American Community Survey (ACS)*#

\# Except where noted, estimates combine male and females * Unweighted with an adjustment for probability of selection. Significance tests assume no sampling error for the ACS.

- For ABS, this consists of weighting for selection within the household and selection for the nonresponse follow-up (NRFU). For RDD, this adjustment had three components: 1) for the landline sample an adjustment was made for selection within the household; 2) for both landline and cellphone frames an adjustment for selection for the Phase 2 NRFU; and 3) for both frames, an adjustment for the probability of selection into the landline or cellphone frame.

Table 5-2.	Comparison of final weighted demographic distributions for RDD, ABS, and American Community Survey (ACS)+#					
	Estimates (percent)			p-value		
	RDD	ABS	ACS	RDD-ACS	ABS-ACS	ABS-RDD
Age [x]						
18-29	21.5	21.6	21.5			
30-44	25.1	25.2	25.1			
45-64	25.7	25.7	25.7			
65+	27.7	27.5	27.7			
Sex [x]						
Male	49.2	48.4	48.7			
Female	50.8	51.6	51.3			
Marital status [x]						
Married	50.6	50.5	50.4			
Never Married	30.1	30.0	30.0			
Other	19.3	19.5	19.5			
Race [x]						
Hispanic	15.8	15.8	15.7			
NH-White	64.0	64.1	64.1			
NH-Black	12.0	12.0	12.0			
NH-Multiracial	1.7	1.7	1.7			
NH-Other	6.5	6.4	6.6			
Education [x]						
High school or less	32.9	36.2	36.4			
Some college	38.3	34.8	34.8			
Bachelor's or higher	28.9	29.0	28.9			
Income						
Less than $25,000	21.5	23.4	15.7	0.001	0.000	0.246
$25,000 - $49,999	19.2	21.6	20.0	0.343	0.080	0.148
$50,000 - $74,999	17.9	15.6	18.3	0.385	0.000	0.137
$75,000+	41.4	39.5	46.0	0.003	0.000	0.203
Born in United States						
Yes	89.5	86.3	81.6	0.000	0.000	0.011
No	10.5	13.7	18.4	0.000	0.000	0.011
Access to internet not including through cellphone						
Yes	89.1	92.2	87.4	0.123	0.000	0.017
No	10.9	7.8	12.6	0.123	0.000	0.017
Access to internet						
Yes	96.4	98.6	87.4	0.000	0.000	0.011
No	3.6	1.4	12.6	0.000	0.000	0.011
Home ownership						
Owned	72.8	73.5	66.8	0.341	0.007	0.352
Rented	25.4	25.6	31.6	0.220	0.046	0.386
Other	1.9	0.9	1.6	0.127	0.017	0.034

Westat®

\# Except where noted, estimates combine male and females; + p-value is the result from a two-sample difference of means z-test.

X – Characteristic is a raking dimension. There are no statistical significance tests computed because the totals were forced to the national distribution.

- Final population weights used for estimates. Significance tests assume ACS has no sampling error.

Table 5-3.	Comparison of final weighted health indicators for RDD, ABS, and national benchmarks+#					
	Estimates (percent)			p-value		
	RDD	ABS	NHIS	RDD-NHIS	ABS-NHIS	ABS-RDD
Any adult in the household have physical, mental, or emotional problem preventing from working						
Yes	12.9	14.8	12.8	0.397	0.015	0.200
No	87.1	85.2	87.2	0.397	0.015	0.200
Any adult in the household been hospitalized overnight in last 12 months						
Yes	19.4	12.3	15.6	0.005	0.000	0.000
No	80.6	87.7	84.3	0.007	0.000	0.000
Doctor, nurse, or other health professional told you that you have asthma						
Yes	13.2	14.1	13.4	0.391	0.279	0.325
No	86.8	85.9	86.5	0.383	0.308	0.325
Doctor, nurse, or other health professional told you that you have any type of depression						
Yes	27.7	24.2	15.7	0.000	0.000	0.022
No	72.3	75.8	84.2	0.000	0.000	0.022
Any adult Injured in household in last 3 months						
Yes	20.7	27.2	n/a	n/a	n/a	0.000
No	79.3	72.8	n/a	n/a	n/a	0.000

n/a – Not applicable. Estimate could not be computed because the public use dataset did not have the relevant variables.

\# Estimates combine male and females

+ p-value is the result from a two-sample difference of means z-test.

Table 5-4.	Lifetime prevalence estimates by Phase 1 and NRFU stages for ABS and RDD surveys+#							
	ABS				**RDD**			
	Before NRFU	**NRFU**	**Difference**	**p-value**	**Before NRFU**	**NRFU**	**Difference**	**p-value**
Contact sexual violence	29.7	34.9	-5.2	0.016	46.8	48.3	-1.4	0.353
Unwanted touching	23.5	30.6	-7.1	0.001	41.4	43.6	-2.2	0.290
Sexual coercion	11.8	14.6	-2.8	0.092	18.1	19.7	-1.6	0.322
Rape	11.0	16.8	-5.8	0.002	17.1	23.1	-6.0	0.014
Made to penetrate*	2.5	2.0	0.5	0.350	3.7	3.2	0.5	0.375
Stalking	8.9	12.6	-3.7	0.034	14.5	11.9	2.6	0.212
Intimate partner violence	27.7	35.2	-7.6	0.001	38.4	41.5	-3.1	0.211
Contact sexual violence by intimate partner	11.6	13.1	-1.4	0.270	15.5	10.6	4.9	0.033
Stalking by intimate partner	4.8	7.4	-2.7	0.083	7.5	3.1	4.4	0.013
Intimate partner physical violence	22.0	28.0	-6.0	0.004	32.4	39.9	-7.5	0.007
Intimate partner psychological aggression	32.1	40.2	-8.1	0.000	46.5	49.7	-3.3	0.221

Estimates combine male and females
+ p-value is the result from a two-sample difference of means z-test.
* Made to penetrate only asked of males.

Table 5-5.	Male lifetime prevalence estimates by Phase 1 and NRFU stages for ABS and RDD surveys+							
	ABS				**RDD**			
	Before NRFU	**NRFU**	**Difference**	**p-value**	**Before NRFU**	**NRFU**	**Difference**	**p-value**
Contact sexual violence	18.0	13.8	4.3	0.077	29.2	35.6	-6.4	0.084
Unwanted touching	12.8	13.5	-0.7	0.380	23.0	31.8	-8.8	0.017
Sexual coercion	5.8	5.6	0.2	0.396	7.7	15.8	-8.1	0.005
Rape	2.8	5.4	-2.6	0.120	2.6	12.2	-9.6	0.000
Made to penetrate	5.2	3.9	1.3	0.275	7.5	6.7	0.8	0.377
Stalking	5.2	4.1	1.0	0.320	7.5	5.2	2.3	0.236
Intimate partner violence	20.5	21.6	-1.1	0.360	33.1	47.0	-13.8	0.000
Contact sexual violence by intimate partner	5.5	3.0	2.5	0.101	5.7	6.8	-1.1	0.358
Stalking by intimate partner	2.0	1.9	0.1	0.398	3.8	2.2	1.6	0.278
Intimate partner physical violence	18.0	20.2	-2.2	0.265	31.4	44.5	-13.1	0.000
Intimate partner psychological aggression	29.1	30.4	-1.3	0.350	42.6	53.2	-10.6	0.007

+ p-value is the result from a two-sample difference of means z-test.

Table 5-6.	Female lifetime prevalence estimates by Phase 1 and NRFU stages for ABS and RDD surveys+							
	ABS				**RDD**			
	Before NRFU	**Nonresponse Follow-up (NRFU)**	**Difference**	**p-value**	**Before NRFU**	**Nonresponse Follow-up (NRFU)**	**Difference**	**p-value**
Contact sexual violence	40.3	56.1	-15.8	0.000	64.1	59.7	4.4	0.196
Unwanted touching	33.3	47.5	-14.2	0.000	59.4	54.3	5.2	0.145
Sexual coercion	17.2	23.3	-6.1	0.010	28.3	23.2	5.1	0.131
Rape	18.5	28.3	-9.8	0.000	31.3	33.0	-1.8	0.352
Made to penetrate	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Stalking	12.3	21.0	-8.7	0.000	21.4	18.0	3.4	0.239
Intimate partner violence	34.2	48.9	-14.7	0.000	43.6	36.5	7.1	0.051
Contact sexual violence by intimate partner	17.2	23.2	-6.0	0.011	25.1	14.1	11.1	0.002
Stalking by intimate partner	7.3	12.9	-5.6	0.009	11.1	3.9	7.3	0.003
Intimate partner physical violence	25.6	35.8	-10.2	0.000	33.3	35.7	-2.3	0.308
Intimate partner psychological aggression	34.9	50.3	-15.5	0.000	50.3	46.6	3.7	0.243

n/a – Not applicable. Not asked of female respondents.
+ p-value is the result from a two-sample difference of means z-test.

	NISVS			Feasibility	
	2010-2012	2015	2016/2017	ABS	RDD
Response rate	33.0%	26.4%	8.5%	33.1%	10.3%
Female					
Contact sexual violence	36.3	43.6	54.3	42.2	63.3
Unwanted sexual contact	27.5	37.0	47.6	35.0	58.5
Sexual coercion	13.2	16.0	23.6	17.9	27.3
Rape	19.1	21.3	26.8	19.7	31.6
Made to penetrate	0.5	1.2	n/a	n/a	n/a
Contact sexual violence by intimate partner	16.4	18.3	19.5	17.9	23.1
Intimate partner physical violence	32.4	30.6	42.0	26.8	33.8
Intimate partner psychological aggression	47.1	36.4	49.4	36.7	49.6
Male					
Contact sexual violence	17.1	24.8	30.7	17.5	30.3
Unwanted sexual contact	11.0	17.9	23.3	12.9	24.5
Sexual coercion	5.8	9.6	10.9	5.8	9.1
Rape	1.5	2.6	3.8	3.1	4.2
Made to penetrate	5.9	7.1	10.7	5.0	7.4
Contact sexual violence by intimate partner	7.0	8.2	7.5	5.2	5.9
Intimate partner physical violence	28.3	31.0	42.3	18.3	33.7
Intimate partner psychological aggression	47.3	34.2	45.1	29.2	44.4

n/a – Not applicable. Not asked of female respondents.

- 2010 - 2012 data taken from: Smith, S.G., Chen, J., Basile, K.C., Gilbert, L.K., Merrick, M.T., Patel, N., Walling, M., and Jain, A. (2017). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010-2012 State Report*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

- 2015 data taken from: Smith, S.G., Zhang, X., Basile, K.C., Merrick, M.T., Wang, J., Kresnow, M., and Chen, J. (2018). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2015 Data Brief – Updated Release.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

- 2016/2017 unpublished data from Centers for Disease Control and Prevention.

	Absolute difference from ABS			Absolute difference from RDD		
	2015	**2016-17**	**2015 < 2016-17**	**2015**	**2016-17**	**2015 > 2016-17?**
Female						
Contact sexual violence	1.4	12.1	yes	19.7	9.0	Yes
Unwanted sexual contact	2.0	12.7	Yes	21.5	10.9	Yes
Sexual coercion	1.9	5.7	Yes	11.3	3.8	Yes
Rape	1.6	7.1	Yes	10.3	4.8	Yes
Made to penetrate	n/a	n/a	n/a	n/a	n/a	n/a
Contact sexual violence by intimate partner	0.4	1.6	Yes	4.8	3.6	Yes
Intimate partner physical violence	3.8	15.2	Yes	3.2	8.2	No
Intimate partner psychological aggression	0.3	12.7	Yes	13.2	0.2	Yes
Male						
Contact sexual violence	7.3	13.2	Yes	5.5	0.3	Yes
Unwanted sexual contact	5.0	10.5	Yes	6.6	1.2	Yes
Sexual coercion	3.8	5.1	Yes	0.5	1.8	No
Rape	0.5	0.6	Yes	1.6	0.4	Yes
Made to penetrate	2.1	5.6	Yes	0.3	3.3	No
Contact sexual violence by intimate partner	3.0	2.3	No	2.3	1.6	Yes
Intimate partner physical violence	12.7	24.0	Yes	2.7	8.6	No
Intimate partner psychological aggression	5.0	15.9	Yes	10.2	0.7	Yes

Table 5-8.   Comparison of lifetime prevalence estimates by sex for ABS and RDD to NISVS 2015 and 2016-17#

n/a – Not applicable. Not asked of female respondents.
Nominal differences; not tested for statistical significance.

6. How Do ABS and RDD Compare with Respect to Other Measures of Data Quality and Two Key Survey Outcomes?

Summary

➢ There is more item-missing data on the web and paper surveys compared to the computer-assisted telephone interview (CATI). The impact of excluding those that are missing on at least one item depends on the particular measure.

➢ Comparisons of lifetime estimates of rape and made to penetrate to the National Survey of Family Growth (NSFG) for 2015-2017 suggest the address-based sample (ABS) estimate is biased in a negative direction, while the random digit dial (RDD) is biased in a positive direction. The ABS is considerably closer to the NSFG estimate than the RDD.

➢ The relationship between age and victimization is similar for the RDD and ABS for both 12- month and lifetime prevalence. The youngest age groups for the RDD have unusually high lifetime estimates relative to other age groups, perhaps suggesting NRB for this group.

➢ The ABS and RDD collect comparable data on the number of perpetrators. The ABS had a higher proportion of perpetrators who were intimate partners.

➢ The RDD respondents reported more serious consequences than ABS respondents.

This chapter compares the ABS and RDD surveys along several measures of data quality, including: 1) item missing data for key measures, 2) comparison to the NSFG, 3) the relationship between age and victimization for the ABS and RDD, and 4) measures of intimate partners and consequences of victimization.

Item-Missing Data

One issue with a self-administered questionnaire is the amount of item-missing data. Only one of the items on the web survey was mandatory for the respondent to enter a legitimate response (age). For any other item the respondent could skip the question. The same applies for the paper survey but missing could also result from failure to follow skip instructions. This section presents several different summary measures for item-missing data on the ABS and RDD surveys. The prevalence estimates provided in Chapter 3 treated item-missing data as if the respondent reported that they were not victimized.[17] This is consistent with how the current National Intimate Partner and Sexual Violence Survey (NISVS) treats item-missing data. This section provides data on how much missing data there are for each of the modes and the implications for the estimates of victimization.

---

[17] The only exception to this is for someone who didn't answer any of the screening items for a particular type of victimization. These individuals were treated as missing when computing the prevalence estimates in Chapter 3.

Two different sets of items were examined. One was whether the respondent answered the items that everyone is asked, hereafter referred to as "required items." The second set of items are follow-ups to the victimization questions, hereafter referred to as "follow-up items." The follow-up items are needed to determine 12-month estimates and estimates for intimate partner violence (IPV). On Web and CATI surveys these follow-ups are administered in the following ways:

- Determining the relationship between victim and offender in cases of reported IPV. For each perpetrator the respondent is asked how they know the person.

- Determining if an incident occurred within the last 12 months. This is a follow-up asked for each lifetime victim on whether any incidents occurred within the last 12 months.

As described in Chapter 2, for a web survey to be defined as a complete, the respondent had to at least complete the section on attempted physically forced sex. In addition, the respondent had to answer at least one of the required victimization questions for at least half of the different types of victimizations. The data provided below are for those who meet these criteria and are defined as a complete or partial complete.

1. The average amount of missing data for the required items on the RDD was 0.6 percent and for ABS it was 1.7 percent (data not shown).

2. For the web, on average, 1.4 percent of the required items were missing. For the paper survey, 6.8 percent of the required items were missing (data not shown). The paper survey included a number of skips that required reading instructions and going to the correct item. The higher rate of missing data with the paper survey is due to this. For the web, the computer controls navigation of the skip patterns.

3. For the web, the amount of missing data for the required items did not differ significantly by type of device (personal computer vs. smartphone) (data not shown).

The web, paper, and CATI surveys differ with respect to how the information on intimate partners and 12-month victimizations are collected. [18] This potentially affects the missing data for measures related to these two characteristics.

