



U.S. Department of Justice
Office of Justice Programs
Bureau of Justice Statistics

Bureau of Justice Statistics Assessment of the Rape and Sexual Assault Pilot Test

Jeffrey H. Anderson, Ph.D., *BJS Director*
Allen J. Beck, Ph.D., *BJS Senior Statistician*

January 2021

The Bureau of Justice Statistics (BJS) funded the methodological research described in the third-party report, *Methodological Research to Support the National Crime Victimization Survey: Self-Report Data on Rape and Sexual Assault – Pilot Test* (NCJ 256011, BJS, January 2021). The purpose of the Pilot Test was to develop and test a new methodology for measuring rape and sexual assault, with the goal of collecting more accurate statistics on the nature and frequency of those crimes. The third-party report detailing the results of the Pilot Test does not provide official government statistics, nor does it provide nationwide statistics.

This assessment by BJS, published in connection with the third-party report, evaluates the methodology of the Pilot Test, and it provides context for interpreting data presented in the third-party report.

During testing, the Pilot Test’s methodology—which was developed and tested by Westat in collaboration with BJS—proved problematic in a variety of ways. It generated extremely high rates of rape and sexual assault that deviated profoundly from statistics published by the Federal Bureau of Investigation (FBI).

The Pilot Test surveyed a total of about 5,800 females ages 18 to 49 across five metropolitan statistical areas (MSAs) in 2014 and 2015: Dallas, Los Angeles, Miami, New York, and Phoenix. Through the Uniform Crime Reporting (UCR) Program, state and local law enforcement agencies convey to the FBI how many rapes are reported each year to law enforcement, including in those five MSAs. A comparison of the Pilot Test with the UCR, as well as with BJS’s National Crime Victimization Survey (NCVS), reveals notable disparities between the numbers generated by the Pilot Test and those published by the UCR and NCVS.

Comparison of UCR, NCVS, and Pilot Test numbers

Over the 2-year span from 2014 to 2015, the UCR published an annualized rate of 0.340 completed or attempted rapes reported to law enforcement per 1,000 persons in the five MSAs in question (with each

MSA weighted equally, as was the case in the Pilot Test).¹ The UCR summary-based counts of rape cannot provide information on how many victims were females ages 18 to 49; however, the FBI’s National Incident-Based Reporting System (NIBRS) does collect that information. NIBRS data, which are not nationally representative,² indicate that 42.4% of rapes reported to participating law enforcement agencies in 2014 and 2015 were committed against females ages 18 to 49, a group that made up 21.3% of the United States population during those years, per the Census Bureau.^{3,4} Therefore, females ages 18 to 49 were victimized at approximately twice the rate (199%) of the general population (42.4% versus 21.3%).

An estimated rate of completed or attempted rapes against females ages 18 to 49 in the five MSAs in question can be generated by applying the percentage of rape victims who were females ages 18 to 49 (42.4%) and the percentage of the population in that age range (21.3%) to the rape victim and population counts published in the UCR for the five MSAs. These calculations produce an estimated rate of completed or attempted rapes in those five MSAs of 0.677 per 1,000 females ages 18 to 49. This reflects the estimated number of completed or attempted rapes per 1,000 females ages 18 to 49 that law enforcement agencies indicated were reported to them.

The NCVS does not break down rapes or sexual assaults in specific MSAs by completed or attempted rapes reported to police. However, survey results from the NCVS MSA

¹This UCR rate is based on an average of 2014 and 2015 statistics for Miami, on 2014 statistics for New York because 2015 statistics were unavailable, on 2015 statistics for Dallas and Phoenix because 2014 statistics were unavailable, and on 2015 statistics for Los Angeles because most reporting in 2014 used the UCR’s legacy definition of rape and therefore was not comparable to reporting based on the updated definition.

²The percentage of law enforcement agencies participating in NIBRS was 35.2% in 2014 and 36.1% in 2015. See <https://www.fbi.gov/services/cjis/ucr/nibrs> for additional coverage information.

