Appendix B:
Survey and Sampling Methodology

Survey and Sampling Methodology

The 2009 survey of the National Capital Region (NCR) was conducted by the Center for Survey Research (CSR) at the University of Virginia using a Computer Assisted Telephone Interviewing (CATI) system and a triple frame sample design. A discussion of the general methodology appears in the introduction of this report. This appendix provides additional details on how the questionnaire was developed, how the sample was created, how the survey was administered, how the data was weighted, and how statistical testing was used to evaluate the results.

Sample

Traditionally, CSR and other research organizations have employed random-digit dialing (RDD) to reach a random sample of the households in a defined geography. RDD methods have been used in the past to produce a more representative sample of the population than other available sampling methods. Households are selected for contact at random so that all households with a working telephone can be reached. Listed and unlisted residential telephones have equal probability of being included in an RDD study. In recent years, the growth in the use of cell phones has been increasingly evident during data collection. A good representative sample should allow contact with a balanced cross-section of the population, and require minimal adjustment of the final results to appropriately reflect the entire population.

Recent research, including a pilot study of cell phones fielded by CSR in January-February 2008[1] and subsequent work in Virginia counties**,** has shown not only the feasibility of cell phone calling, but demonstrated that the demographics of those reached via cell phone are quite different from those currently reachable via landline phone.

Analysis of the occurrence of key demographic groups through the years has shown decreased representation of key groups, particularly younger respondents in landline RDD samples. Including cell phone contact serves to increase the under-represented groups in the final sample and ensures that completed interviews accurately reflect the actual population of the target areas. Our 2008 cell phone experiment demonstrated that using a $10 incentive with cell phones increases the completions per hour rate and makes calling more affordable. However, it is still 2-3 times more expensive to call cell phones than landlines because of the inefficiency of the number blocks, even compared to list-assisted random digit dialing. To balance the added cost of cell phone calling, the sample was augmented by numbers taken randomly from directory listings. Listed sample is cheaper to administer because the proportion of numbers that are eligible households is much higher and time does not need to be devoted to screening out ineligible or non-working phone numbers.

In summary, a landline RDD (Random B) sample of 15,112 telephone numbers (49% of the total) randomly generated from five-digit call groups known to be in operation in the NCR, and a second, general directory-listed sample from electronic white pages of 6,127 telephone numbers (20% of the total) were combined with a cell phone sample of 9,585 numbers (31% of the total) in an effort to ensure greater targeting of harder to reach populations and geographies. The cell phone numbers were randomly drawn from telephone exchanges assigned to cell phone usage at the rate centers which serve the localities in the study area. The cell phone sample does not attempt to distinguish working and nonworking number blocks. Respondents contacted by cell phone were offered a $10 gift card to either Target or Wal*Mart stores as partial compensation for any telephone charges incurred by participating in the interview.

[1] Abdoulaye Diop, Young-Il Kim, John Lee Holmes, and Thomas M. Guterbock. *Prince William County Cell Phone Pilot Survey [A Supplement to the 2007 Citizen Satisfaction Survey]: Summary Report of Results.* Center for Survey Research, March 2008.

All samples were purchased from Survey Sampling, Inc. of Fairfield, CT, a commercial sampling company that uses state-of-the-art methodologies. Table B-1 on the following page summarizes the sample purchased and the numbers of completions for the different sample types.