- CATI. For each perpetrator, the relationship to the victim is collected using the detailed codes available to the interviewer. Once collecting all the relationships, a single question is asked whether any of the incidents occurred within the last 12 months.

- Web. For each perpetrator, the relationship to the victim is collected by two questions. The first question has eight broad categories:

  1. Someone involved with romantically or sexually *at the time*
  2. Someone *previously* involved with romantically or sexually
  3. A friend or acquaintance
  4. Knew through work
  5. Person of authority
  6. Knew less than 24 hours

---

[18] If at least one of the targeted items indicated an intimate partner the case was not considered missing. For example, if someone skipped the relationship question for one perpetrator but reported another perpetrator was an intimate partner, the case was not counted as missing.

7.   A complete stranger
8.   Someone else

If any category is selected other than "a complete stranger" or "someone else," a second more detailed list is shown. The specific detailed list shown is based on the broad relationship category selected and the sex of the respondent. The data on whether it occurred in the last 12 months is then collected for this perpetrator.

- Paper. Once the information on lifetime victimization is collected, the respondent is provided a list of relationships considered to be an intimate partner. A follow-up question is asked if any of the incidents occurred in the last 12 months. The respondent is then asked if any of the incidents occurred with an intimate partner as defined by a list shown on the questionnaire.

The design for the collection of relationship on the web and paper were adaptations for each mode. The lists used by the interviewer on the CATI are too long and complicated to be presented on the web or paper. The web adapted this list using an unfolding design. The paper simplified the information to accommodate navigation and space on the hard-copy instrument. The intent of the design for collecting the 12-month information on the web is to make the task easier for the respondent by not asking them to summarize across all perpetrators.

For the lifetime non-intimate partner prevalence measures (combining male and female victimizations), the amount and possible impact of not answering at least one of the lifetime questions is provided in Table 6-1.

4.   Web. The amount of missing data ranges from .6 percent to 6.1 percent. Rape (5.1%) and Made to Penetrate (6.1%) have the highest estimates. If missing data are excluded, Contact Sexual Violence for web respondents increases by 1.8 percentage points (33.6% to 35.4%). The other measures increase by less than 1 percentage point, most less than 0.5 percentage point.

5.   CATI. The amount of missing data ranges from 0.1 percent to 3.5 percent. If missing data is excluded the contact sexual violence estimate increases by .9 percentage points.

6.   Paper. The amount of missing data ranges from .5 percent for stalking to 20.3 percent for contact sexual violence. The estimates for contact sexual violence increase by 8.2 percentage points if missing data are excluded.

For the 12-month estimates of non-intimate partner violence, the amount and impact of excluding missing data because the 12-month question was not answered is summarized in Table 6-2.

7.   Web. The amount of missing was .5 percent or less. Excluding missing data has no noticeable effect on the prevalence estimates.

8.   CATI. There was virtually no missing data because the 12-month question was not answered.

9.   Paper. The amount of missing data ranged from .6 percent (Rape) to 3.4 percent (Contact Sexual Violence). If the missing data are excluded, the Contact Sexual Violence estimate goes up by 0.2 percentage points.

For the estimates for IPV, the amount and impact of excluding missing data because the relationship information was not provided for at least one of the perpetrators is summarized in Table 6-3.

10. <u>Web</u>**.** The percentage missing the relationship for at least one perpetrator ranged from .4 percent for stalking to 4.0 percent for IPV. If the missing is excluded the IPV estimate goes up 1.5 percentage points.

11. <u>CATI</u>**.** The percentage missing the relationship for at least one perpetrator ranged from .4 percent for stalking to 1.8 percent for IPV. If the missing is excluded, the IPV estimate goes up 0.5 percentage points.

12. <u>Paper</u>**.** The percentage missing the relationship for at least one perpetrator ranged from 0.0 percent for stalking to 5.9 percent for IPV. If the missing is excluded, the IPV estimate goes up 2.2 percentage points.

Comparison to the National Survey of Family Growth (NSFG)

Both the ABS and RDD surveys included questions from the NSFG. The purpose was to be able to compare estimates of sexual violence victimization between the different surveys. The NSFG is an in-person survey that administers questions on nonconsensual sexual contact to both men and women age 15-49. The NSFG response rate was 67 percent for females and 64 percent for males.[19] To minimize problems with social desirability bias, it administers these questions with a self-administered survey using audio computer-assisted interviewing (ACASI). The data used in this analysis is from 2015 to 2017.[20]

Differences between the NSFG and the Feasibility Study surveys may be indicative of NRB. The response rate for the NSFG is considerably higher than the Feasibility Study surveys. The differences may also reflect variations in the measurement procedures. As noted above, the NSFG is an in-person contact and uses a self-administered survey. Neither of the Feasibility Study surveys are in-person. The RDD uses an interviewer-administered instrument. The differences may also reflect the context of the surveys. The NISVS placed the NSFG items at the end of the questionnaire, after all of the sexual, physical and psychological violence questions were asked. For some respondents, the items may have sounded redundant. This may have changed how they interpreted the NSFG questions. These items on the NSFG are placed in the context of a health survey on topics related to reproduction and sexual activity (among other things).

For the NISVS, the NSFG question for women was:

> NSFG1   *At any time in your life, have you ever been forced by a male to have vaginal intercourse against your will?*

---

[19] See NSFG_2015-2017_UserGuide_Main Text (cdc.gov) p. 7.

[20] The most recent public release of the NSFG data is for the 2017-2019 data collection. However the variables needed for this analysis have been suppressed in the public use dataset due to modified disclosure risks. They will only be accessible through the NCHS Research Data Center (RDC). The RDC has not been operating during the COVID-19 period. For years prior to 2017-2019, NCHS has published the relevant statistics for these items. For the 2017-2019 data, this publication is scheduled for release some time in the spring of 2021.

Two questions were included for males:

> NSFG2   *At any time in your life, have you ever been forced by a female to have vaginal intercourse against your will?*

> NSFG3   *At any time in your life, have you ever been forced by a male to have oral or anal sex against your will?*

The wording on the NSFG for females is slightly different to account for questions that had previously been asked on the survey about involuntary and forced intercourse. Prior to the above question, NSFG respondents are asked:

> JE-2   *Would you say then that this first vaginal intercourse was voluntary or not voluntary, that is, did you choose to have sex of your own free will or not?*

Those who say "yes" to this question are asked a slightly different version of NSFG1:

> JE-5   *Besides the time you already reported, have you ever been forced by a male to have vaginal intercourse against your will?*

Those who said "yes" to JE-2 are automatically counted as having had forced vaginal intercourse against their will in the NSFG published estimates.

However, the use of the word "force" in NSFG1/JE-5 is not implied by the wording of JE-2. Involuntary sex might be interpreted as including tactics that are not considered "force," such as verbal pressure.[21] If that is true, then some NSFG respondents who report an involuntary encounter in JE-2 may not consider it by "force" and would answer "no" to JE-5 if it was asked in the form included on the NISVS.

The NSFG does include follow-up questions to JE-2 which ask about the tactic used the first time involuntary intercourse occurred. These tactics include

- being given alcohol or drugs;
- the person was bigger or was a grown up;
- threatened ending the relationship;
- pressured by words or actions without threats of harm;
- threatened physical hurt or injury;
- physical hurt or injured; and
- physically held down.

---

[21] We are not arguing that verbal pressure should not be considered "force." Our point is that some respondents may not interpret the word force as including verbal pressure.

We calculated three different estimates from the NSFG for females. One estimate replicates that published for the NSFG. This counts anyone who either said: 1) their first time having intercourse was involuntary intercourse and was not of their own free will (JE-2 = yes) or 2) they had been forced to have intercourse against their will (JE-5 = yes). The second estimate excludes from the first estimate anyone who only said their first time having intercourse occurred because he or she was pressured by words or actions without threats of harm. The third estimate takes out from the second estimate those who only said their first time was because of drugs or alcohol. The estimates from both the NSFG and the feasibility survey are for those age 18 to 49.

13. The published NSFG estimate of forced vaginal intercourse is 18.8 percent (Table 6-4). When removing from this estimate persons who reported being pressured by words or actions without threats of harm, the estimate drops to 17.5 percent. It goes down to 17.0 percent when further removing incidents when the victim was given alcohol or drugs.

14. The RDD estimates of forced vaginal intercourse by a male are significantly above the NSFG. For females, the RDD estimate for women is 29.2 percent compared to the published NSFG estimate of 18.8 percent. Similarly, for males reporting forced vaginal intercourse the RDD estimate is 6.9 percent compared to 4.3 percent for the NSFG. The estimate for being forced to have oral or anal sex for RDD is also above the published NSFG estimate, although the difference is not statistically significant (3.3% vs. 2.3%).

15. Two of the three ABS estimates are below the NSFG estimates. For females involving forced vaginal intercourse by a male, the estimate for the ABS is 12.3 percent compared to 18.8 percent for the published NSFG estimate. The three alternative estimates described above are somewhat closer but still significantly above the ABS. Similarly, the ABS estimate for forced vaginal intercourse for males is significantly below the NSFG estimate (2.6% vs. 4.3%). The estimate for males involving forced oral or anal sex with another male is above the NSFG, but the difference is not statistically significant.

16. The ABS estimate for females is considerably closer to the NSFG than the RDD (Table 6-4). For the female estimate, the difference between the ABS and the three NSFG estimates range from 4.7 to 6.5 percentage points. This compares to the differences for the RDD of 10.4 to 12.2 percentage points.

Comparing the Relationship Between Age and Victimization for ABS and RDD

The relationship between age and victimization is an important one for analysis of victimization. For very recent experiences (e.g., past 12 months), it is expected that victimization prevalence will be negatively associated with age. The relationship of age with lifetime victimization is not as clear. Aging increases cumulative exposure to risk. This should be reflected in higher lifetime estimates with age. When measured on a cross sectional survey, this assumes that the risk as one ages is similar over time. For example, the risk of a young person in 1960 is the same as that in 2020. Survey error might also affect the measurement of this relationship. As one ages it will be more difficult to recall victimizations that occurred in the distant past. There may also be differential NRB by age. Younger adults are more difficult to contact and gain cooperation on a survey. The weights adjust for age distributions, but if there are differences within the adjustment cells, there may still be NRB.

Given the importance of age for analyzing victimization, the analysis examined whether it was different for the RDD and ABS surveys. First, the relationship between age and the NISVS measures of lifetime victimization were compared to that of the NSFG (Table 6-5). This provides an external measure of the relationship with a survey with a high response rate. This analysis is restricted to those age 18-49, as this is the age range of the NSFG.

17. There is not a difference in the relationship between age and victimization for the three different surveys. Nominally there are a few exceptions, but none of the differences are statistically significant. For example, the estimate for forced intercourse for 18-29 year old males for RDD is noticeably higher for the youngest vs oldest age group (9.2 vs. 3.5). However, these are based on very small sample sizes and the difference is not statistically significant.

In order to look at the age-victimization relationship for the full age range, the relationship was estimated for the different forms of non-intimate partner violence and stalking for those age 18+. Because of small sample sizes, the 12-month victimization estimates were examined once combining across sex. For lifetime victimization the estimates were broken out by sex.

18. For both RDD and ABS, the relationship between age and past 12-month estimates is negative for all types of contact sexual violence and stalking. This is indicated by the highly significant age effects (Table 6-6). There is a significant interaction between the frame and age for contact sexual violence and unwanted sexual contact. For RDD, there is a large drop between the youngest age group (18-29) and the next two groups (e.g., 30-44, 45-59). There is a more gradual decline for ABS especially for unwanted sexual contact.

19. For female lifetime victimization (Table 6-7), the relationship between age and victimization is generally negative, with the oldest age group (60+) having the lowest victimization estimates. This relationship is significantly different between the two frames for rape. For RDD, the two younger age groups have very high estimates relative to the older age groups. This pattern is not evident for ABS, where the differences between the younger and older age groups are not as great. Looking more closely at the RDD estimates, almost half of the RDD youngest age group report a lifetime rape. One possible explanation is NRB where the response rate for the two youngest age groups was significantly lower for RDD (see Table 5-1).

20. For male lifetime estimates the relationship is either flat (non-significant) or negative, depending on the age. There are no significant interactions between age and the sample frame.

Comparison of ABS and RDD on Intimate Partners and Consequences of Victimization

Two important characteristics collected on the NISVS is whether the perpetrator is an intimate partner and the consequences of the victimizations reported on the survey.

21. For females, the ABS and RDD do not differ with respect to the average number of perpetrators for each type of Contact Sexual Violence and for Stalking (Table 6-8).

22. For males, the average number of perpetrators is consistently larger for the RDD. However, the differences between RDD and ABS are very small and are not statistically significant.

23. For females, the average percent of perpetrators t*hat are intimate partners are generally lower for the RDD, with two that are statistically significant (Alcohol and drug facilitated rape; Stalking). There is a similar pattern for males, but none of the differences are statistically significant.

24. The RDD respondents were more likely to report each of the consequences included on the survey (Table 6-9). For example, 18.6 percent of the RDD victims report being in fear or concerned for their safety. This compares to 13.9 percent of the ABS victims. A similar pattern occurs for post-traumatic stress disorder (PTSD), mental or emotional harm, physical injuries, seeking help, and missing work or school.

Table 6-1.   Number and percent missing at least one required item for non-intimate partner prevalence measures by survey mode#

	Web				CATI (RDD)				Paper			
	Missing at least one		Lifetime prevalence		Missing at least one		Lifetime prevalence		Missing at least one		Lifetime prevalence	
	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing
Contact sexual violence	170	5.1%	33.6%	35.4%	26	1.8%	45.2%	46.1%	38	20.3%	32.1%	40.3%
Unwanted touching	21	0.6%	27.8%	28.0%	13	0.9%	40.3%	40.6%	7	3.7%	28.2%	29.5%
Coerced sexual contact	31	0.9%	13.5%	13.6%	16	1.1%	15.1%	15.3%	10	5.3%	18.1%	19.3%
Rape	170	5.1%	12.6%	13.3%	24	1.6%	15.2%	15.5%	30	16.0%	16.7%	19.9%
Alcohol and drug facilitated	87	2.6%	6.9%	7.0%	17	1.2%	5.9%	6.0%	10	5.3%	10.8%	11.5%
Physical force	67	2.0%	6.8%	7.0%	12	0.8%	8.8%	8.9%	22	11.8%	12.4%	14.1%
Attempted physical force	53	1.6%	4.8%	4.9%	9	0.6%	8.0%	8.1%	2	1.1%	7.7%	7.7%
Made to penetrate*	91	6.1%	5.4%	5.7%	4	0.6%	6.6%	6.6%	11	14.1%	6.4%	7.5%
Alcohol or drug facilitated	36	2.4%	3.4%	3.4%	4	0.6%	3.4%	3.5%	2	2.6%	5.3%	5.5%
Physical force	20	1.3%	2.2%	2.2%	2	0.3%	1.9%	1.9%	5	6.4%	3.9%	4.2%
Attempted physical force	54	3.6%	1.5%	1.5%	1	0.1%	1.0%	1.0%	9	11.5%	2.7%	3.1%
Stalking	37	1.1%	10.05%	10.17%	51	3.5%	11.91%	12.34%	1	0.5%	7.95%	8.00%

Estimates combine male and females
* Made to penetrate was only asked of males.