³See U.S. Census Bureau. (2020). *2014 Annual Social and Economic Supplements*. <https://www.census.gov/data/datasets/2014/demo/cps/cps-asec-2014.html>

⁴See U.S. Census Bureau. (2020). *2015 Annual Social and Economic Supplements*. <https://www.census.gov/content/census/en/data/datasets/2015/demo/cps/cps-asec-2015.html>

file, 2000-2015,⁵ found that on an annualized basis there were 6.509 rape or sexual assault victimizations per 1,000 female residents ages 18 to 49 in the five MSAs in question in 2014-15 (based on a simple average of the five MSAs, as in the Pilot Test).⁶ Over that same time period, in urban or suburban counties nationwide, 12.0% of the rape or sexual assault victimizations captured by the NCVS against female residents ages 18 to 49 were completed or attempted rapes reported to police.⁷ Applying that percentage to results from the five MSAs, the estimated rate of completed or attempted rapes reported to police in those areas was 0.781 per 1,000 female residents ages 18 to 49—or about 15% higher than the UCR number.⁸

In comparison, the annualized rate generated by the Pilot Test was 6.345 completed or attempted rapes reported to police in the five MSAs per 1,000 female residents ages 18 to 49.⁹ That is 9.4 times the rate calculated based on available UCR data. Based on these numbers, one of two things must be true: either state and local law enforcement agencies, in connection with the FBI, recorded only 11% of reported rapes (0.677 out of 6.345), or else the Pilot Test generated inflated estimates—perhaps by nearly an order of magnitude.¹⁰

⁵To facilitate access to subnational data, BJS created extract files for the 52 largest metropolitan statistical areas (MSAs) covering the 2000-2015 survey years. A third-party report describing this methodology and the accompanying extract files is currently under review at BJS.

⁶Because these data are derived from smaller geographical areas and therefore smaller sample sizes than the national NCVS, they are typically aggregated over multiple years to bolster reliability. For comparability purposes, this estimate is presented as a simple average of rates in the five MSAs over a 2-year period. This estimate should be interpreted with caution. Its component estimates are each based on 10 or fewer sample cases, or the coefficient of variation is greater than 50%.

⁷See Bureau of Justice Statistics. (2020). *National Crime Victimization Survey, Concatenated File [United States], 1992-2019 (ICPSR 37689)*. NACJD. <https://doi.org/10.3886/ICPSR37689.v1>

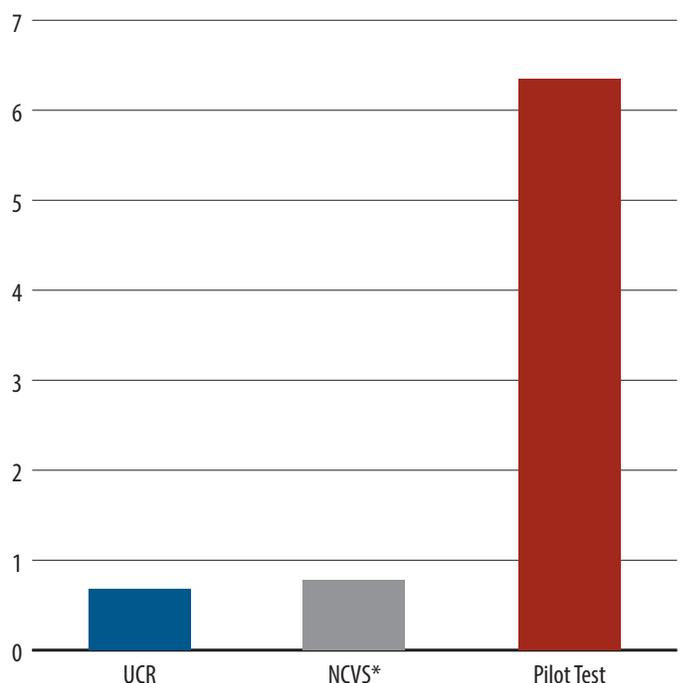
⁸This rate was heavily influenced by the effects of series victimizations (multiple, largely indistinguishable victimizations experienced by the same victim, which are capped at 10 victimizations), without which the rate would only have been about one-third as high—or about three-eighths of the UCR number.

⁹This statistic, as well as some others provided herein, is not published in the accompanying Pilot Test report but was computed by BJS from the Pilot Test data. The standard error for the estimate of 6.345 is 1.364. This rate includes 17 respondents (0.3% of the total) who completed the behaviorally specific questions but not the detailed incident form. All subsequent numbers referencing the Pilot Test were computed by BJS from the Pilot Test data, unless they are cited as having come from the Pilot Test report.

¹⁰Some researchers in the 1970s and 1980s suggested that 30% to 50% of calls to the police either did not get recorded in the final UCR counts or were re-classified as different types of crimes; however, improvements in reporting processes over the past few decades have likely reduced that percentage (Westat, *Rape and Sexual Assault Pilot Study 2014 - 2015*, from briefing to BJS on key findings, April 17, 2018). Even in the event that this outer-limits number of 50% of calls not being recorded as reported has been true in recent years, the Pilot Test's rate of rapes reported to police would still outpace the UCR rate of rapes reported to police by nearly a factor of five.