Table B-1: Summary of Survey Sample Types Used, 2009

Phone Type	Sample	(%)	Completed	(%)	Ratio (sample: completes)
RDD – no address	7481	(24.3%)	349	(13.1%)	21:1
RDD – address match	7631	(24.8%)	968	(36.4%)	8:1
Listed	6127	(19.9%)	898	(33.8%)	7:1
Cell_$10	9585	(31.1%)	442	(16.6%)	21:1
TOTAL	**30824**		**2657**		**12:1**

Telephone surveys risk biases owing to variation among members of a household in the likelihood of answering the telephone. For example, persons who do not work may be more likely to be available to answer the phone than are those who are employed. Various methods have been developed to randomize respondents within households in order to reduce these biases. CSR used a "minimally intrusive method" which combines random selection (between two adults) by computer with the "last-birthday" method (if a household has three or more adults), in which interviewers ask to speak to the adult in the household who had the most recent birthday or, if last birthday is unknown, with the Kish selection process of enumerating first names or initials of eligible household members for random selection by the computer.[2] This protocol was applied to all households reached via the RDD or listed samples. Cell phone adults, however, were considered to be sampled as individuals. Prior research by others has shown that the percentage of cell phones actively shared by more than one adult is low and that it is very difficult in practice to accomplish a 'hand-off' of the cell phone from one adult to another randomly selected user of the phone.[3] Therefore, no within-household selection was attempted in the cell phone interviews for this study.

Random and cell phone sample can include numbers that fall outside the boundaries of the NCR, so each number was screened for location at the beginning of the interview to exclude ineligible contacts. Persons under 18 years of age were not interviewed.

Questionnaire Development

Significant work went into the development of a questionnaire that would serve the many and various needs of the project stakeholders. This section will describe the steps taken that resulted in the final survey instrument.

- Literature review
- Behavioral-focused workshop with grant participants
- Questionnaire design
- Evacuation behavior modeling
- Focus groups

[2] Programmed by CSR into the CATI system based on the method's description in Louis Rizzo, J. Michael Brick and Inho Park "A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys," *Public Opinion Quarterly*, Vol. 68, No. 2 (2004), pp. 267-274

[3] J. Michael Brick, W. Sherman Edwards, and Sunghee Lee."Sampling Telephone Numbers and Adults, Interview Length, and Weighting in The California Health Interview Survey Cell Phone Pilot Study." *Public Opinion Quarterly (* 2007) 71: 793-813.

Literature review

CSR conducted a focused literature review on studies of evacuation behavior that seemed applicable to the NCR behavioral study. The review eventually included ten studies, with summaries of methods and findings. The review was crucial in helping CSR survey researchers to understand the current state of research on the issues.

Behavioral-focused workshop with grant participants

A major impetus to the success of the questionnaire development and ultimately to the success of the study was to bring together representatives from the partner states and organizations. A workshop was convened at the Boar's Head Inn near Charlottesville, Virginia on March 25, 2009.

The intent of the workshop was to present the competing needs and desires of the participants and to try to gain some consensus on a broad outline of the questionnaire content for the 2009 survey. Major goals for the study included decisions about which events should be studied and how to present them to respondents, the level of geographic specificity that would yield sufficient responses for analysis and determination of what other variables should be included in the survey to help understand evacuation behaviors.

Key decisions made as a result of the workshop included focusing on radiologic dispersion or "dirty" bombs and on anthrax events. Shelter in place and evacuation behaviors were of interest, particularly focusing on how people made a decision between the two possible choices.

The use of escalating hazard levels described in hypothetical scenarios was proposed, beginning with a minimal threat and ratcheting up to what could be considered a catastrophic level for some respondents.

Questionnaire design

CSR started with the script that had been used in a June 2005 report, *Community Shielding in the National Capital Region: A Survey of Citizen Response to Potential Critical Incidents.*[4] This served as a basis for a conceptual outline of the topic areas that had proven to be of interest to the team.

Shortly after the workshop the team decided to drop the anthrax scenario because it was too difficult to create one that would be plausible, accurate and simple enough to communicate to respondents in a telephone interview. Thereafter, dirty bomb scenarios were the focus of the survey.

Once the conceptual outline was ratified by the team, CSR developed a draft questionnaire. CSR's 2005 study of community shielding was the source of some questions, while other questions came from other surveys or were developed especially for use in this survey. The draft questionnaire went through several revisions with two main themes at its core: evacuation behavior modeling, and factorial scenario design.