Table 6-2. Number and percent missing at least on question asking if incident occurred in the last 12 months for non-intimate partner prevalence measures by survey mode*#

	Web				CATI				Paper			
	Missing at least one date		Lifetime prevalence		Missing at least one date		Lifetime prevalence		Missing at least one date		Lifetime prevalence	
	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing
Contact sexual violence	16	0.5%	5.3%	5.3%	0	0.0%	5.4%	5.4%	5	3.4%	3.2%	3.4%
Unwanted touching	10	0.3%	2.3%	2.3%	0	0.0%	4.0%	4.0%	2	1.2%	2.6%	2.6%
Coercion	3	0.1%	2.6%	2.6%	0	0.0%	2.1%	2.1%	3	1.9%	0.9%	0.9%
Rape	5	0.2%	1.3%	1.3%	0	0.0%	1.1%	1.1%	1	0.6%	0.4%	0.4%
Made to penetrate+	1	0.1%	0.7%	0.7%	0	0.0%	0.5%	0.5%	0	0.0%	0.0%	0.0%
Stalking	0	0.0%	3.0%	3.0%	1	0.1%	3.4%	3.4%	0	0.0%	1.0%	1.0%

* Excludes missing on items to determine lifetime prevalence.
# Estimates combine male and females
+ Made to penetrate was only asked of males

	Web				CATI				Paper			
	Missing at least one		Lifetime prevalence		Missing at least one		Lifetime prevalence		Missing at least one		Lifetime prevalence	
	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing	Number	Percent	With missing	Without missing
Intimate partner violence	127	4.0%	30.3%	31.8%	26	1.8%	39.7%	40.2%	9	5.9%	29.2%	31.4%
Contact sexual violence	36	1.2%	12.7%	12.8%	35	2.4%	14.9%	15.2%	7	4.7%	13.0%	13.6%
Stalking	13	0.4%	5.3%	5.3%	6	0.4%	6.9%	6.9%	0	0.0%	3.1%	3.1%
Intimate physical violence	108	3.3%	22.8%	23.8%	5	0.3%	33.7%	33.8%	5	2.7%	18.9%	19.7%

* Excludes missing on items to determine lifetime prevalence for non-intimate partner prevalence items.
# Estimates combine male and females

			NSFG (2015 - 2017)		
weighted	**RDD**	**ABS**	**Published**	**Alternate 1**	**Alternate 2**
Vaginal intercourse by a male (females only)	29.2%+	12.3%+	18.8%	17.5%	17.0%
Vaginal intercourse by a female (males only)	6.9%*	2.6%*	4.3%	n/a	n/a
Oral or anal sex by a male (males)	3.3%	3.1%	2.3%	n/a	n/a

Table 6-4. Comparison of estimates of forced sex for those age 18-49 for RDD, ABS, and NSFG by sex

n/a – Not applicable. + different from all NSFG estimates at p<.001; *different from NSFG estimate at p<.05.

- Alternate 1 – excludes from published estimate those who said the first involuntary time having sexual intercourse was because of being pressured by words or actions without threats of harm.

- Alternate 2 – excludes from Alternate 1 estimate those who said the first involuntary time having sexual intercourse was because of being given drugs or alcohol.

		Survey					
		RDD		ABS		NSFG	
	Age	Rate	SEE	Rate	SEE	Rate	SEE
Forced vaginal intercourse by a male (females only)	18-29	27.3	5.7	11.1	1.9	17.4	1.3
	30-39	37.3	7.4	11.8	2.3	18.6	1.5
	40-49	21.6	4.8	14.5	2.1	20.7	1.5
Forced vaginal intercourse by a female (males only)	18-29	9.2	3.8	3.2	1.4	4.4	0.5
	30-39	5.9	2.7	2.6	0.8	4.6	0.7
	40-49	3.5	2.6	1.5	0.6	4.1	0.9
Forced oral or anal sex by a male (males)	18-29	3.6	2.3	4.9	1.9	1.8	0.6
	30-39	2.2	2.1	2.6	1.0	1.9	0.5
	40-49	3.7	2.3	1.4	0.5	3.5	1.0

SEE = Standard error of the estimate.

+ The difference between the age groups for RDD are not statistically significant for any of the measures. Differences between age groups for ABS are not statistically significant for any of the measures. Differences between age groups for NSFG are not statistically significant for any of the measures.

Table 6-6.	Non-intimate partner 12-month prevalence estimates by age and sample frame#				
				Significance of*	
	Age	**ABS**	**RDD**	**Age**	**Age by frame interaction**
Contact sexual violence	18-29	13.8	16.6		
	30-44	3.8	3.1		
	45-59	2.4	1.0		
	60+	1.0	2.5		
				<.0001	0.046
Unwanted sexual contact	18-29	5.8	12.0		
	30-44	2.5	2.3		
	45-59	1.3	0.5		
	60+	0.5	2.4		
				<.0001	0.014
Coerced sex	18-29	7.8	7.9		
	30-44	2.2	1.1		
	45-59	0.8	0.4		
	60+	0.1	0.0		
				<.0001	0.449
Rape	18-29	3.3	4.8		
	30-44	0.8	0.2		
	45-59	0.6	0.0		
	60+	0.3	0.0		
				NC	NC
Made to penetrate	18-29	0.6	1.0		
	30-44	0.3	0.0		
	45-59	0.1	0.0		
	60+	0.2	0.0		
				NC	NC
Stalking	18-29	6.4	4.6		
	30-44	2.8	4.2		
	45-59	2.7	3.5		
	60+	0.6	1.5		
				0.001	0.432

Estimates combine male and females
* Significance tests estimated with a logistic regression predicting the victimization estimate with age, sample frame and the interaction between age and the sample frame.

NC – Logistic model could not be estimated because of empty cells in the RDD frame.

	Age	Male ABS	Male RDD	P – Value of* Age	P – Value of* Age by Frame	Female ABS	Female RDD	Significance of Age	Significance of Age by frame
Contact sexual violence	18-29	21.4	34.3			50.7	79.1		
	30-44	20.8	33.7			41.5	68.3		
	45-59	14.9	27.9			42.8	56.4		
	60+	13.9	24.6			35.6	56.6		
				0.011	0.985			0.000	0.155
Unwanted sexual contact	18-29	13.6	28.1			43.3	69.2		
	30-44	14.9	27.5			35.0	63.1		
	45-59	10.9	20.0			34.9	54.0		
	60+	12.2	21.7			28.3	52.9		
				0.145	0.972			0.000	0.604
Coerced sex	18-29	8.8	14.2			25.8	47.8		
	30-44	6.9	10.4			18.5	34.0		
	45-59	4.5	6.9			17.0	23.0		
	60+	3.6	4.6			11.6	14.2		
				0.001	0.972			0.000	0.136
Rape	18-29	7.3	2.6			23.3	48.6		
	30-44	2.3	3.6			18.4	38.7		
	45-59	1.4	6.5			21.9	26.5		
	60+	2.4	4.5			15.5	21.0		
				0.900	0.153			0.000	0.0428
Made to penetrate	18-29	6.7	9.3						
	30-44	6.3	8.6						
	45-59	3.9	5.5						
	60+	3.7	5.9						
				0.236	0.997			n/a	n/a
Stalking	18-29	9.7	7.3			16.1	24.9		
	30-44	4.1	10.1			15.8	33.4		
	45-59	3.9	8.0			16.3	20.8		
	60+	2.9	2.7			5.5	8.9		
				0.033	0.241			0.001	0.379

* Significance tests estimated with a logistic regression predicting the victimization estimate with age, sample frame, and the interaction between age and the sample frame.
n/a – Not applicable. Not asked of female respondents.

Table 6-8.	Average number of perpetrators and average percentage that are intimate partners for different forms of contact sexual violence and stalking by sample frame and sex+					
	Average number of perpetrators			Average % that are intimate partners		
	RDD	ABS	p-value	RDD	ABS	p-value
Female						
Unwanted touching	2.5	2.4	0.368	55.0	58.8	0.309
Alcohol and drug facilitated rape	1.8	1.6	0.200	67.2	84.0	0.048
Physically forced rape	1.6	1.5	0.243	76.0	86.6	0.091
Attempted physically forced rape	1.4	1.5	0.381	92.3	84.7	0.209
Stalking	1.5	1.5	0.342	81.1	91.7	0.022
Male						
Unwanted touching	2.1	1.7	0.065	46.9	73.9	0.104
Alcohol and drug facilitated rape and made to penetrate	2.0	1.6	0.313	100.0	63.1	0.046
Physically forced rape and made to penetrate	2.4	1.5	0.219	67.9	90.0	0.336
Attempted physically forced rape and made to penetrate	2.6	2.2	0.351	51.6	82.9	0.082
Stalking	1.8	1.5	0.156	90.6	89.6	0.396

+ p-value is the result from a two-sample difference of means z-test.

Table 6-9.	Percentage of victims reporting selected consequences#		
		RDD	ABS
Fear or concern for safety		18.6	13.9
Post-traumatic stress disorder		21.5	14.8
Mental or emotional harm		18.8	8.8
Physical injuries		21.5	14.6
Seeking help		13.7	8.0
Missing work or school		9.8	4.9

\# Estimates combine male and females

7. How Do ABS and RDD Compare on Burden, Privacy, and Confidentiality?

Summary

➢ It took more than twice as long to complete the RDD survey than the web survey. The median time to complete was 38 minutes and 15 minutes for the RDD and web, respectively. The time to complete the survey varied significantly by the number of different types of victimizations that were reported.

➢ Some of the difference in time to complete the survey is related to reading the questions out loud by an interviewer (RDD) versus reading silently (web). It may also reflect web respondents not reading the entire question or skimming items.

➢ Despite the shorter time it took to fill out the survey, a larger percentage of ABS vis-à-vis RDD respondents reported the survey was burdensome and too long. Significantly more RDD respondents reported being willing to take the survey again.

➢ About one-fifth of RDD respondents said they experienced intense emotions and one-third agreed the survey caused them to think of unwanted things. Reactions by ABS respondents was very similar.

➢ About one-fifth of the ABS respondents took the survey when someone was in the room. Someone in the household knew what the survey was about for about one-fifth of ABS respondents. RDD respondents reported having somewhat more privacy on both of these measures.

➢ About one-quarter of the web respondents filled out the survey on a smartphone. Most of the characteristics reviewed in this chapter (e.g., perceptions of burden, privacy, etc.) did not differ between those who filled out the survey on a PC/tablet or a smartphone.

This chapter compares the ABS and RDD surveys with respect to respondent perceptions of burden, privacy, and confidentiality.

Measures of Burden

1. A total of 1,461 RDD extended surveys were completed (Table 7-1). The median time to complete the survey was 38 minutes. The mean time was considerably higher (46 minutes), indicating there were some respondents who took a very long time to complete. The maximum time to complete was more than 3 hours. The minimum was 9 minutes.

2. A total of 3,306 respondents completed the survey on the web. The median time to complete was 15 minutes with a mean of 18 minutes.

3. Approximately 72 percent of the web surveys were completed on either a PC or a tablet. Twenty-eight percent were completed on a smart phone.

4. The longer time to complete with a telephone interview partly reflects the ability of most individuals to be able to read faster than interviewers can read the questions. Some of the difference may also be due to web respondents not reading or skimming questions/introductions.

5. As one might expect, the time to complete the survey was affected by the number of different types of victimizations reported (Table 7-2). Those that did not report any victimizations had a median time to complete of 30 and 12 minutes for RDD and web respondents, respectively. The time to complete increases with each additional type of victimization that is reported. For three or more different types of victimizations, the RDD had a median time of 50 minutes compared to 25 minutes for the web.

After completing the survey, respondents were asked several questions related to the burden of the survey (Table 7-3).

6. Across all three measures of burden, RDD respondents found the survey less burdensome. They were less likely than ABS respondents to say it was at least a little burdensome (39.8% vs. 46.2%; p<.001), that it was too long (27.2% vs 36.5%; p<.001) and more likely to say they would do the survey again (68.7% vs. 52.3%; p<.001).

7. When compared to the nonvictims, the RDD victims are slightly more likely to report that the survey was at least "a little burdensome" (41.4% vs. 36.7%; p<.20) and the survey was too long (28.7% vs. 24.5%; p<.05). However, 10 percent more victims said they were willing to take the survey again when compared to non-victims (72.6% vs. 61.5%; p<.001).

8. The ABS victims and nonvictims were somewhat different. There was very little difference in the percentage who said it was at least "a little burdensome" (46.8% vs. 45.6%). Somewhat more of the nonvictims said the survey was too long (33.1% vs. 39.9%; p<.001). As with the RDD sample, 10 percent more victims said they were willing to take the survey again when compared to the nonvictims (57.9% vs. 46.8%; p<.001).

9. For the ABS respondents, there was some variation in these measures by the mode. Fewer of those using the call-in computer-assisted telephone interviewing (CATI) when compared to the web thought the survey was burdensome (27.6% CATI vs. 45.8% web; p<.06) and that it

took too long (21.9% CATI vs. 36.8% web; p<.10). A very high percentage of these individuals said they would do the survey again (81.2% CATI vs. 52.2% for web). It should be noted that these CATI respondents voluntarily called in to do the survey, indicating they were highly motivated to do it. For those filling out paper survey (n=187), a higher percentage than the web respondents say the survey was burdensome (55.4% paper vs. 45.8% web; p<.03) and that it was too long (49.6% paper vs 35.8% web; p<.01).

Perceived Sensitivity

10. When asked about the sensitivity of the survey (Table 7-4) about one-fifth of the RDD respondents (19.6%) said they experienced intense emotions and about one-third (33.2%) agreed that the survey caused them to think of things they did not want to think about. These results are very similar for the ABS respondents.

11. More RDD than ABS respondents agreed that participating on the survey helps others (65.3% vs. 44.4%; p<.001).

12. A larger percentage of the victims said the survey brought out intense emotions and caused them to think about things they did not want to think about. For example, more RDD victims were likely to say they experienced intense emotions than nonvictims (24.4% vs. 10.7%; p<.001).

Privacy Conditions Surrounding the Survey Administration

13. Most of the respondents completed the survey at home. Slightly fewer of the RDD respondents did it at home than ABS respondents (RDD 78.5%; ABS 83.2%; p<.01).

14. The ABS survey was done in slightly less private conditions. More ABS respondents did the survey when someone else was in the room (22.3% vs. 13.2%; p<.001). More ABS respondents had someone in the room for at least 5 minutes (14.5% vs. 7.5%; p<.001). Slightly more ABS respondents had a spouse or partner in the room (11.6% vs. 8.0%; p<.01).