The Pilot Test's numbers of crimes reported to police were this high despite the fact that its percentage of crimes reported to police was quite low. In general, roughly 1 in 3 rapes and 1 in 4 sexual assaults captured by the NCVS are reported to police, compared to only about 1 in 7 rapes and 1 in 25 sexual assaults in the Pilot Test. Therefore, in comparison to the NCVS, the Pilot Test's number of unreported rapes or sexual assaults was much higher even than its number of reported rapes or sexual assaults. Overall, for all rapes or sexual assaults against female residents ages 18 to 49 in the five MSAs in 2014-15—whether reported to law enforcement or not—the Pilot Test's rates were 12.4 times higher than those

Completed or attempted rapes reported to police per 1,000 females ages 18 to 49 in five MSAs, annualized rate for 2014-2015, per the UCR, NCVS, and Pilot Test



Note: See text for numbers and specifics. The Pilot Test does not provide official government statistics.

*This rate is derived by multiplying the total rate of rape or sexual assault victimization against females ages 18-49, restricted to 5 MSAs, 2014-2015 (6.509 per 1,000 females ages 18-49, with standard error 4.890), by the nationwide percentage (12.0%) of rape or sexual assault victimizations against females ages 18-49 in urban or suburban areas that were reported to police (using the NCVS's old definitions of urban and suburban areas). In comparison, the 2014-2015 NCVS reports a national victimization rate of completed or attempted rapes reported to police of 0.464 per 1,000 females ages 18-49 (standard error 0.141). Over the most recent decade (2010-2019), the NCVS reports a national victimization rate among the larger population (including both sexes) of 0.228 (standard error 0.025) completed or attempted rapes reported to police among persons age 12 or older, which is five-eighths (62.5%) as high as the 0.365 rate reported by the FBI's UCR (for all ages) over the same period.

Source: Federal Bureau of Investigation, Crime in the United States, 2014 and 2015; Bureau of Justice Statistics, National Crime Victimization Survey Metropolitan Statistical Area file, 2000-2015; and Westat, *Methodological Research to Support the National Crime Victimization Survey: Self-Report Data on Rape and Sexual Assault – Pilot Test*, 2018.

reported in the NCVS (81.0 versus 6.5 rapes or sexual assaults per 1,000 female residents ages 18 to 49).¹¹

When BJS sought applicants in 2011 to undertake the Pilot Test project, the solicitation stated that “because under-reporting of victimization is assumed to be substantial, designs that yield more reports of rape and sexual assault would be considered ‘better,’” all other things being equal.¹² However, the true test of a new survey methodology is not whether it generates higher numbers but whether it generates more accurate numbers. The accuracy of the methodology utilized by the Pilot Test is called into question by the nearly tenfold disparity between the number of rapes that the Pilot Test says were reported to law enforcement and the number of rapes that law enforcement says were reported to law enforcement.

The Pilot Test’s numbers are also high in relation to other numbers. Before the Pilot Test was conducted, its design assumptions predicted that the portion of females ages 18 to 49 in the five MSAs who were victims of rape within the prior year would be 0.45%. The survey did not perform as expected. Its actual results were more than 7.5 times as high—3.4%. (The portion of sexual-assault victims was only about 70% as high as expected: 2.19% versus an expected 3.15%. This was largely because the design assumptions predicted seven times the prevalence rate for sexual assault as for rape, far in excess of the less-than-one-to-one ratio between sexual assaults and rapes usually found in the NCVS.)¹³

The Pilot Test’s numbers for completed rape were also high in relation to numbers from the National Intimate Partner and Sexual Violence Survey (NISVS). The NISVS is a survey that (1) uses a “health frame” rather than a crime frame, (2) uses broader definitions in many cases than those based on the legal definitions of crimes, and (3) uses a single-step process that does not substantiate

affirmative responses. Accordingly, the NISVS is known for its high estimates of victims. As the Pilot Test report notes, the NISVS said that 0.8% of females ages 18 to 49 in 2011 were victims of completed rape using force during the prior year (nationally), whereas the Pilot Test (based on the five MSAs) reported that 1.7% to 1.9% of females ages 18 to 49 in 2014-15 were victims of completed rape using force during the prior year.¹⁴