15. More ABS respondents reported that someone in the household knew what the survey was about (19.0% vs. 7.8%; p<.001).

16. None of the above characteristics vary much by whether the respondent reported a victimization.

	RDD	Web		
		Total	PC or tablet	Smartphone
Number of completes	1461	3306	2389	917
Percent of completes	100%	100%	72%	28%
Timing in minutes				
Median	38	15	15	15
Mean	46	18	18	17
Minimum	9	4	4	4
5th percentile	23	7	7	7
25th percentile	30	11	11	11
75th percentile	56	21	21	21
95th percentile	96	37	37	38
Maximum	196	111	111	85
Standard deviation	25	10	10	11

Estimates combine male and females

Table 7-2.	Median length of ABS and RDD surveys by number of victimizations reported and type of device for the web (minutes)#				
			Web		
	RDD	Total	PC or tablet	Smartphone	
No victimizations reported	30	12	13	11
Only one type of victimization reported	34	14	15	14
Two different types of victimizations reported	38	18	18	17
Three or more different types of victimizations reported	50	25	25	25

Estimates combine male and females

	RDD	ABS					
		Total	**Web**			**Paper**	**CATI**
			Total	**PC/Tablet**	**Smartphone**		
Overall							
Survey was very burdensome, somewhat burdensome, or a little burdensome	39.8	46.2	45.8	46.4	44.4	55.4	27.6
Survey was too long	27.2	36.5	35.8	34.3	38.9	49.6	21.9
Strongly agree or agree to making the same choice to participate	68.7	52.3	52.2	52.0	52.6	49.3	81.2
Among respondents who reported at least one victimization							
Survey was very burdensome, somewhat burdensome, or a little burdensome	41.4	46.8	46.5	48.3	43.3	55.9	37.9
Survey was too long	28.7	33.1	32.8	31.2	35.6	42.4	25.7
Strongly agree or agree to making the same choice to participate	72.6	57.9	57.5	56.3	59.7	61.0	79.7
Among respondents who did not report any victimization							
Survey was very burdensome, somewhat burdensome, or a little burdensome	36.7	45.6	45.0	44.7	45.8	55.2	11.0
Survey was too long	24.5	39.9	38.9	37.3	42.7	53.4	15.8
Strongly agree or agree to making the same choice to participate	61.5	46.8	46.9	48.0	44.4	43.1	83.5

Estimates combine male and females

Table 7-4. Perceived sensitivity of the survey for RDD and ABS by type of web device (percent)#

		ABS					
			Web				
	RDD	**Total**	**Total**	**PC/Tablet**	**Smartphone**	**Paper**	**CATI**
Overall							
Agreed or strongly agreed about experiencing intense emotions	19.6	19.3	19.1	18.1	21.2	20.2	27.1
Agreed or strongly agreed about survey causing things you did not want to think about	33.2	32.1	32.4	31.1	35.3	28.1	25.5
Agreed or strongly agreed about helping others by participating in the survey	65.3	44.4	44.3	42.2	48.7	41.9	74.1
Reported questions being very sensitive or somewhat sensitive	56.2	49.5	52.9	55.1	48.3	n/a	36.0
Among respondents who reported at least one victimization							
Agreed or strongly agreed about experiencing intense emotions	24.4	26.3	26.4	25.4	28.2	26.1	16.3
Agreed or strongly agreed about survey causing things you did not want to think about	41.0	42.9	43.0	41.8	45.1	47.0	18.1
Agreed or strongly agreed about helping others by participating in the survey	69.8	51.5	51.2	48.5	56.0	54.7	71.3
Reported questions being very sensitive or somewhat sensitive	61.7	54.4	56.9	58.4	54.2	n/a	49.0
Among respondents who did not report any victimization							
Agreed or strongly agreed about experiencing intense emotions	10.7	12.3	11.7	11.2	12.9	17.1	44.5
Agreed or strongly agreed about survey causing things you did not want to think about	18.8	21.5	21.7	20.9	23.6	18.0	37.3
Agreed or strongly agreed about helping others by participating in the survey	56.8	37.5	37.4	36.3	40.0	35.0	78.7
Reported questions being very sensitive or somewhat sensitive	46.1	44.7	48.7	52.0	41.3	n/a	15.0

n/a = Not applicable. Not asked.
Estimates combine male and females

Table 7-5.	Privacy conditions when taking the survey for RDD and ABS by type of web device (percent)#						
			ABS				
			Web				
	RDD	**Total**	**Total**	**PC/Tablet**	**Smartphone**	**Paper**	**CATI**
Overall							
At home	78.5	83.2	88.4	89.3	86.6	n/a	98.2
Someone else in room	13.2	22.3	23.1	19.3	31.0	12.0	5.4
Someone sitting or standing in the room for at least 5 minutes	7.5	14.5	15.5	11.9	22.8	n/a	5.4
Spouse or partner in room	8.0	11.6	12.4	10.4	16.4	n/a	5.4
Someone else in household knew what the survey was about	7.8	19.0	20.4	19.3	22.7	n/a	0.0
Among respondents who reported at least one victimization							
At home	80.1	84.2	87.8	87.6	88.0	n/a	100.0
Someone else in room	14.4	22.7	23.0	19.5	29.4	19.7	8.7
Someone sitting or standing in the room for at least 5 minutes	7.8	14.9	15.6	11.5	23.3	n/a	8.7
Spouse or partner in room	8.3	12.4	13.0	10.9	16.8	n/a	8.7
Someone else in household knew what the survey was about	7.0	16.8	17.8	16.1	20.9	n/a	0.0
Among respondents who did not report any victimization							
At home	75.6	82.2	89.1	90.9	84.9	n/a	95.2
Someone else in room	11.1	21.9	23.2	19.1	32.8	7.8	0.0
Someone sitting or standing in the room for at least 5 minutes	7.2	14.1	15.4	12.4	22.3	n/a	0.0
Spouse or partner in room	7.3	10.8	11.7	9.9	15.9	n/a	0.0
Someone else in household knew what the survey was about	9.1	21.1	23.1	22.3	24.8	n/a	0.0

n/a = Not applicable. Not asked.
Estimates combine male and females

8. Results of ABS experiments

Several different field experiments were carried out on the field test. For the address-based sample (ABS), samples were randomly assigned to two different respondent selection methods. One was a probability method and the other a non-probability method. The second experiment varied the mode offered to non-respondents for the extended interview.

Respondent Selection

An experiment was conducted that compared two different respondent selection procedures. One was a full probability method using the Rizzo-Brick-Park (RBP) method.[22] This is the same method used on the current National Intimate Partner and Sexual Violence Survey (NISVS) computer-assisted telephone interviewing (CATI) survey. This procedure is a full probability method that gives all adults in the household a known non-zero chance of selection. When the first request was sent to the household, the respondent was asked to go to the web to complete the household screener. The screener included selecting an adult for the extended survey using the RBP method. The screener respondent was instructed to ask the identified individual to take the extended survey. If no screener is filled out, a reminder postcard was sent to the household. For those still not responding after the postcard, a third mailing, sent express delivery, was sent that asked someone in the household to complete the screener on the web, but also included a paper version of the screener that asked for a roster of adults. If the paper screener was returned, an individual was randomly selected by project staff. The selected individual was sent a request to fill out the survey.

The second experimental condition was a non-probability method using the Youngest Male Oldest Female (YMOF) approach.[23] This method puts an emphasis on collecting data from younger people and from males. These two groups are traditionally under-represented in ABS mail surveys. To implement the YMOF, households were randomly assigned to one of four groups: (1) youngest adult male, (2) youngest adult female, (3) oldest adult male, and (4) oldest adult female. Households are randomly assigned to one of these groups. The screener respondent is asked to select that particular individual (e.g., youngest adult male) as the respondent. If the household does not fit the particular profile (e.g., no adult males in household), additional instructions are given to select a respondent. For example, for the youngest adult male group (Group 1), if there is only one adult male in the household, the respondent is told to select that individual. If there are no adult males, then the person is asked to select the youngest adult female. This is not a probability method because there are a small number of individuals who are not given a chance of selection. In particular, in households that have three or more adults of the same sex, those in the middle age group cannot be selected. The base-weight is computed using the reported number of adults in the household. Nonresponse and post stratification (or raking) adjustments are made to adjust for coverage and nonresponse. The assumption is that this last adjustment compensates for the

---

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

Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys | Public Opinion Quarterly | Oxford Academic (oup.com)

[23] Yan, T., Tourangeau, R. and R. McAloon (2015) A Meta-analysis of Within-Household Respondent Selection Methods on Demographic Representativeness. Paper presented at the 2015 Federal Committee on Survey Methodology Research Conference,

https://nces.ed.gov/FCSM/pdf/H3_Yan_2015FCSM.pdf

relatively small number of households that have three or more same-sex adults. According to the American Community Survey (ACS), approximately 3.9% of the households have this particular age-sex profile.

The results of the experiment include:

1. There was no difference between the response rates for the two methods. The response rate for the YMOF was 33.8 percent and the response rate for RBP was 32.4 percent (data not shown).

2. Table 8-1 compares the demographic distribution from the two selection methods with the ACS. The distributions were estimated using weights that reflect the respondent's probability of selection and do not include any adjustments for nonresponse or raking to the national population. There were no significant differences between the methods with respect to distributions by demographic groups, including age, sex, marital status, race, education, income, born in the United States, access to the internet, and home ownership (Table 8-1). Of particular note, the YMOF did not have a significant effect on the proportion of young people and the proportion of males completing the survey. This may be because most of the prior studies were done with surveys that did not use the web as a response mode. For at least age, the web may bring in more young people than more traditional modes, such as paper mail surveys. It isn't clear why the method did not boost the proportion of females.

3. The lifetime and 12-month prevalence estimates do not differ by the method used to select a respondent (Tables 8-2 to 8-5). The one exception is for the lifetime and 12-month prevalence for stalking by an intimate partner among females. The non-probability method has a rate that is significantly higher than the probability method (lifetime prevalence 9.9 vs. 5.9; p<.009). It is not clear how to interpret this difference, given the number of comparisons among these tables, one would expect 1 or 2 to be significant by chance.

Overall, there is not a discernable difference between the probability and non-probability method, either with respect to response rates, demographic distributions or victimization estimates.

ABS Experiments – Mode Choice

This experiment offered an alternative mode to nonrespondents to complete the extended survey. If the selected respondent did not complete the survey in response to the initial request to do the web survey, a second request was sent that gave the respondent the opportunity to complete the survey with one of two alternative modes. Half the non-respondents were told they could complete an abbreviated paper version of the survey, while the other half was told they could call in and complete the survey with a telephone interviewer. In both cases, the respondent was still be able to complete the survey by web.

The paper version was offered because prior research has found that respondents are more likely to fill out a paper survey than other possible modes (web, telephone).[24] With respect to response rates, paper and pencil instruments tend to yield the highest response rates.[25] Mixing web and

---

[24] Montaquila, J.M., Brick, J.M., Williams, D., Kim, K., and Han, D. (2013). A study of two-phase mail survey data collection methods. *Journal of Survey Statistics and Methodology*, 1(1), 66–87.

[25] Messer, B.L., and Dillman, D.A. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly,* 75(3), 429–457.

paper surveys can also yield high response rates, although even here a significant number of respondents end up using the paper instrument. For example, without offering a bonus to complete the survey on the web, approximately two-thirds of respondents complete by paper in mixed-mode surveys.[26] In order to encourage use of the web, the respondent was offered $15 to complete by web and $5 to complete the paper survey.

The second condition for this experiment is an inbound CATI interview. Respondents were given the opportunity to call into an 800 number to complete the CATI version of the survey. One of the disadvantages of the paper survey described above is that it requires simplifying the NISVS instrument. Without a computer to drive the skip patterns, the paper instrument did not collect as much of the detail as the CATI or web surveys. For example, it was not possible to collect data on each of the different perpetrators for each type of victimization. Testing an inbound CATI option has the advantage of collecting all of the NISVS data, rather than an abbreviated version as done on the paper version. The inbound CATI also provides an extra layer of privacy from other household members when taking the survey relative to the web or a paper version.

The inbound line was staffed to cover (all times are Eastern Standard) 9 am to 12 am Monday through Friday, 10 am to 6 pm on Saturday and 2 pm to 10 pm on Sunday. Respondents calling outside these hourswere asked to leave a voicemail and the best time for a call back.

1. The response rates for the web/paper treatment was higher than the web/CATI group (Table 8-6). Of the requests that offered the web/paper option, 27.8 percent completed the main NISVS survey compared to 21.8 percent of the requests for the web/CATI group.

2. This difference in response rate resulted in approximately 61 more surveys for the web/paper option when compared to the web/CATI option. This represents approximately one percentage point difference in the overall response rate between the two options.

3. More respondents used the paper survey than the call-in CATI. Of the requests with the web/paper option, 187 filled out the paper survey and 158 used the web (unweighted). Or about 55 percent of those responding to the web/paper group used the paper. Of the requests involving the web/CATI option, 33 completed by CATI and 251 did it by web. Or about 12 percent of the completes in the web/CATI group were from CATI.

4. With respect to the types of respondents who used the alternative modes (Table 8-7):

    - The web respondents are very close to the national profile, as indicated by the ACS. The difference between the web and the ACS for particular age groups are statistically significant, but the actual proportions are very close (e.g., with 1 or 2 percentage points). For example, the proportion of web resondents 18-29 was 2.3 percentage points higher than the ACS. Those completing by either paper or CATI were much older than those filling the survey out by the web. The proportion of respondents that were 60 and over was 64.9 percent and 56.0 percent for the paper and CATI, respectively. This compares to 25 percent for the web.

    - The web was much less representative of the national population (ie ACS) for income, born in the US and access to internet.

---

[26] Messer, B.L., and Dillman, D.A. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly,* 75(3), 429–457.

- Those completed by paper were more likely to be in the lower education group (55.8% paper, 43.3% CATI, 35.0% web).

- The paper and CATI respondents were less likely to have access to the internet (99.1% for web, 69.3% for paper, and 62% for CATI).

To assess the impact of the alternative mode choices on key outcomes, victimization estimates were calculated using two alternative design assumptions. One is a design that used the web/paper option group. To simulate this, weights were developed for the full sample but excluded the web/CATI group. The second design is one which no alternative modes are used. To simulate this, weights were developed for those that just responded by the web. This includes all web responses regardless of which of the mode-choice group the respondent may have been assigned to. The victimization estimates were calculated for these two groups and compared to the estimates for the full sample (Tables 8-8 to 8-11).

Comparing the full sample to the web/paper group provides an indication of the effect of using the web/paper option group relative to the web/CATI group. If the web/paper group is less than the full sample this is an indication that the web/paper group produces estimates that are lower than the web/CATI option group, since the latter is included in the full sample but not the web/paper option estimate. As noted below, to assess the significance of the difference between the two option groups, we compared them using the full sample.

Testing for statistical significance between these estimates is not straightforward, since a large percentage of the sample contribute to each of the estimates. For example, approximately 80 percent of the full sample responded before being offered an alternative mode. These same respondents are part of the estimates calculated for the web/paper design. When comparing the full sample to the web/paper design, statistical significance was estimated using a logistic regression for the full sample. The regression predicted a dictomous variable coded as '0' for those not victimized and '1' for those that were victimized. The predictor in the regression included a three-category variable: 1) responded prior to when non-respondents were assigned to one of the two optional mode groups an optional of mode, 2) responded in the web/CATI group, and 3) responded in the web/paper group. If the difference between groups two and three was significant, the estimates were marked as being statistically significant. The discussion below concentrates on the sex-specific estimates.

5. For the lifetime prevalence estimates (Table 8-9):

- Males. The difference between the web/paper and full sample are very small, ranging in absolute value from 0 to .5 percentage point. The biggest difference (.5) is for intimate partner psychological aggression (IPPA) (29.2% full sample vs. 28.7% web/paper p<.05).

- Females. The differences between the web/paper and full sample range in absolute value from .1 to 1 percentage point. The web/paper is consistently lower than the full sample. The three largest differences are for unwanted sexual touching (35.0% vs. 34.1%; p<.05); rape (19.7% vs. 18.8%; p<.05); and contact sexual violence by an intimate partner (17.9% vs. 17.0%; p<.05).

6. For 12-month prevalence estimates (Table 8-11):

- Males. The differences range in absolute value from 0 to .2. The web/paper is lower than the full sample in most cases. The differences for intimate partner violence (IPV) (2.8 vs. 2.6; p<.05) and intimate partner physical violence (IPPV) (2.0 vs. 1.9; p<.05) are statistically significant.

- Females. The differences range in absolute value from .1 to .8. Four of the differences are statistically significant, sexual coercion (4.3 vs. 3.7; p<.05); contact sexual violence by an intimate partner (3.9 vs. 3.3; p<.05); IPPV (3.2 vs. 2.4; p<.05); and IPPA (6.1 vs. 5.6; p<.05). For all of the significant differences, the full sample is greater than the web/paper.

Overall, these results suggest that if the web/paper option were used, the estimates would be slightly lower when compared to a design that used web/CATI design. It is not clear why the web/paper group tends to produce somewhat lower estimates of victimization. One explanation is that it brings in respondents who tend to have lower victimization estimates. This is indicated by the higher proportion of older respondents that the web/paper option brought in. The other explanation is that it is related to the mode used to do the survey. However, the bulk of the web/CATI surveys were completed by web, which is generally considered comparable to the paper survey. Both are self-administered. This *suggests* that the differences are related to who is responding, rather than the mode. It is also a possible that the abbreviated paper survey led to lower estimates, but it isn't clear why the abbreviated survey would lead to this.

The estimates that approximate just using the web are even closer to the full sample discussed above. This is logical because only about five percent of the respondents used an alternative mode. This suggests that if no alternative mode were offered, the estimates would be similar to what has been presented in prior chapters.

Table 8-1.	Demographic distributions for ABS sample for probability and non-probability respondent selection methodsX and the American Community Survey*#		
	Probability	**Non-probability**	**Benchmark (ACS)**
Age			
18-29	18.8	20.1	21.5
30-44	26.4	25.0	25.1
45-59	24.2	24.7	25.7
60+	30.6	30.2	27.7
Sex			
Male	46.6	46.7	48.7
Female	53.4	53.3	51.3
Marital status			
Married	56.7	57.8	50.4
Never married	19.8	20.1	30.0
Other	23.6	22.0	19.5
Race			
Hispanic	11.9	12.4	15.7
NH-White	71.2	71.4	64.1
NH-Black	6.5	6.7	12.0
NH-Multiracial	3.1	2.2	1.7
NH-Other	7.4	7.3	6.6
Education			
High school or less	21.1	21.1	36.4
Some college	34.7	33.9	34.8
Bachelor's or higher	44.2	45.0	28.9
Income			
Less than $25,000	17.2	17.4	15.7
$25,000 - $49,999	19.3	17.8	20.0
$50,000 - $74,999	16.4	15.4	18.3
$75,000+	47.1	49.4	46.0
Born in U.S.			
Yes	87.8	87.0	81.6
No	12.2	13.0	18.4
Access to internet not including through cellphone			
Yes	94.1	94.3	75.6
No	5.9	5.7	24.4
Any access to internet			
Yes	98.2	98.9	87.4
No	1.8	1.1	12.5
Home ownership			
Owned	72.5	74.4	66.8
Rented	26.7	24.5	31.6
Other	0.8	1.1	1.6

* Unweighted with an adjustment for the probability of selection. For ABS this adjusts for the selection for the nonresponse bias follow-up. For RDD this adjustment has three components:1) for the landline sample an adjustment for selection within the household, 2) for both landline and cell phone frames an adjustment for selection for the Phase 2

nonresponse follow-up and 3) for both frames an adjustment for the probability of selection into the landline or cell phone frame.

American Community Survey; # Estimates combine male and females

X Probability Method = Rizzo-Brick-Park; Non-probability = Youngest Male, Oldest Female

Table 8-2.	Lifetime prevalence estimates for ABS sample for selected measures of sexual, physical, and emotional abuse by respondent selection methodX+#		
	Probability	Non-probability	p-value
Contact sexual violence	29.1	31.6	0.158
Unwanted touching	22.6	26.1	0.055
Sexual coercion	12.0	12.3	0.387
Rape	11.6	11.9	0.388
Made to penetrate*	2.5	2.3	0.4
Stalking	8.3	10.4	0.083
Intimate partner violence	28.4	28.8	0.389
Contact sexual violence by intimate partner	11.0	12.6	0.175
Stalking by intimate partner	3.7	6.5	0.002
Intimate partner physical violence	23.4	22.1	0.317
Intimate partner psychological aggression	32.2	33.9	0.277

X Probability Method = Rizzo-Brick-Park; Non-probability = Youngest Male, Oldest Female

Estimates combine male and females

+ p-value is the result from a two-sample difference of means z-test.

* Made to penetrate was only asked of males

	Probability	Non-probability	p-value
Table 8-3. Twelve-month prevalence estimates for ABS sample for selected measures of sexual, physical, and emotional abuse by respondent selection method x+#			
Contact sexual violence	4.8	4.8	0.399
Unwanted touching	2.2	2.4	0.376
Sexual coercion	2.7	2.4	0.375
Rape	1.2	1.1	0.382
Made to penetrate*	0.5	0.1	0.042
Stalking	2.5	3.2	0.252
Intimate partner violence	4.5	5.3	0.293
Contact sexual violence by intimate partner	2.7	2.1	0.274
Stalking by intimate partner	0.5	1.6	0.011
Intimate partner physical violence	2.4	2.9	0.331
Intimate partner psychological aggression	4.9	4.9	0.397

x Probability Method = Rizzo-Brick-Park; Non-probability = Youngest Male, Oldest Female
+ p-value is the result from a two sample difference of means z-test.
# Estimates combine male and females
* Made to penetrate was only asked of males

Table 8-4.	Lifetime prevalence estimates For ABS sample for selected measures of sexual, physical, and emotional abuse by respondent selection method[X] by sex[+]					
	Males			Females		
	Probability	Non-probability	p-value	Probability	Non-probability	p-value
Contact sexual violence	16.3	18.6	0.273	40.5	43.9	0.168
Unwanted touching	11.6	14.1	0.208	32.4	37.4	0.075
Sexual coercion	5.5	6.1	0.356	17.7	18.1	0.393
Rape	2.9	3.3	0.372	19.3	20.0	0.376
Made to penetrate	5.4	4.7	0.350	0.0	0.0	
Stalking	4.6	5.5	0.307	11.7	15.0	0.080
Intimate partner violence	21.1	20.2	0.372	34.9	37.0	0.288
Contact sexual violence by intimate partner	4.6	5.7	0.277	16.6	19.2	0.196
Stalking by intimate partner	1.1	2.8	0.059	5.9	9.9	0.009
Intimate partner physical violence	19.5	17.3	0.260	26.9	26.7	0.398
Intimate partner psychological aggression	28.9	29.5	0.388	35.2	38.1	0.222

[X] Probability Method = Rizzo-Brick-Park; Non-probability = Youngest Male, Oldest Female
[+] p-value is the result from a two-sample difference of means z-test.

Table 8-5.	Twelve-month prevalence estimates for ABS sample for selected measures of sexual, physical, and emotional abuse by respondent selection method[X] by sex[+]					
	Males			Females		
	Probability	Non-probability	p-value	Probability	Non-probability	p-value
Contact sexual violence	2.3	1.9	0.363	7.1	7.6	0.383
Unwanted touching	0.8	1.5	0.238	3.5	3.4	0.396
Sexual coercion	0.7	0.4	0.278	4.3	4.3	0.398
Rape	0.7	0.1	0.079	1.7	2.1	0.350
Made to penetrate	1.1	0.2	0.040	n/a	n/a	n/a
Stalking	2.0	1.8	0.385	3.0	4.7	0.166
Intimate partner violence	2.6	3.0	0.367	6.1	7.6	0.290
Contact sexual violence by intimate partner	0.8	0.5	0.304	4.3	3.5	0.329
Stalking by intimate partner	0.2	0.5	0.253	0.8	2.7	0.014
Intimate partner physical violence	1.6	2.4	0.270	3.1	3.3	0.392
Intimate partner psychological aggression	3.2	3.9	0.320	6.3	5.9	0.388

[X] Probability Method = Rizzo-Brick-Park; Non-probability = Youngest Male, Oldest Female
[+] p-value is the result from a two-sample difference of means z-test.
n/a – Not applicable. Not asked of female respondents.

	Web/paper			Web/CATI		
	Total	**Phase 1**	**NRFU***	**Total**	**Phase 1**	**NRFU***
Requests	1,241	911	330	1,300	925	375
Complete	345	273	72	284	205	79
Web	158	113	45	251	174	77
Alternative	187	160	27	33	31	2
Percent complete	27.8%	30.0%	21.8%	21.8%	22.2%	21.1%

* Nonresponse follow-up. A subsample of nonrespondents were mailed a request to complete the survey after being offered a higher incentive.
\# Estimates combine male and females

		Web	Paper	CATI	ACS	p-value For web v ACS
	Age					
18-29		22.8	4.6	1.8	21.5	.000
30-44		26.7	4.0	6.7	25.1	.000
45-59		25.6	26.5	35.5	25.7	.361
60+		25.0	64.9	56.0	27.7	.000
	Sex					
Male		48.0	46.5	46.1	48.7	.144
Female		52.0	53.5	53.9	51.3	.144
	Marital status					
Married		50.8	50.7	24.3	50.4	.210
Never married		30.5	21.1	36.8	30.0	.107
Other		18.8	28.2	38.9	19.5	.009
	Race					
Hispanic		16.0	12.5	7.0	15.7	.038
NH-White		64.1	65.9	49.8	64.1	.396
NH-Black		11.5	16.1	36.3	12.0	.102
NH-Multiracial		1.7	1.2	3.9	1.7	.363
NH-Other		6.6	4.2	3.0	6.6	.392
	Education					
High school or less		35.0	55.8	43.4	36.4	.000
Some college		34.9	32.1	38.5	34.8	.360
Bachelor's or higher		30.1	12.1	18.1	28.9	.000
	Income					
Less than $25,000		22.3	37.1	53.4	15.7	.000
$25,000 - $49,999		20.8	34.6	21.3	20.0	.255
$50,000 - $74,999		15.9	11.0	6.1	18.3	.003
$75,000+		40.9	17.3	19.3	46.0	.000
	Born in U.S.					
Yes		86.4	84.0	98.3	81.6	.000
No		13.6	16.0	1.7	18.4	.000
	Access to internet with no cell data plan					
Yes		93.9	69.3	62.0	75.6	.000
No		6.1	30.7	38.0	24.4	.000
	Access to internet					
Yes		99.1	84.9	74.2	87.4	.000
No		0.9	15.1	25.8	12.5	.000
	Home ownership					
Owned		68.7	73.1	44.3	66.8	.015
Rented		30.3	24.8	50.4	31.6	.083
Other		1.0	2.1	5.3	1.6	.003

Table 8-7. Demographic distributions for ABS sample by mode of response*#

* Unweighted data with an adjustment for the probability of selection. For ABS this adjusts for the selection within the household and for the nonresponse bias follow-up. For RDD this adjustment has three components: 1) for the landline

sample an adjustment for selection within the household, 2) for both landline and cell phone frames an adjustment for selection for the Phase 2 nonresponse follow-up, and 3) for both frames an adjustment for the probability of selection into the landline or cell phone frame.

Except where noted, estimates combine male and females

	Full sample	With web/ paper option	Web only
Contact sexual violence	30.4	29.9	30.3
Unwanted touching	24.4	24.0	24.3
Sexual coercion	12.1	11.7	11.8
Rape	11.7	11.2[x]	11.5
Made to penetrate	2.4	2.4	2.5
Stalking	9.4	9.3	9.4
Intimate partner violence	28.6	28.4	28.8
Contact sexual violence by intimate partner	11.8	11.2[x]	11.7
Stalking by intimate partner	5.1	4.9	5.1
Intimate partner physical violence	22.7	22.6	22.9
Intimate partner psychological aggression	33.1	32.6[x]	33.6

Table 8-8. Lifetime prevalence estimates for ABS sample for selected measures of sexual, physical, and emotional abuse for full sample, web/paper option group and those completing by the web*#

* Full Sample = Includes all who participated on the feasibility study and weighted to the national population; Web/Paper Option = Includes all who participated before respondents were given the choice of modes plus those randomly assigned to the mode/paper choice group. Weighted to national population; Web only = Includes those who participated using the web for the full sample, including both groups given a choice of mode. Weighted to the national population.
# Estimates combine male and females
[x] Significance estimated by comparing those in the Web/Paper and Web/CATI groups. Difference is significant at p<.05. See text on how test was computed

Table 8-9. Lifetime prevalence estimates for ABS sample for selected measures of sexual, physical, and emotional abuse for full sample, web/paper option group and those completing by the web by sex*

	Males			Females		
	Full sample	With web/paper	Web-only	Full sample	With web/paper	Web-only
Contact sexual violence	17.5	17.7	17.7	42.2	41.2	41.9
Unwanted touching	12.9	13.1	12.9	35.0	34.1[x]	34.7
Sexual coercion	5.8	5.6	5.8	17.9	17.3	17.3
Rape	3.1	3.0	3.3	19.7	18.8[x]	19.1
Made to penetrate	5.0	4.9	5.1	0.0	0.0	0.0
Stalking	5.0	4.8	5.0	13.4	13.4	13.3
Intimate partner violence	20.6	20.7	20.8	36.0	35.4	36.0
Contact sexual violence by intimate partner	5.2	5.0	5.1	17.9	17.0[x]	17.8
Stalking by intimate partner	2.0	1.7	1.9	7.9	7.9	8.0
Intimate partner physical violence	18.3	18.4	18.6	26.8	26.6	26.9
Intimate partner psychological aggression	29.2	28.7[x]	29.9	36.7	36.2	37.0

* Full Sample= Includes all who participated on feasibility weighted to the national population; With Web/Paper Option = Includes all who participated before respondents were given the choice of modes plus those randomly assigned to the mode/paper choice group. Weighted to national population; Web only = Includes those who participated using the web. Weighted to the national population.

[x] Significance estimated by comparing those in the Web/Paper and Web/CATI groups. Difference is significant at p<.05. See text on how test was computed

	Full sample	With web/ paper option	Web only
Table 8-10. Twelve-month prevalence estimates for ABS sample for selected measures of sexual, physical and emotional abuse for full sample, web/paper option group and those completing by the web*#			
Contact sexual violence	4.8	4.6	4.8
Unwanted touching	2.3	2.2	2.2
Sexual coercion	2.5	2.2^x	2.5
Rape	1.2	1.1	1.2
Made to penetrate$^+$	0.3	0.2	0.3
Stalking	2.9	2.8	2.9
Intimate partner violence	4.9	4.4	5.0
Contact sexual violence by intimate partner	2.4	2.0^x	2.4
Stalking by intimate partner	1.1	1.1	1.0
Intimate partner physical violence	2.6	2.1^x	2.7
Intimate partner psychological aggression	4.9	4.6^x	4.9

*Full Sample = Includes all who participated on feasibility weighted to the national population; With Web/Paper Option = Includes all who participated before respondents were given the choice of modes plus those randomly assigned to the mode/paper choice group. Weighted to national population; Web only = Includes those who participated using the web. Weighted to the national population.

\# Estimates combine male and females

+ Made to penetrate was only asked of males

^{x} Significance estimated by comparing those in the Web/Paper and Web/CATI groups. Difference is significant at p<.05. See text on how test was computed.

Table 8-11. Twelve-month prevalence estimates for ABS sample selected measures of sexual, physical, and emotional abuse for full sample, web/paper option group and those completing by the web by sex*

	Males			Females		
	Full sample	With web/paper	Web-only	Full sample	With web/paper	Web-only
Contact sexual violence	2.1	2.1	2.1	7.4	6.8	7.3
Unwanted touching	1.1	1.2	1.1	3.4	3.1	3.2
Sexual coercion	0.6	0.5	0.6	4.3	3.7[x]	4.2
Rape	0.4	0.3	0.4	1.9	1.8	1.9
Made to penetrate	0.6	0.5	0.7	n/a	n/a	n/a
Stalking	1.9	1.6	1.9	3.8	3.9	3.9
Intimate partner violence	2.8	2.6[x]	2.9	6.8	6.1[x]	6.9
Contact sexual violence by intimate partner	0.7	0.6	0.7	3.9	3.3[x]	3.9
Stalking by intimate partner	0.3	0.4	0.3	1.7	1.8	1.7
Intimate partner physical violence	2.0	1.9[x]	2.1	3.2	2.4[x]	3.2
Intimate partner psychological aggression	3.6	3.7	3.7	6.1	5.6[x]	6.1

* Full Sample = Includes all who participated on feasibility weighted to the national population; With Web/Paper Option = Includes all who participated before respondents were given the choice of modes plus those randomly assigned to the mode/paper choice group. Weighted to national population; Web only = Includes those who participated using the web. Weighted to the national population.

n/a - Not applicable. Not asked of female respondents.

[x] Significance estimated by comparing those in the Web/Paper and Web/CATI groups. Difference is significant at p<.05. See text on how test was computed.

9. Recommendations for the Next NISVS

This chapter reviews the analyses presented in the previous chapters and makes a recommendation on the use of address-based sampling (ABS), random digit dial (RDD), or both for the next National Intimate Partner and Sexual Violence Survey (NISVS).

Data Quality

When assessing data quality of the different methodologies from a total survey error perspective, it is important to consider the representation of the population and the measurement of the construct. Representation refers to how a sample of responders can be generalized to the population of interest. Measurement refers to how well the survey responses accurately reflect the underlying constructs. Both can contribute to the variability and bias of the estimates. The two designs considered in the feasibility study differ on both representation and measurement.

Representation and Variability of Estimates. The response rate for the ABS was more than three times higher the rate of the RDD. For a fixed cost, this leads to significantly better precision for the ABS design. Our initial estimate is that two to three times more data can be collected with the ABS design when compared to RDD. For a survey that measures some very rare phenomena (e.g., rape and made to penetrate), collecting more data is an important advantage of the ABS methodology. This also translates to being able to collect more data to generate local area- or state-level estimates.