The Pilot Test (like the NISVS but unlike the NCVS) was unbounded, meaning that no earlier version of the survey was administered to establish a baseline in time. It was therefore susceptible to “telescoping,” whereby victims report crimes that occurred outside of (before) the reporting timeframe (in this case, outside of the prior year). No adjustment was made to account for the survey’s being unbounded, even though telescoping has been shown to inflate results greatly. In addition, the Pilot Test, which was administered by Westat, a private corporation, had much lower response rates than the NCVS, which is administered by the Census Bureau. These low response rates may, in turn, have inflated the Pilot Test’s estimates, as it could be that victims were more likely than non-victims to respond. As the Pilot Test report says, “If non-victims are more likely to participate on the NCVS relative to the RSA Pilot Test, then some of the difference between the estimates may be related to non-response bias.”¹⁵

Methodological concerns

Beyond these basic limitations, the Pilot Test’s methodology proved problematic in the following eight ways:

I. Pressure to respond. Women were told that they would be paid to take the survey (\$20), unlike for the NCVS, and then were asked a large battery of questions solely about rape and sexual assault. As a result both of the payment and of the large battery of questions, some respondents may have felt pressured to say yes to at least some of these items. As is stated in the Pilot Test report, “By asking multiple questions about related behavior, respondents may feel pressure to answer in the affirmative.”¹⁶ The report adds, “Survey methodologists have found that asking more questions that mention the targeted or related behaviors will produce higher rates of reporting.”¹⁷

¹¹See table 7-12 (p. 158) in Cantor et al. (2018). *Methodological Research to Support the National Crime Victimization Survey: Self-Report Data on Rape and Sexual Assault – Pilot Test*. Westat (hereafter “Pilot Test report”). The standard error for the Pilot Test’s estimate of 81.0 was 4.9. See table 7-14 (p. 18) in Canton et. al. (2018). *Methodological Research to Support the National Crime Victimization Survey: Self-Report Data on Rape and Sexual Assault – Pilot Test: Appendices*. Westat (hereafter “Pilot Test appendices”). The standard error for the NCVS estimate of 6.509 was 4.890.

¹²See Bureau of Justice Statistics. (2011). *Methodological Research to Support the National Crime Victimization Survey: Self-Report Data on Rape and Sexual Assault – Pilot Test Solicitation* (Grants.gov No. 2011-BJS-3011). https://www.bjs.gov/content/pub/pdf/cvssdrsapt_sol.pdf

¹³See table C3-1 (p. 343) in Pilot Test appendices.

¹⁴See table 7-15 (p. 166) in Pilot Test report.

¹⁵See p. xlvii in Pilot Test report.

¹⁶See p. xl in Pilot Test report.

¹⁷See p. xlvi in Pilot Test report.

II. Not framed as a crime survey. The Pilot Test was not administered as a crime survey. The Pilot Test report says that the “Pilot Test was introduced as a survey about health and safety, and the questions do not refer to any acts as being criminal.”¹⁸ This is problematic for a survey about crime, specifically about the crimes of rape and sexual assault. Since the survey did not portray events in a crime context, many women reported actions that they did not consider to be criminal; that they did not regard as rape or sexual assault.

III. Not incorporating women’s views on crime.

Women’s own views about whether they had experienced a crime were disregarded by the survey’s classification scheme. This raises concerns about whether the survey adequately took into account women’s thoughts about the nature of the incidents in which they were involved. Indeed, most incidents that the survey classified as rape or sexual assault were not thought by respondents to have been rape or sexual assault. Less than one-third (32%) of incidents that the survey described as rape or sexual assault were thought by the women involved, at the time those acts occurred, to be crimes.¹⁹ Breaking out rape and sexual assault separately, 41% of incidents that the survey classified as rapes, and 19% of incidents that the survey classified as sexual assaults, were regarded as crimes by those who actually experienced them. Among certain types of crimes, 18% of incidents that the survey classified as attempted rape in the form of oral sex, 11% of incidents classified as attempted rape in the form of digital penetration,²⁰ and 0% of incidents classified as attempted or threatened sexual assault involving inability to consent, were regarded as crimes by those who experienced them.²¹

¹⁸See p. xlv in Pilot Test report.

¹⁹The remaining 68% were not all incidents in which the women said that they thought it was not a crime. In 94.5% of all rape or sexual assault incidents, women answered either yes or no as to whether they thought it was a crime at the time of the incident. In the other 5.5% of incidents, women either said they did not know (0.7%), refused to answer (0.4%), or did not give any response to the question (4.4%).

²⁰This refers to attempted rape in the form of oral sex where the incident did not also involve attempted rape in the form of vaginal or anal sex, and to attempted rape in the form of digital penetration where the incident did not also involve attempted rape in the form of vaginal, anal, or oral sex.