The evidence on the comparative representation of particular subgroups of the two frames was mixed. Holding constant the fact that the ABS collected much more data as the RDD and whether those who responded are different from those who did not respond, the profile of the RDD and ABS respondents were similar. They both represented the sex and income groups well. They both underrepresented those in lower education groups and over-represented non-Hispanic Whites. The one area that ABS stood out was the extent that younger respondents, age 18-29, participated. This is an important age group, especially for 12-month estimates where these individuals have the highest risk of many types of sexual violence.

The feasibility study did not include Spanish-speaking respondents. Somewhat surprisingly, for the ABS, the percentage of respondents who identified as Hispanic was very close to the national average. Nonetheless, the ABS methodology, at least for those using mail paper surveys, have struggled with getting responses from those who are not fluent in English. Most of the research on non-English speakers has been with paper mail surveys. However, English literacy tends to be correlated with education. The fact that there is under-enumeration of those in the lower education group is likely indicative of some difficulties getting responses from those who cannot speak English. Unlike the paper mode, a web instrument can easily accommodate multiple languages. The issue for future research is how to best recruit non-English and lower educated respondents.

The web survey, as expected, over-represents those who have access to the internet. Offering alternative modes corrects for this somewhat, but not entirely. As internet penetration increases, this may be less of a problem. Nonetheless, a future area of research is how to bring in more respondents who are either not comfortable responding on the internet or who do not have ready access. Offering an alternative mode (e.g., phone or paper) is one step in that direction.

Representation and Bias. Judging how the RDD and ABS surveys compare with respect to nonresponse bias (NRB) is not straightforward. The two surveys were very close when looking at potential correlates of nonresponse and victimization, such as particular demographic characteristics and the health of the responders relative to the national population. The level of effort analysis indicated that lifetime prevalence estimates for ABS is subject to a negative bias for females (i.e., the estimates are low because of nonresponse).[27] The RDD results indicated a negative bias for selected measures for males and a positive bias for selected measures for females. However, with response rates between 10 and 30 percent for the two surveys, it is difficult to generalize from the nonresponse follow-up (NRFU) group to the remaining 70 to 90 percent of those who did not respond at all. Level of effort analyses like this are difficult to interpret in the absence of other evidence because the basic assumption of the continuum of resistance model has not been found to be true in some cases.[28] [29] When the NISVS was first designed and the response rate was significantly higher, a negative bias was indicated by a similar analysis.[30]

Several other measures suggest the NRB for the RDD is significant in the positive direction (estimates are too high). One is the large jump observed in NISVS estimates between the 2015 and 2016-17 surveys. The ABS results closely tracked the 2015 survey results, while the RDD tracked the 2016-17 results. It is possible to attribute the jump in 2016-17 to other changes during this time period. For example, it might have occurred because of an increased willingness to report sexual violence. However, several other surveys did not show a similar jump during this time period. The jump in the NISVS may also be due to a change in the design of the questionnaire. The 2016-17 version had fewer follow-up questions which might have reduced the number of respondents who did not report a victimization to avoid these additional questions.

Having said this, a negative bias for ABS and a positive bias for RDD is consistent with the comparison to the National Survey of Family Growth (NSFG) questions. The NISVS ABS estimates using the NSFG questions were lower than the NSFG estimate from the NSFG survey. The opposite was the case for the RDD estimates. Finally, the RDD age-specific estimates for some lifetime measures seemed to show respondents in the younger age groups to have very high estimates of lifetime victimization. The younger age groups were also under-represented in the RDD, which may be indicative of NRB specifically for these individuals.

Measurement and Bias. There was a large difference in the lifetime estimates between the two surveys. There was no consistent difference for the 12-month estimates. As noted above, the comparisons to the NSFG suggested that the ABS lifetime estimates may be too low. While NRB cannot be ruled out, it is also possible that measurement of at least some lifetime events are under enumerated for the ABS. The ABS was almost entirely a web-administered mode (93% of all responses), which may be more susceptible to error related to respondent motivation in the form of satisficing or inattention. This refers to the tendency of respondents not seeking to understand the question completely; just well enough to provide enough answers to get through the survey. The

---

[27] Note the estimates presented in this report, with exceptions explicitly noted, incorporate the NRFU portion of the survey. This should lessen the effects of any nonresponse bias.

[28] Lin, I-F, and Schaeffer, N.C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly*, 59(2), 236-58.

[29] Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias and total bias. *Public Opinion Quarterly*, 70(5), 737-758.

[30] Carley-Baxter, L., Black, M.L., and Twiddy, S. (2007). "The Impact of Incentives on Survey Participation and Reports of Intimate Partner and Sexual Violence." Paper presented at the American Association for Public Opinion Research Annual Conference, Anaheim, CA.

higher levels of missing data on the web, as well as some respondents getting through the survey in a very short period of time, is indicative of this. The definition of a completed survey did set minimum standards with respect to answering enough questions to provide the topline estimates.[31] Nonetheless, even after defining a completed survey, ABS respondents had higher levels of item-missing data, reported higher levels of burden and less willingness to take the survey again. While some of the difference related to the measures of burden with RDD may be a mode effect,[32] the pattern is also consistent with ABS respondents being less motivated. Recall of lifetime events is a particularly difficult cognitive task that may be subject to this type of satisficing.

Several of the other important measures collected on the NISVS compared favorably for the ABS relative to the RDD. This included the proportion of victims reporting intimate partners and a greater proportion of victims who report incidents occurring within the last 12 months. The latter is a more sensitive question, which perhaps the self-administered web does a better job measuring.

There are methods to reduce satisficing behavior. For example, the survey can build in prompts to catch those who answer questions too quickly. It is also possible to build in more or different questions used to measure satisficing. The feasibility survey included two "trap" questions (see Chapter 3). Almost every respondent responded to these items in the "correct" way. However, additional items that include measures that ask for more detailed responses, could also be inserted into the instrument.

Privacy, Confidentiality, and Minimizing Harm

One of the concerns with conducting a sensitive survey like NISVS is privacy and confidentiality. For CATI, the interviewer is in control over the dissemination of the information about the survey topic. During the design phase of the feasibility surveys, there was some question about whether a totally self-administered survey would be able to keep the topic private. For example, the graduated consent procedure relied on a strategic release of information in the advance letters, the web pages and the paper survey. The ABS procedures did not reveal the critical details on the topic of the survey until the respondent was into the first few sections of the survey. If someone in the household opened the mail, they would not know about the survey.

A higher percentage of ABS respondents, compared to those on the RDD, reported that someone else knew about the topic of the survey and that someone else was in the room when they were taking the survey. However, the overall percentages on these were still low (20% for ABS) and this did not vary much by whether someone reported a victimization on the survey.

A second concern related to a self-administered survey is minimizing harm if a respondent gets upset when taking the survey. In an RDD mode, interviewers are available to provide help-line information to respondents who may become upset. The ABS relies on respondents reading the written material provided in the advance letters, web page and paper survey. By all indications, the extent respondents got upset when taking the survey, at least to the point that would be considered more than minimum harm, is not high. The percentage of respondents who reported thinking of things they would not like to think about or experiencing intense emotions was the same between the two frames. The RDD survey only observed one instance that a respondent expressed an emotional response that the interview had to be interrupted. This was considered a low-level

---

[31] See Chapter 3 – Definition of a Completed Survey.

[32] For example, the self-administration may lead to a greater willingness to be critical of the survey.

emotional response. While an equivalent measure is not available for the ABS, the survey did not receive any calls from respondents that were emotionally upset or in need of help.

Recommendation for Future NISVS

When considering a recommendation for the design of the NISVS, it is important to keep in mind the context that motivated the redesign work. The NISVS response rates have been dropping like those for all other RDD surveys. With this drop in response rates the cost is getting higher. As noted above, our initial calculation is that the RDD component of the study was between two to three times more expensive per completed survey than the ABS. All signs are pointing to RDD response rates continuing to drop. This threatens the viability of this methodology as a way to collect data. At the very least, it leads to an expectation that this methodology will become even more expensive when moving forward.

Use the ABS sample frame with push-to-web methodology

With this context in mind, as well as the measures of quality summarized above, the best design for NISVS is the ABS, push-to-web design. There is very little reason to include RDD. The prevalence measures for the RDD are much higher than equivalent surveys and it is much more expensive to implement. With response rates continuing to decline, these problems will only get more severe.

The above analyses also suggest the ABS design may have a negative bias, perhaps from measurement rather than nonresponse. The evidence is preliminary and is based primarily on comparisons with the NSFG measures. As with any comparisons, this measure is somewhat flawed because of the inherent differences between the ABS and NSFG designs. Nonetheless, the higher levels of missing data on the web suggest that some web respondents may not be fully engaged. Of course, measurement error is also present in interviewer-administered surveys as well. So, it should not be seen as a unique problem with the push-to-web methodology. Further research should be conducted on ways to promote respondent engagement on the web and methods to measure it.

Use additional mailings

The design used in the feasibility study was limited by calendar time. As a result, it was not possible to make four contact attempts at each stage of the survey. Using four attempts is standard procedure for most surveys using postal mail as the contact method. The screening survey had three contacts: 1) initial recruitment letter, 2) reminder postcard, and 3) follow-up with paper screener. The follow-up for extended respondents had two contact attempts: 1) initial request and 2) offer of alternative mode.

The nonresponse follow-up did add a contact, but this is different because it changes the essential survey conditions by offering respondents an additional incentive. We recommend that the survey add one additional contact to both the screening and extended phase of the survey: 1) for the screener, an additional follow-up which includes the paper screener and 2) for the extended, an additional follow-up that offers the respondent the choice of the mode.

Use the Web/CATI optional Group

There are three primary advantages of the web/paper option group. The first is that it resulted in a slightly higher response rate (1 point) than the web/CATI group. Second, it brought in individuals with lower education, which is a group that is generally underrepresented in this type of ABS design. The

small sample sizes make it difficult to assess whether the observed differences in who responded to each of the option groups was significant. But it is the case, based on prior research, that different types of respondents tend to use the paper mode. Furthermore, the results suggest that this difference in representation may lead to respondents who are less likely to report a victimization. When comparing estimates between the web/paper group to the web/CATI group, the former was consistently lower. The third advantage is that the optional mode (paper) is self-administered, which is compatible with the web survey. This should minimize any effects that mode of administration (interviewer vs. self-administered) has on responses.

There are three advantages of the web/CATI option group. The first is that it yields a complete dataset for all respondents. The paper instrument was shortened and did not include all of the items on the full NISVS questionnaire. For example, the paper instrument did not collect the detail on each perpetrator as on the web and CATI instrument. The second advantage is the web/CATI has more control over the privacy of the interview. Individuals can only be exposed to the web questionnaire by signing onto the account, using the appropriate username and password. The paper instrument was included in the postal package addressed to the particular respondent. The third advantage is that there was less item-missing data for the web and CATI as compared to the paper. The computerized instruments are able to provide more guidance to respondents when navigating the skip instructions. This results in lower rates of item-missing data.

Approximately 80 percent (weighted) of the surveys were completed before the respondent was given the choice between modes. This limits the effect this feature of the design can have on the final results. This might change once the additional mailing recommended above is instituted. The two yield very similar prevalence estimates. The decision between these two options comes down to trading off the slightly better representation of the web/paper with the slightly more complete dataset that the web/CATI offers. The web/CATI also offers some advantages with respect to item-missing data and privacy. We recommend using the web/CATI option. In our view, the completeness of the data outweighs the small increases in representation the web/paper offers. However, this is a close call. It might be worth considering running a similar experiment on the national study to collect more data when all mailings are instituted.

Use the probability method of respondent selection

The differences between the two selection methods were negligible. Of note, both of the methods produced the same distributions for age and sex. The primary objective of the YMOF method was to boost participation of young people and males. Both of these groups are typically under-enumerated in surveys involving postal contacts. However, both the YMOF and the Rizzo-Brick-Park (RBP) methods were similar with respect to these two demographic groups. In addition, there was no difference in the lifetime and 12-month prevalence estimates.

The RBP probability method is recommended for the larger study. The probability method is preferred because it maintains full coverage of all individuals within the household. In a small number of households, the YMOF does not assign a non-zero probability of selection to every member. Given there is no difference in either response or victimization estimates between the two methods, the RBP method is recommended so that all adults have a non-zero chance of selection.

Include items on the NISVS questionnaire that allow assessment of representation and bias

The feasibility study included a number of additional measures to compare against other surveys, including the ACS (internet use, born in United States, home ownership); the NHIS (mental/emotional problems, hospitalization, have asthma, have been depressed, any adult injured), and the NSFG (forced vaginal intercourse by male, by female and oral/anal sex by a male). These items were useful for the analysis of data quality on the NISVS. These measures, or at least a subset of these measures, should be included in the survey. The measures used on the NISVS should be coordinated with the most recent versions available from each respective survey.

Include more items to measure the attention of the respondent on the web survey

The Feasibility instrument included two items to assess whether the respondent was carefully reading the questions. One was placed at the beginning of the survey and the other at the end. More, or at least different items, should be placed on the instrument. Other approaches might include an additional attention check placed in the middle of the survey (e.g., pick "x" from the list displayed). It could also include putting an open-ended item at the end asking about a general topic (e.g., Health policy issues). This would be used to see if respondents put in coherent answers, put in a non-sequitur or skip it entirely.[33]

---

[33] Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., and D. Asare-Marfo (2020) Assessing the risks to online polls from bogus respondents. Pew Research Center, Washington DC. Available at: Bogus respondents and online polls - Pew Research Center Methods | Pew Research Center.
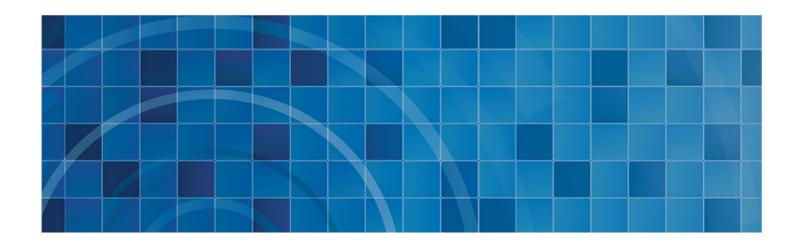
Appendix

National Intimate Partner and Sexual Violence Survey (NISVS) Redesign – Weighting Plan

National Intimate Partner and Sexual Violence Survey (NISVS) Redesign – Weighting Plan

Feasibility Study Weighting Plan
Contract Number GS00F009DA

October 21, 2020

Table of Contents

Introduction

The National Intimate Partner and Sexual Violence Survey (NISVS) Feasibility Study will produce national data on the prevalence of sexual violence, intimate partner violence (IPV), stalking, and a variety of measures related to the consequences of these forms of victimization. The Feasibility Study tests alternative sampling frames, modes of data collection, and experimental data collection features. The study will generate estimates from each frame, compare data quality measures across frames and make recommendations on a final design for the NISVS moving forward.

The purpose of this document is to describe the weighting procedures used on the NISVS Feasibility Study for both the random digit dialing (RDD) sample and address-based sample (ABS). The RDD sample and ABS are treated as two independent samples, and weighting is performed separately for both samples. Since the probability of a person being selected for both samples is marginal, we do not make any considerations for the probability of being selected for both the ABS and RDD samples. Figure 1 outlines the overall weighting process and lists the weights produced for both the ABS and RDD sample.

The remainder of the introduction is a high-level overview of the weighting approach. The details of the weighting are included in the subsequent sections. Although relevant information about sample design is included in this document, more detailed information on the sample design can be found in the Feasibility Study Plan.

The RDD design is a dual frame national survey, with approximately 65% of the frame being cell phone numbers, and approximately 77% of the sampled numbers being cell phone numbers. Approximately 80 percent of completes are expected from cell phones and 20 percent from landlines (LL). For the RDD sample, the weighting process begins by creating base weights

separately for the cell phone and LL samples. The base weights undergo an unknown eligibility adjustment that accounts for the proportion of cases with unknown eligibility that are assumed to be eligible. Next, we perform an adjustment to account for the nonresponse follow-up (NRFU) subsample. These weights are then adjusted for nonresponse and the landline weights are given an additional adjustment to account for the number of phone lines in the household. We then calculate person-level weights based on the number of people in the household. Composite weights are calculated that combine the separate cell phone and LL adjusted weights. Finally, we rake and trim the composite weights to be consistent to external totals.