²¹More women thought that these incidents were crimes after taking the survey, but this may have been a product of the survey’s having led them. If women said an incident was a crime, the survey did not ask them why. If they said it was *not* a crime, then the survey asked them why they did not think it was a crime, taking them through a series of reasons why they might not have thought it a crime. It then gave them one more chance, asking, “Now, looking back, do you consider this incident to be a crime?” Some respondents answered differently at that point.

IV. Classification inconsistencies. The Pilot Test survey involved three parts: behaviorally specific questions (BSQs), a detailed incident form (DIF), and in some cases a narrative. Of the 14 BSQs, only the first 5 asked about use of force or inability to consent. These 14 BSQs ranged from “has a male used force or threats of force to make you have vaginal sex against your will?” (among the first 5 questions) to “has anyone, male or female, kissed you in a sexual way against your will?” (among the subsequent 9 questions).²² The DIF, administered to the roughly one-eighth of respondents who answered yes to a BSQ, asked a wide range of more detailed questions.²³ The narrative allowed respondents to describe incidents in their own words.

These three parts of the Pilot Test produced three different levels of classifications: classification from the BSQs; initial classification from the DIF; and final classification (based on the DIF, including any answers to open-ended DIF questions, and sometimes a narrative).

There was remarkably low consistency across these three levels of classification. Among incidents characterized in the final classification as completed vaginal rape using force, just over a one-third were consistently classified across the three classifications—36%. For final classifications of forcible completed anal rape (31%) and completed forcible rape in the form of digital penetration (35%), the level of agreement was similarly low.

It is problematic that approximately two-thirds of these final classifications disagreed with classifications based on the DIF or (more often) the BSQs. While these inconsistencies reinforce the importance of having a screener (BSQs), DIF, and narrative in such a survey, they also highlight the need to improve the questions and the classification scheme, especially on the DIF. The final classification was driven principally by the DIF, which suggests that the DIF questions were written or utilized in such a way as to classify many incidents as rape that would not have been classified as such based on the BSQs or the narratives.

V. Inconsistencies across interviews. Of respondents who were interviewed in the Pilot Test survey, approximately 15% were re-interviewed a few weeks later to see whether the survey would produce consistent results upon repeated administrations. Once again, remarkably low consistency was found. Among those whom the survey classified as victims of completed

²²See table 2-1 (p. 27) in Pilot Test report.

²³See p. 298 in Pilot Test appendices.

rape, half (50%) were classified as victims of completed rape based on the re-interview, and half (50%) were not. Among those whom the survey classified as victims of completed rape due to inability to consent, fewer than one in four (23%) were classified as victims of completed rape due to inability to consent based on the re-interview.

These inconsistencies may have resulted in part from DIF questions that asked about incomplete pieces of incidents, which when combined did not provide complete pictures of those incidents conducive to yielding clear classifications. Some assumptions made to fill in the gaps may have inflated the estimates. This suggests that classification based on a series of close-ended questions about portions of incidents, pieced together to represent the whole, is difficult.

VI. Problems with inability to consent. The Pilot Test report states that “the two elements that define an incident as a crime” are “the use of physical force” or the “inability to consent” due to alcohol intoxication or drug use.²⁴ The survey’s classifications based on inability to consent, however, sometimes differed from what the respondents said about their own ability to consent. In 21% of all incidents that the survey classified as completed rape due to inability to consent because of alcohol or drug use, the survey respondent said she was *not* unable to consent.

At other times, the survey’s classifications based on inability to consent led it to characterize actions as crimes that the legal system would unlikely regard as crimes. One respondent to the survey said that while she was drunk, a man was “trying to have sex with me. I told him to stop and he did.” Elsewhere on the survey, she said that no force had been used, that she had not been the victim of a crime because he “stopped when [I] resisted,” and that he had stopped “immediately.” Nevertheless, because she was drunk, the survey classified this incident as attempted rape involving inability to consent.

VII. Inconsistent answers overriding narratives and other answers. Seemingly inconsistent answers on the DIF were allowed to override the BSQs and the narrative, as well as respondents’ own thoughts on whether they had been the victim of a crime. This seems particularly true for question D1a, the first question on the DIF.

For example, one respondent did not answer yes to any of the five BSQs that asked about rape (completed penetration involving force or inability to consent). She instead answered yes to the BSQ that said someone

had “fondled, groped, grabbed, or touched you against your will.” In her narrative, she described the following incident: “I was standing on a bus...when a man came up behind me and put his hand on my butt, using his thumb to feel around it. I moved my body away from him and stood facing towards him. He avoided my eye contact.” On the DIF, she answered no to any force being involved except for reiterating that the man “grabbed me without permission.” However, on DIF question D1a—“Did the person threaten to, try to, or actually put his penis in your vagina when you didn’t want it to happen?”—she picked the answer that said “the person actually did do this.” As a result, the Pilot Test survey classified this incident as completed rape using force—despite the fact that this classification contradicted the BSQs, the narrative, and the DIF question about use of force, all of which were in agreement.