The ABS sample is nationally representative and multi-mode. Selected addresses are first asked to complete the survey on the web. Initial nonrespondents are then given the option to complete the survey over the phone (in-bound CATI) or by paper. Like the RDD weighting process, we begin by creating base weights based on the inverse of the household sampling rate and then creating unknown eligibility adjusted weights. A subsample of initial screener nonrespondents and extended nonrespondents are selected for NRFU subsamples and followed with extra effort. To adjust for this, the two NRFU subsamples are weighted by the inverse of the subsampling rates. Then, we adjust the weights for nonresponse. These nonresponse adjusted household weights are used to create person-level weights that account for within household selection probability. Finally, we rake and trim the person-level weights to be consistent to external totals. Along with the full-sample weight (regardless of mode of data collection), we create two sets of specialized weights that follow the same adjustments as the full-sample weight:
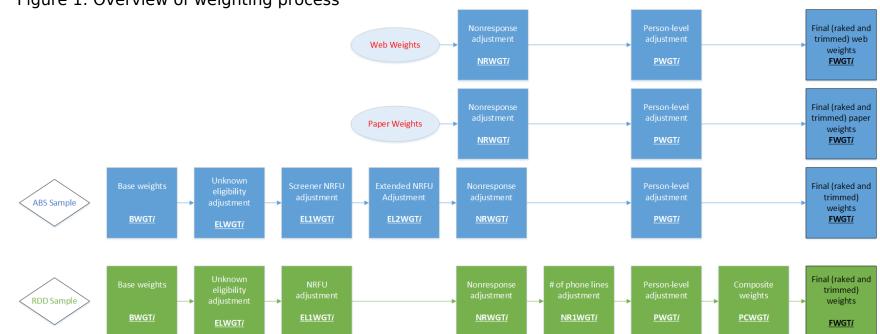
1) One set considers only those who respond by web
2) A second set considers only those who respond by either 1) the initial web request or 2) are assigned to paper follow-up (i.e. excludes those

who do not respond to initial web request and are assigned to CATI
follow-up)

For both the ABS and RDD samples, we provide replicate weights, variance
stratum, and variance unit variables that allow for variance estimation using
either Jackknife replication methods or Taylor Series estimation, described in
later sections.

Figure 1: Overview of weighting process

Web Weights → Nonresponse adjustment **NRWGT*i*** → Person-level adjustment **PWGT*i*** → Final (raked and trimmed) web weights **FWGT*i***

Paper Weights → Nonresponse adjustment **NRWGT*i*** → Person-level adjustment **PWGT*i*** → Final (raked and trimmed) paper weights **FWGT*i***

ABS Sample → Base weights **BWGT*i*** → Unknown eligibility adjustment **ELWGT*i*** → Screener NRFU adjustment **EL1WGT*i*** → Extended NRFU Adjustment **EL2WGT*i*** → Nonresponse adjustment **NRWGT*i*** → Person-level adjustment **PWGT*i*** → Final (raked and trimmed) weights **FWGT*i***

RDD Sample → Base weights **BWGT*i*** → Unknown eligibility adjustment **ELWGT*i*** → NRFU adjustment **EL1WGT*i*** → Nonresponse adjustment **NRWGT*i*** → # of phone lines adjustment **NR1WGT*i*** → Person-level adjustment **PWGT*i*** → Composite weights **PCWGT*i*** → Final (raked and trimmed) weights **FWGT*i***

1. Random Digit Dial Weighting Plan

1.1 Base Weights

The sampling frame for the LL sample is formed through compiling all 100-banks of telephone numbers that have at least one listed residential number in published telephone directories. The Comprehensive Screening Service (CSS) is used to screen for non-residential businesses and non-working numbers on the LL frame, which are ineligible for this survey. The sampling frame for cell phones is formed similarly through 1000-banks of numbers dedicated to wireless service. The sampling provider is able to screen for non-active cell phone numbers on the cell phone frame, which are ineligible for this survey. Table 1 shows the frame size and sample size for both landlines and cell phones. The frame size includes all phone numbers obtained for the frame, including those pre-screened and identified as ineligible.

The base weights are calculated separately for landline and cell phone samples based on the sampling rate. The sampling rate is calculated as the sample size divided by the frame size, separately for landlines and cell phones. Table 1 shows the base weights, calculated as the inverse of the sampling rate, given to the LL and cell phone sampled cases. The base weight ($BWGT0_i$) is a constant for each sampled number (i) within its respective sampling frame.

Table 1: Frame size, sample size, and base weights for RDD Sample

Sample Frame	Frame Size	Sample Size	Base Weight
Landline	289,962,700	6,572	44,120.92
Cell Phone	526,057,000	22,914	22,957.89

1.2   Weighting Disposition Codes

Each sampled number is assigned a weighting disposition code that is used to create the various weights described throughout section 1. The weighting disposition codes are eligible respondents (ER), eligible nonrespondents (ENR), ineligible numbers (IN), and numbers with unknown eligibility (UNK). An eligible respondent either has a 'completed' interview (i.e. all applicable questions answered including consequences and debriefing), a 'completed through CQ' interview (i.e. all applicable questions answered through consequences), or a 'partial complete' interview (i.e. all applicable questions answered through stalking section questions ST1-ST18). All other cases are assigned a weighting disposition code of ENR, IN, or UNK based on their field disposition code. Appendix Item A shows how the field disposition codes are grouped into weighting disposition codes for RDD.

1.3   Adjusting for Unknown Eligibility

The unknown eligibility adjusted weight (ELWGT0$_i$) accounts for the proportion of cases with unknown eligibility assumed to be eligible. This is achieved by reallocating the base weights of those with unknown eligibility (UNK) to those with known eligibility (ER, ENR, IN), separately for LL and cell phone samples. This reallocation is performed within adjustment cells formed using Census region and Census division. The adjustment factor can be written as

$$EL_{AD}J_c = \left\{ \frac{T_C}{T_{ER_C}+T_{EN R_c}+T_{I N_c}} \, for \, ER, ENR, \in \, i \, 0 \, for \, UNK \right.$$

where c is the adjustment cell, T$_C$ is the sum of weights for all cases in adjustment cell c, T$_{ERc}$ is the sum of weights for eligible respondents in adjustment cell c, and T$_{ENRc}$ and T$_{INc}$ can be defined similarly for eligible nonrespondents and ineligible cases, respectively. The unknown eligibility adjusted weight is defined as ELWGT0$_i$ = BWGT0$_i$ * EL_ADJ$_c$. All cases with a

weighting disposition code of ER, ENR, and IN will have positive, non-zero values for ELWGT0$_i$.

1.4 Adjusting for NRFU Subsample

After the regular data collection process is completed, initial nonrespondents are selected for the NRFU with subsampling rate *f*, and followed with extra effort. For RDD, this subsampling rate (f) is 0.323, and is consistent throughout the data collection period. We allocate the weights of cases eligible and not selected for the NRFU to those that were selected by weighting selected cases by the inverse of the subsampling rate. The adjustment factor for this NRFU adjustment is

$$EL1_{ADJ} = \begin{cases} \dfrac{1}{f} \text{ selected for NRFU subsample} \\ 0 \text{ eligible but not selected for NRFU subsample} \\ 1 \text{ ineligible for NRFU subsample} \end{cases}$$

The NRFU subsample adjusted weight is EL1WGT0$_i$ = ELWGT0$_i$ * EL1_ADJ, and it reflects all previous adjustments along with an adjustment for the NRFU subsample. All cases with a weighting disposition code of ER, ENR, or IN can have a positive, non-zero weight for EL1WGT0$_i$.

1.5 Adjusting for Nonresponse

Next, we create weights that adjust for nonresponse of the extended interview (NRWGT0$_i$). This is achieved by reallocating the weights of nonrespondents (ENR) to respondents (ER). This adjustment is done separately for LL and cell phone samples, and the reallocation is performed within adjustment cells created by Census region and Census division. The adjustment factor is defined as

$$NR_{ADJ_c} = \begin{cases} \dfrac{T_{ER_c} + T_{ENR_c}}{T_{ER_c}} \text{ for } ER \\ 0 \text{ for } ENR, \in, UNK \end{cases}$$

The nonresponse adjusted weight is $NRWGT0_i = EL1WGT0_i * NR\_ADJ_c$ and it reflects all previous adjustments along with a nonresponse adjustment. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $NRWGT0_i$.

1.6 Adjusting for Number of Phone Lines

The nonresponse adjusted weights ($NRWGT0i$) for cases in the LL sample are adjusted for the number of reported phone lines dedicated to voice communication in the household. Separately for the LL and cell phone samples, we use hot-deck imputation with donors coming from within age groups (i.e. 18-35, 36-64, 65+) to impute a handful of missing values on the question asking for the number of phone lines in the household. Since very few households reported having more than two phone lines dedicated to voice communication, this adjustment caps the number of phone lines at two to minimize effects on the weights. The adjustment factor is defined as

$$NR1_{AD}J_i = ¿.$$

The adjusted weight is calculated as $NR1WGT0_i = NRWGT0_i * NR1\_ADJ_i$, and it reflects all previous adjustments along with an adjustment for the number of phone lines in the household. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $NR1WGT0_i$.

1.7 Person-Level Adjustment

Next, we create person-level weights ($PWGT0_i$). This adjustment is simply the inverse of the number of adults in the household, capped at four. For the small number of completed cases in the landline sample with missing household size information (26), the number of adults in the household is imputed using a hot-deck with donors coming from within states. For the cellphone sample, the person that answers the phone is automatically

selected, so there is no within-household adjustment needed. The adjustment is defined as

$$P_{AD}J_i = \begin{cases} 1/Min(\text{¿ } Adlts_i, 4) \text{ for the } \ll sample \\ 1 \text{ for the cell sample} \end{cases}$$

The person-level weight is calculated as $PWGT0_i = NR1WGT0_i * P\_ADJ_i$ and it reflects all previous adjustments along with an adjustment for the number of adults in the household. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $PWGT0_i$.

1.8 Compositing

Next, following the standard NISVS RDD weighting procedures, we create four domains of household members:

a) Adults with only landlines (a)
b) Adults with landlines and cell phones in the household, reached on their landline (b)
c) Adults with landlines and cell phones in the household, reached on their cell phone (c)
d) Adults with only cell phones (d)

Those in (b) and (c) could have been sampled from either frame, so their weights are composited. Let A be the size of the landline frame and B be the size of the cell phone frame. The compositing factor is $\lambda = \dfrac{A}{A+B}$. The estimates for the size of each telephone population are based on the NHIS estimates (i.e., Table 2 of the early release report by Blumberg and Luke, 2019). The estimate is $\lambda = \dfrac{38.2}{38.2+96.4} = .28$.

Since the composite weight adjustment is determined by the type of phone(s) in the household and no survey question asks the landline sample

whether they have a cell phone, we explored using logistic regression to impute cell phone status for the landline sample using the data in Blumberg and Luke (2019). The tables in that report only show the univariate distribution, so we chose age, which is the variable most correlated with the probability of having a cell phone. The logistic regression model below is based on the probability of having a cell phone for those below 65 years of age (Blumberg and Luke, 2019).

$$\log\left(\frac{p}{1-p}\right)=\beta_0+\beta_1 x_1$$

Where p is the probability of having a cell phone and $x_1 = \begin{cases} 0 \, when \, age < 65 \\ 1 \, when \, age \geq 65 \end{cases}$. The imputation based on this gave those below 65 years of age a probability of 0.9825 of having a cell phone, and a probability of 0.886 for those above 65 years of age.

The adjustment for composite weights is

$$PC_{ADJ}=\begin{cases} \lambda \, group(b) \\ (1-\lambda) \, group(c) \\ 1 \, groups(a) \wedge (d) \end{cases}$$

The composite weight is calculated as PCWGT0$_i$ = PWGT0$_i$ * PC_ADJ. All cases with a weighting disposition code of ER will have a positive, non-zero weight for PCWGT0$_i$. The composite weight reflects the base weight with all adjustments described in previous sections. The final expression of the composite weight can be defined as:

$$PCWGT0_i = BWGT0_i * EL_{AD}J_C * EL1_{ADJ} * NR_{AD}J_C * NR1_{AD}J_i * P_{AD}J_i * PC_{ADJ}.$$

1.9   Raking and Trimming

The final step in weighting adjustments is calibrating the composite weights so the weighted totals in several dimensions matches control totals from external sources.

The first four dimensions are the composition of adults according to age (18-29, 30-44, 45-59, 60+), race (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic multi-racial, non-Hispanic other), marital status (never married, married, other), and education (high school or below, some college, Bachelor's degree or higher). We obtain external control totals for these four dimensions from the 2014-2018 5-year ACS tables. The ACS 5-year data provide the most precise estimates for these dimensions compared to ACS 3-year or 1-year data. The fifth raking dimension is the estimated composition of adults by telephone status and gender. Since these data are only available from Blumberg and Luke (2019) we compute the proportions from that source and apply the ACS population totals to get the estimated totals. The table below gives the domain and proportions, which are from Table 2 of the early release report by Blumberg and Luke, 2019.

Table 2: Proportion of adults by telephone status and gender

Domain	Proportion
Males	
Cell only	0.3041
Dual	0.1705
Landline-only	0.0177
Females	
Cell only	0.3210
Dual	0.1701
Landline-only	0.0166

A relatively small number of responding records have missing data for one or more of these demographics. To be able to rake the weights, these missing values need to be imputed. We use hot-deck imputation to impute missing demographic information separately for the landline and cell phone samples.

Raking is a calibration method that allows controlling to a large set of auxiliary control totals or raking dimensions. The raking adjustment rakes the weights until they converge and then trims them. The goal of the trimming is to avoid large weights that could increase the variance of the estimates or make estimates of subdomains very unstable. Any raked weight that is greater than the cutoff of 4.5 times the median weight is trimmed to equal that cutoff value. This process of raking and trimming repeats until the raked weights are below this trimming cutoff and are consistent with the control totals for each dimension, which we define as having an absolute difference of less than five in each dimension's level. The minimum cell size allowed within the raking dimensions is set to 30 per level, and then is lowered if the adjustment factor yielded is not problematic. No more than 15 raking iterations are allowed per cycle, and no more than ten raking and trimming cycles are allowed in this process.

The raking adjustment produces a common adjustment factor for each of the 1,080 combinations of the levels of the five raking dimensions (4 x 5 x 3 x 3 x 6 levels). The raked and trimmed weight is $FWGT0_i = PCWGT0_i * RT\_ADJ_c$, where RT_ADJ_c is the common adjustment in the specific combination, c. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $FWGT0_i$. This final weight is composited and at the person-level, and is adjusted for unknown eligibility, the NRFU subsample, nonresponse, the number of phone lines in the household, and is raked to auxiliary control totals.

1.10 Variance Estimation

Since we do not observe the entire population, all estimates produced from the sample have errors due to sampling variance. The two methods of computing sampling errors for complex sample surveys are Taylor series and replication. The final weighting data set contains the following that allow for variance estimation:

1) replicate weights, and

2) variance stratum and variance unit identifiers.

We implement a stratified jackknife (JK2) variance estimator (Rust and Rao 1996) that involves creating a set of replicate weights that are produced by applying the same adjustments made to the full sample weight. The JK2 method has nice statistical properties as discussed in Rust and Rao and it permits combining samples easily because the constant multiplier of the squared error terms is unity and is not dependent on the number of replicates. We use the JK2 method with 100 replicates and produce 100 replicate weights that allow for computations of standard errors of estimates. Each of the replicate weights is generated by setting the replicate weight to zero for a grouped portion of sampled numbers and then reweighting the non-zero sampled numbers. The replicates are created by first randomly collapsing the sampled numbers into pairs. The first pair is assigned a variance stratum code of 1, and each number in the pair is randomly assigned a variance unit of either 1 or 2. The next pair is assigned a variance stratum of 2 and variance units 1 and 2, etc. The pairs are assigned variance stratum and variance units similarly, with the variance stratum restarting at 1 after the 100th pair.