Another respondent answered yes to a BSQ about whether someone had “kissed you in a sexual way against your will.” (The registered answer on the BSQs always represents the most serious incident that the person experienced, so she had not said yes to any of the BSQs about rape.) Her complete narrative was as follows: “I was at work and a coworker came up and gave me a hug and I thought he was just being friendly and then he kiss[ed] me on the neck and then a little bit later in the day he put my hand on his crotch, and then blocked my way so that I couldn’t go to where I needed to go, and I said I have work to do and he moved.” On the DIF, she said that she had not been the victim of a crime. The Pilot Test classified this incident as a completed rape using force—because in response to question D1a (“Did the person threaten to, try to, or actually put his penis in your vagina when you didn’t want it to happen?”), she picked the answer that said yes, “the person actually did do this.”

An additional respondent answered on the BSQ that someone had “groped” her against her will. Her complete narrative was as follows: “I was at a BBQ and my friend I hadn’t seen since high school came up behind me and grabbed my butt. And that was it. I told him to stop.” On the DIF, she said that he had stopped immediately and that she had not been the victim of a crime. The Pilot Test classified the incident as completed rape using force, even though that contradicted both her narrative and her answers to the BSQs. It did so because in response to question D1a (“Did the person threaten to, try to, or actually put his penis in your vagina when you didn’t want it to happen?”), she picked the answer that said “the person actually did do this.”

²⁴See p. xl in Pilot Test report.

In a final example, a respondent said yes to the BSQ about groping (again, only the most serious incident is noted on the BSQs, so she had not said yes to any of the BSQs involving rape). In her narrative, she said that she was coming back from the city with her kids “and it was very tight on the train, rush hour kind of situation,” and a man “grabbed at me and tried to kiss me and then just as I was fighting him and cursing him out, he jump[ed] off the train like it was a joke....So that’s what happened.... He took advantage and grabbed as much as he could in a minute, not even, and it was like before it was started it was over.” The Pilot Test survey classified this as completed rape using force—in contradiction to the BSQs and narrative—because in response to D1a (did he “put his penis in your vagina?”), she said that the person “actually did do this.”

As these examples suggest, the wording of question D1a may have been problematic.²⁵ In full, the phone-interview version of question D1a read as follows:

Did the person threaten to, try to, or actually put his penis in your vagina when you didn’t want it to happen?

- *Say one if the person verbally threatened to do this but did not physically try to do it*
- *two if the person physically tried to do this but did not actually do it*
- *three if the person actually did do this or*
- *four if this did not happen.*²⁶

It could well be that this question is overly long and confusingly worded. Perhaps by the time some respondents heard the answer, “Say...three if the person actually did do this,” they had forgotten the stem of the question and were treating this answer as if it were simply affirming that what they had described elsewhere, on the BSQs, was actually true. Regardless, answers to this question were allowed to override other information acquired on the survey and, thereby, to cause incidents to be classified as rapes—in contradiction to the narratives, the BSQs, and sometimes other parts of the DIFs.

VIII. Problems with use of force. The survey’s use-of-force indicator was not clearly tied to the crime in question and often was not indicative of criminal intent. Despite this, it was used as an indicator of criminality.

²⁵On more than one occasion, a respondent said, in response to question D1a, that another female either had, had tried, or had threatened to “put his penis in your vagina.”

²⁶See p. 298 in Pilot Test appendices.

The DIF’s only indicator of use of force was question D4. That question had four parts (in addition to two open-ended questions), which asked whether the person did any of the following: “Hold or pin you so you had difficulty moving”; “Use a weapon, or threaten to use a weapon”; “Physically attack you or threaten to attack you, but not with a weapon”; or “Physically attack, or threaten to attack, someone else.”²⁷ In less than one-third (32%) of incidents classified by the survey as forcible rape or sexual assault, the respondents answered yes to any of the three questions about an attack or threatened attack (with or without a weapon).

Question D4 was not necessarily tied to the moment in question or to the crime in question. Take a respondent who said yes to the question about a person trying to kiss her when she did not want it to happen, and who also said yes to the question (D1a) about the person trying to engage in sexual intercourse with her at some point during that same incident. If the respondent answered that the person held or pinned her, with the respondent meaning that this happened during the attempt to kiss her, the classification scheme would apply that answer to the second question as well—the question about trying to engage in sexual intercourse. As a result, it would categorize the incident as attempted rape—even if no holding or pinning occurred in the context of trying to get her to engage in sexual intercourse; even if the person had no intention of trying to *force* her to engage in sexual intercourse; and even if she entirely agreed with this assessment of the event. As the Pilot Test report says, “[I]n situations where there is unwanted sexual contact...and an attempt/threat of penetrative contact..., it is not possible to assess whether the threat or attempt was associated with the force or threat of force.”²⁸

In short, the survey never directly asked women whether someone had attempted to rape them, and it did not use their answers about whether they had been the victim of a crime. But the survey often treated holding or pinning—even if that holding or pinning occurred at a different chronological moment than an attempt to get someone to engage in sexual intercourse—as evidence of intent to rape.

Notably, in incidents that the survey classified as rape or sexual assault based on the respondents having answered yes to the question about holding or pinning (and no to the questions about being attacked, with

²⁷See p. 301 in Pilot Test appendices.

²⁸See p. 303 in Pilot Test report.

or without a weapon), the women involved regarded themselves, at the time of the incident, as having been victims of any sort of crime in only 25% of the incidents. In comparison, in 80% of the incidents that the survey classified as rape or sexual assault based on the respondents having answered yes to the question about the use of a weapon, the women involved regarded themselves as having been the victim of a crime.

In each of the examples in section VII, it was not merely the inconsistent answer to D1a that caused the incident to be classified (in contradiction to the narratives and the BSQs) as an attempted or completed rape. It was this in combination with an affirmative answer about use of force on D4. But in each of the instances described, the use of force indicated on D4 was consistent with, and was likely linked to, the actions described in the narrative, rather than to any effort to engage in sexual intercourse. Despite this, the use-of-force indicator was used to classify these incidents as attempted or completed rapes.

In the instance where the woman was groped on a bus, the sole use of force she reported was that the offender “grabbed...without permission,” which confirms the groping. Yet this use of force, in combination with her answer to D1a, was used to classify the incident as a completed rape. In the instance where the woman was blocked by her coworker at work, who then moved when she said that she had work to do, the sole use of force she reported was that he had held or pinned her so that she “had difficulty moving,” which could refer to the blocking. In the instance where the woman was groped at a barbeque, the sole use of force she reported was that the offender had held or pinned her, which could refer to the groping. In the example where the woman was groped on a train, the sole use of force she reported was that the offender physically attacked her but not with a weapon, which was consistent with her report that he “grabbed as much as he could.” Yet these uses of force, in combination with the answers to D1a, were used to classify the incidents as completed rapes—even though this was contradicted both by the women’s narratives and by their answers to the BSQs.

To provide other specific examples, a respondent answered yes to a BSQ about whether someone “tried” (with no indication of force) “but did not succeed at making you have any type of sex against your will.” Her narrative was as follows: “[M]y boyfriend stayed over the night before. It was the morning, he...started touching and fondling me. I sort of resisted, not, you know, not forcefully I didn’t push his hand away but I did like roll away and he kept me there with his hand. He tried to, or he did, put his penis in my vagina and started to

have intercourse...and I told him to stop and he did.” She said on the DIF that he stopped “immediately,” and two reasons she gave for why she did not view herself as having been the victim of a crime were that she “didn’t think [he] knew what [she] wanted to happen” and that he “stopped when [she] resisted.” Nevertheless, the survey took her answer to D4 (holding or pinning), along with her answer that her boyfriend had attempted to engage in sexual intercourse with her, as evidence of intent to rape by force, and the incident was classified as completed rape by force.

In another instance, a respondent answered yes to the same BSQ as the person above. She said that she was “in a relationship” with a man she “loved,” and her narrative conveyed that he was drunk, she “had his little girl in a buggy,” and “he just came up behind me and grabbed by butt and then pulled my hair and for me doing that in public is a big no-no.” Elsewhere on the DIF, she said, “We were in a public place at grocery.” She did not think she had been the victim of a crime. She answered no to the four D4 questions about use of force but reiterated in an open-ended D4 question that her boyfriend had engaged in “pulling hair.” Because she responded affirmatively to two parts of question D1 (did he “put his penis in your vagina?—and did he “put his penis in your anus”), this hair-pulling was used to classify the incident in the grocery store as attempted vaginal and anal rape using force.

These examples point to the problem of interpreting use of force in a binary manner, largely divorced from timing and context. They suggest the importance of more fully taking into account respondents’ perceptions of the circumstances, so as to better reflect respondents’ views of the incidents.