Next, we use the variance strata (ranging from 1-100) and variance units (ranging from 1-2 within each variance stratum) to create 100 replicate base weights for each sampled number. Replicate 1 base weights are set equal to the full sample weight for those in variance stratums 2-100. For those in variance stratum 1, replicate 1 weights are double the full sample weight for those in variance unit 1, and zero for those in variance unit 2. Replicate weights 2-100 are created similarly using the relevant variance stratum, and the replicate base weights are BWGT1-BWGT100. This process of creating the variance stratum, variance unit, and replicate base weights is computed separately for the cell phone and LL samples. The 100 replicate base weights

Westat®

are adjusted in the same manner as described above, and 100 replicate weights exist for each weight described in the previous sections. The final full-sample weight is FWGT0$_i$ and the final replicate weights are FWGT1$_i$ – FWGT100$_i$.

The jackknife variance estimator can be computed with these replicate weights using any statistical package that handles survey data (e.g., R, SAS, SUDAAN, STATA). The full-sample variance estimate is computed as

$v(\hat{y}) = \sum_{r=1}^{100} ¿¿$, where $\hat{y}$ is the estimate of the outcome y using the full sample weights, and $\hat{y}_{(r)}$ is the estimates of outcome y using using the replicate weight r.

We also include variance stratum and variance unit variables for computing variance estimates using Taylor series linearization. These are the same variance stratum and variance unit variables described earlier in this section used to create replicate weights. Unlike the replicate weights, the Taylor series method does not reflect all the steps in the weighting process, but it generally gives estimates of sampling errors for means and proportions that are very similar to those from replication. Similar software used for computing the jackknife variance estimator can be used to produce variance estimates based on Taylor series linearization.

2.   Address - Based Sample Weighting Plan

2.1   Base Weights

The sampling frame for ABS is drawn from a database of addresses from the Computerized Delivery Sequence File used by Marketing Systems Group (MSG) to provide random samples of addresses. This database is estimated to cover approximately 98 percent of all households in the country, and it includes all non-vacant residential addresses in the United States, including post office (P.O.) boxes, throwbacks (i.e., street addresses for which mail is redirected by the United States Postal Service to a specified P.O. box), and seasonal addresses. The sampling frame has 130,162,901 addresses, and a sample of size n=12,566 is drawn with equal probability using implicit stratification sorted by geography using zip code +4 within states. The base weight, ($BWGT0_i$), is calculated as the frame size divided by the sample size, and is constant for each sampled unit *i*. Therefore, the base weight for all records is 10,358.34.

2.2   Weighting Disposition Codes

Each sampled address is assigned a weighting disposition code that is used to create the various weights described throughout section 2. The weighting disposition codes are eligible respondents (ER), eligible nonrespondents (ENR), ineligible numbers (IN), and numbers with unknown eligibility (UNK). An eligible respondent is __. All other cases are assigned a weighting disposition of ENR, IN, or UNK.

2.3 Adjusting for Unknown Eligibility

We adjust the base weights to account for the proportion of addresses with unknown eligibility assumed to be eligible. First, we estimate the proportion of cases with unknown eligibility assumed eligible, e, using the "backing out" method by DeMatteis (2019). We calculate e as

$$e = \frac{1}{T_{UNK}} \left(\hat{T}_{ext} - T_{elig} \right)$$

where T_{UNK} is the sum of weights for cases with unknown eligibility (UNK), T_{elig} is the sum of weights for addresses in the survey with known eligibility (i.e., either ER, ENR, or IN), and $\hat{T}_{ext}$ is an external estimate of the total number of households nationwide. We use the 2014-2018 ACS 5-year estimate of number of households as our external estimate, $\hat{T}_{ext}$. In the case where e is negative or greater than 1, we will set e equal to 0 or 1, respectively.

The American Association of Public Opinion Research's report on methods for computing response rates gives the formula for the proportion of eligible addresses in the frame as

$$e = \frac{N_{ER} + N_{ENR} + p * N_{UNK}}{N_{ER} + N_{ENR} + N_{¿} + N_{UNK}}$$

where N_{ER} is the number of known eligible responding addresses in the frame (ENR), and N_{ENR}, N_{IN}, and N_{UNK} can be defined similarly. Using this formula, we can solve for p, which is the proportion of base weights of those with unknown eligibility (UNK) that needs to be reallocated to those with known eligibility (ER, ENR, IN). This reallocation is done within weighting adjustment cells formed using Census division, Hispanic status, and whether the

sampled person owns or rents their home. The adjustment factor is defined

$$
as \ EL_{AD}J_c = \begin{cases} \dfrac{T_{ER_c} + T_{ENR_c} + (p * T_{UNK_c})}{T_{ER_c} + T_{ENR_c}} for \ ER, ENR \\ \dfrac{T_{ER_c} + T_{ENR_c} + ((1-p) * T_{UNK_c})}{T_{ER_c} + T_{ENR_c}} for \in \ \textit{¿} \ 0 \ for \ UNK \end{cases}
$$

where c is the adjustment cell, T_{ERc} is the sum of weights for eligible respondents in adjustment cell c, and T_{ENRc} and T_{UNKc} can be defined similarly. The unknown eligibility weight is calculated as $ELWGT0_i = BWGT0_i * EL\_ADJ_C$. All cases with a weighting disposition code of ER, ENR, and IN will have a positive, non-zero weight for $ELWGT0_i$.

2.4 Adjusting for NRFU Subsamples

After the regular data collection process is complete, initial screener nonrespondents are selected for the NRFU with subsampling rate f_1 and followed with extra effort. For the ABS screener NRFU, this subsampling rate (f_1) is 0.5, and is constant for all screener nonrespondents throughout the data collection period. We allocate the weights of cases eligible and not selected for the screener NRFU to those that were selected by adjusting selected cases by the inverse of the subsampling rate. The screener NRFU adjustment factor is

$$
EL1_{ADJ} = \begin{cases} \dfrac{1}{f_1} \ selected \ for \ screener \ NRFU \\ 0 \ eligible \ but \ not \ selected \ for \ screener \ NRFU \\ 1 \ ineligible \ for \ screener \ NRFU \end{cases}
$$

The screener NRFU adjusted weight is calculated as $EL1WGT0_i = ELWGT0_i * EL1\_ADJ$, and it reflects all previous adjustments along with an adjustment for the screener NRFU subsample.

Similarly, initial extended nonrespondents are selected for the NRFU with subsampling rate f_2 and followed with extra effort. For the ABS extended NRFU, this subsampling rate (f_2) is 0.5, and is constant for all extended nonrespondents throughout the data collection period. The extended NRFU adjustment factor is

$$EL2_{ADJ} = \begin{cases} \dfrac{1}{f_2} \text{ selected for extended NRFU} \\ 0 \text{ eligible but not selected for extended NRFU} \\ 1 \text{ ineligible for extended NRFU} \end{cases}$$

The extended NRFU adjusted weight is calculated as $EL2WGT0_i = EL1WGT0_i *$ EL2_ADJ, and it reflects all previous adjustments along with an adjustment for the extended NRFU subsample. All cases with a weighting disposition code of ER, ENR, or IN can have a positive, non-zero weight for $EL1WGT0_i$ and $EL2WGT0_i$.

2.5 Adjusting for Nonresponse

Next, we create weights that adjust for nonresponse of the extended interview ($NRWGT0_i$). This is achieved by reallocating the weights of nonrespondents (ENR) to respondents (ER) within adjustment cells created by Census division, Hispanic status, and whether the sampled person owns or rents their home. The adjustment factor is defined as

$$NR_{AD}J_c = \begin{cases} \dfrac{T_{ER_c} + T_{ENR_c}}{T_{ER_c}} \text{ for ER} \\ 0 \text{ for ENR}, \in, UNK \end{cases}$$

The nonresponse adjusted weight is $NRWGT0_i = EL2WGT0_i * NR\_ADJ_C$ and it reflects all previous adjustments along with the nonresponse adjustment. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $NRWGT0_i$.

2.6   Person-Level Adjustment

Next, we create person-level weights ($PWGT0_i$). This adjustment is simply the inverse of the number of adults in the household, capped at four. For the small number of completed cases with missing household size information, the number of adults in the household is imputed using a hot-deck with donors coming from within states and mode of data collection (i.e. web, in-bound CATI, or paper). The adjustment is defined as $P\_ADJ_i = 1/Min(\#Adults_i, 4)$. The person-level weight is calculated as $PWGT0_i = NRWGT0_i * P\_ADJ_i$ and it reflects all previous adjustments along with an adjustment for the number of adults in the household. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $PWGT0_i$.

2.7   Raking and Trimming

The final step in weighting adjustments is calibrating the composite weights so the weighted totals in several dimensions matches control totals from external sources. The five dimensions used are the composition of adults according to age (18-29, 30-44, 45-59, 60+), race (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic multi-racial, non-Hispanic other), marital status (never married, married, other), education (high school or below, some college, Bachelor's degree or higher), and gender (female, male). We obtain external totals for these dimensions from the American Community Survey 2014-2018 5-year data.

A relatively small number of complete records have missing data for one or more of these demographics. To be able to rake the weights, these missing values need to be imputed. We use hot-deck imputation to impute missing demographic information with donors coming from within mode (i.e. web, in-bound CATI, or paper).

As described in section 1.9, we iteratively rake and trim the weights to these control totals until the raked weights are below a trimming cutoff (i.e. below

4.5 times the median weight) and are consistent with the control totals for each dimension, which we define as having an absolute difference of less than five in each dimension's level. The minimum cell size allowed within the raking dimensions is set to 30 per level, and then is lowered if the adjustment factor yielded is not problematic. No more than 15 raking iterations are allowed per cycle, and no more than ten raking and trimming cycles are allowed in this process.

 The raking adjustment produces a common adjustment factor for each of the 360 combinations of the levels of the five raking dimensions (4 x 5 x 3 x 3 x 2 levels). The raked and trimmed weight is $FWGT0_i = PWGT0_i * RT\_ADJ_c$, where RT_ADJ_c is the common adjustment factor in the specific combination c. All cases with a weighting disposition code of ER will have a positive, non-zero weight for $FWGT0_i$. This final weight is at the person-level, and is adjusted for unknown eligibility, the screener and extended NRFU subsamples, nonresponse, and is raked to auxiliary control totals.

2.8　Variance Estimators

 The final weighting data set contains the following that allow for variance estimation:

1) replicate weights, and

2) variance stratum and variance unit identifiers.

We implement a stratified jackknife (JK2) variance estimation method with 100 replicates, similar to the process described for RDD. The replicates are created by first sorting the sample by zip+4 and collapsing the addresses into pairs. The first pair is assigned a variance stratum code of 1, and the pair is randomly assigned a variance unit of either 1 or 2. The next pair is assigned a variance stratum of 2 and variance units 1 and 2, etc. All pairs are assigned variance stratum and variance units similarly, with the variance

Westat®

stratum restarting at 1 after the 100th pair. These variance stratum and variance unit variables can be used to calculate the full-sample variance estimate using Taylor Series linearization. Replicate base weights, $BWGT1_i$-$BWGT100_i$ are created similarly to that of the RDD replicate base weights outlined in section 1.10. The 100 replicate base weights are adjusted in the same manner as the full-sample weights, and 100 replicate weights exist for each weight described previously in section 2. The final full-sample weight is $FWGT0_i$ and the final replicate weights are $FWGT1_i$ – $FWGT100_i$. Section 1.10 outlines how to calculate the full-sample variance estimate using the replicate weights.

2.9   Specialized Weights

The series of weights described previously in section 2 includes respondents from all modes (i.e. web, paper, in-bound CATI), which we call these the set of full-sample weights. We also produce two sets of specialized weights, which we call web weights and paper weights. For web weights, only those who responded by web are considered eligible respondents (ER). This set of weights can be used to produce estimates as if we did not offer any alternative mode. For paper weights, only those who respond either to the initial web request or responded by either web or paper after being assigned to paper follow-up are considered eligible respondents (i.e. excluding those who did not respond to the initial web request and were assigned to in-bound CATI follow-up). This set of weights can be used to analyze the data as if only web and paper collection had been done. Since very few cases respond by in-bound CATI, we do not provide a set of specialized weights for this mode.

We provide three weighting files with identical variables, one for the set of full-sample weights, one for the set of web weights, and one for the set of paper weights. The base weight (BWGT0), unknown eligibility adjusted weight (ELWGT0), and NRFU adjusted weights (EL1WGT0, EL2WGT0) are

Westat®

identical across the three files. The nonresponse adjusted weight (NRWGT0), person-level weight (PWGT0), and final raked and trimmed weight (FWGT0) will differ across the files. For all three sets of weights, the formulas for adjustment factors and weights remains the same as described previously in section 2. The only difference in calculations lies in the classification of eligible respondents for the three sets of weights.

3. References

Blumberg, S. and J. V. Luke, "Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2019," https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless202009-508.pdf.

DeMatteis, J. M. (2019). Computing "e" in Self-Administered Address-Based Sampling Studies.

Rust, K. F., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research*, *5*(3), 283-310.

4.   Appendix

Appendix Item A: RDD Field and Weighting Disposition Codes

Weighting Disposition Code	Field Disposition Code	Label	Description
ENR	LH	Final Language Problem - Hearing/Speech	Two calls to this respondent resulted in a hearing or speech communication problem.
	LM	Max Calls - Language	Questionnaire had an additional language problem and has reached the maximum calling algorithm.
	LP	Final Language Problem - Non-English	Two calls to this respondent resulted in a non-English communication problem.
	MC	Max Calls	The calling algorithm has been fulfilled. At least one "human" contact has been made at the number and there are no refusals or language problems in the call history for the household.
	NL	Not Locatable	The sampled person was not located. If the project is tracing, this code is assigned after the use of tracing resources.
	NP	Not available in Field Period	Nonresponse: subject not available in field period
	NS	Subject Sick/ Incapacitated	Nonresponse: subject physically or mentally incapable of completing interview
	RB	Final refusal	Refusal - On at least two calls, the respondent refused to be interviewed or broke off during the interview and refused to continue.
	RD	Final refusal - Do Not Call	Final refusal. The R asked to be removed from calling. The case is not called for conversion.
	RG	Final refusal - Refused Gender	Final refusal. The R refused to provide their gender.

	RH	Hostile Refusal - Voxco	A refusal designated as hostile is finalized and not called for conversion.
	RM	Max Calls - Refusal	Questionnaire had an additional refusal code and has reached the maximum calling algorithm.
	RN	Inbound Refusal	An inbound caller refuses the interview. There is no refusal conversion. Attempt.
IN	IE	Ineligible Screener - no one 18+	Ineligible Screener. No one who lives in the HH is at least 18, or no one who uses the cell phone is at least 18.
	I4	Ineligible - Lives outside the US	Ineligible Interview. R lives outside the United States.
	OD	Duplicate case	Duplicate phone number or phone number for previously enumerated HH
	OO	Other Out of scope	Other out of scope - The questionnaire is out of scope and no other field code applies.
	NR	Non-Residential Phone Number	The number called was not a residential number. Included are businesses, institutions, agencies, modems, public facilities, vacation homes, group quarters.
	NW	Non-Working Phone Number	A call attempt reached a telephone company recording that indicated the telephone number is not working; or on three separate call attempts the case was a Questionable Ring (7). NW is a final code for screeners and a phone result and possibly final code for extended interviews.
	I1	Ineligible – less than 18 years	Ineligible Interview. R is not at least 18 years of age.
UNK	N1	Refielded NA - Voxco	A refielded Voxco NA case that has already cycled through a pre-determined number of calls.
	N2	Refielded NM - Voxco	A refielded Voxco NM case that has already cycled through a pre-determined number of calls.
	NA	No Contact -	The calling algorithm has been

| | | RNA only | fulfilled (time slices are filled) with no "human" or answering machine contact.   For List Sample Extd, NA can be a phone result or Mainrslt. |
| | NM | No Contact - Answering Machine in history | The calling algorithm has been fulfilled (time slices are filled) and only answering machine contact was made. For List Sample Extd, NM can be a phone result or Mainrslt. |