Further information gleaned from the Pilot Test

While the Pilot Test revealed a wide range of methodological concerns, it also proved informative in at least four additional ways:

- The Pilot Test found that the relative anonymity of an audio computer-assisted self-interview (ACASI) did not produce rates for rape and sexual assault that were different to a statistically significant degree from those produced by the more interpersonal interaction of a computer-assisted telephone interview (CATI). The Pilot Test report says that “it was expected that the increased anonymity of a self-administered survey mode, such as ACASI, would produce higher estimates than an interviewer-administered survey mode, like CATI,” but “there were very few differences.”²⁹

²⁹See p. xxxvi in Pilot Test report.

- Relatedly, and also unexpectedly, the Pilot Test, with its increased privacy versus the NCVS (especially on the ACASI), reflected a *lower* percentage of rapes or sexual assaults committed by intimate partners such as spouses, ex-spouses, boyfriends, and girlfriends. The Pilot Test report states that a higher percentage of “Pilot Test incidents occurred at a friend’s house,” while a higher percentage of “NCVS incidents occurred at the respondent’s home”—where the NCVS survey often takes place.³⁰
- The Pilot Test concurred with prior analysis by BJS that college students are not more likely to be victims of rape or sexual assault than those of the same age who are not college students.³¹ The Pilot Test report says that “women enrolled in college did not exhibit higher rates of victimization than non-college students,” and “college enrollment did not increase risk.”³²
- The Pilot Test confirmed the importance of a two-step interview process for reducing false positives, which is “a significant advantage for the NCVS,” according to the Pilot Test report, “given the goals of counting criminal events.”³³

Summary

The Pilot Test produced extremely high rates of completed or attempted rapes reported to law enforcement, which deviated—by nearly an order of magnitude—from those based on the number of completed or attempted rapes that law enforcement agencies said were reported to them. In comparison to the NCVS, the Pilot Test’s numbers of rapes and sexual assaults *not* reported to law enforcement were even higher than its numbers of rapes and sexual assaults reported to law enforcement.

³⁰See p. xlviii in Pilot Test report.

³¹See <https://www.bjs.gov/content/pub/pdf/rsavcaf9513.pdf>.

³²See p. xlvii in Pilot Test report.

³³See p. 10 in Pilot Test report.

These extremely high rates resulted from a survey methodology that was problematic: The Pilot Test survey asked a wide battery of questions about rape and sexual assault, which may have led some respondents to feel pressure to respond to at least some questions in the affirmative. It did not present these questions in a crime context, and it disregarded women’s answers about whether they thought they had been the victim of crime. It sometimes allowed inconsistent answers to a particular question to override answers both from other questions and from narratives about the events involved. The Pilot Test’s classifications based on inability to consent sometimes defied respondents’ own assessments of their ability to consent and often led to something being labeled as a crime that the legal system would unlikely regard as a crime. The Pilot Test’s use-of-force measure was insufficiently tied both to the moment in question and to the crime in question. This use-of-force measure often did not provide evidence of criminal intent—more specifically, of intent to rape or sexually assault—but was utilized as if it did. Finally, there was remarkably low consistency across the survey’s three levels of classification, and also between classifications based on original interviews and those based on re-interviews.

Despite these serious methodological concerns, the Pilot Test provided a number of meaningful insights about the comparative results generated by different survey modes and among different survey populations. The lessons learned from the Pilot Test have helped inform BJS’s ongoing efforts to provide the most accurate possible statistics on rape and sexual assault, as well as other crimes. They have underscored the importance of ensuring that the views of respondents are fully and properly measured, rather than producing survey results with which respondents would often disagree.³⁴

³⁴Nearly two-thirds (63%) of incidents that the Pilot Test classified as rape or sexual assault were not regarded as rape, sexual assault, or any other crime, at the time of the incident, by those who actually experienced them.



The Bureau of Justice Statistics of the U.S. Department of Justice is the principal federal agency responsible for measuring crime, criminal victimization, criminal offenders, victims of crime, correlates of crime, and the operation of criminal and civil justice systems at the federal, state, tribal, and local levels. BJS collects, analyzes, and disseminates reliable statistics on crime and justice systems in the United States, supports improvements to state and local criminal justice information systems, and participates with national and international organizations to develop and recommend national standards for justice statistics. Jeffrey H. Anderson is the director.

This report was written by Jeffrey H. Anderson and Allen J. Beck. BJS statisticians produced and verified estimates in this report.

Edrienne Su copy-edited the report. Carrie Epps-Carey and Morgan Young produced the report.

January 2021, NCJ 256009



NCJ 256009

Office of Justice Programs
Building Solutions • Supporting Communities • Advancing Justice
www.ojp.gov