Evacuation behavior modeling

Research on evacuation behavior suggests that a multi-step process is involved as a person evaluates whether to stay or go in response to an event. This iterative process means the decision can change. Individuals may bring to the process not only current information about the event, but also past experiences with events, disaster training they might have had, self-perception about their role in the community, socio-economic, contextual or attitudinal attributes, along with many other possible factors.

The questionnaire needed to account for these various steps in the process by which an individual makes the decision to stay or go during an emergency. We are grateful to Professor Joseph Trainor of the University of Delaware for his insights into this process.

---

[4] Monnica T. Williams, Gregory B. Saathoff, Thomas M. Guterbock, Anna McIntosh, and Robin Bebel. *Community Shielding in the National Capital Region: A Survey of Citizen Response to Critical Incidents.* Center for Survey Research, University of Virginia, June 2005. The study was funded by the Department of Homeland Security via a grant to George Mason University.

Factorial scenario design

In addition to the variables indicated by models of evacuation decision-making, attributes of the event itself would impact decisions to stay or go. To measure these a factorial design was employed in which four aspects of each scenario would be randomly varied for the survey respondents so as to test the relative impact of each factor on the evacuation decision.

The first aspect involved whether prior warning of the event had been given: some respondents were presented with scenarios in which no prior notice of the impending event was provided whereas other respondents' scenarios were presaged with prior warnings from terrorists coupled with dirty bombs having exploded already in London and New York.

The hazard level of the scenario varied across three levels: minimum, moderate and maximum intensity. These levels were determined by the number and proximity of dirty bombs to the respondent. Each respondent received two of the three possible scenarios to avoid undue repetition during the survey. In each case, however, the first scenario was less dangerous than the second so that every respondent experienced escalation. So some respondents went from the minimum scenario to the moderate, others from moderate to maximum, and a third group from minimum to maximum.

The third aspect, the location of the respondent when a hazardous event occurred, varied across two choices: inside and at home, or inside a building but not at home, usually at the workplace. Respondents randomly selected to be at home would be at home for both scenarios during their interview. Respondents selected to be at work would be at work for both scenarios. Respondents who did not have a workplace but were selected to be at work were asked to think of a building away from their home where they might be during the day and this proxy location was used for both of their scenarios during the interview.

Finally, the source of messaging about the event varied across four choices: the local emergency manager, the local fire chief, the local chief administrative officer, or the governor/mayor. As with the location, respondents were randomly selected for one of these four messengers and that information source was used for both of their scenarios during the interview. Overall, this design creates 48 different versions of the dirty bomb scenario.

In addition, because evacuation behavior was important to understand and because the official instructions in all scenarios would be for people to shelter in place, respondents who chose to shelter in place in response to both scenarios presented to them were also asked about their evacuation behaviors if they were told by authorities that they had to evacuate.

Focus Groups

Three focus groups were held to test the draft survey instrument developed after the kick-off workshop in March. A group convened in Fairfax, VA on the evening of Monday, June 15, followed by a second during the afternoon on Thursday, June 18 in Washington, DC and a final focus group that evening in Landover Hills, MD. Participants were randomly recruited from residents of the immediate areas in which each group was held. About 12 people attended per group with each participant receiving a $40 cash incentive in a sealed envelope along with $5 to cover transportation and parking costs.

It is critical to our understanding of potential behaviors in an emergency that we understand the needs and concerns of NCR residents. The three focus groups were an effective aid in making sure the final survey would be relevant to those living in the area and that the questions could be answered in a way that is meaningful and unambiguous.

During the focus group, participants completed a paper version of the survey but in the interest of replicating the telephone interview experience as much as possible, the description of the fictitious scenarios to be considered was read aloud to them. Afterward, the CSR/CRMES team leading the focus group solicited comments about the clarity of the questions, their relevance to the participants' lives and whether participants felt able to answer the questions in a meaningful way.

The feedback from these focus groups was effective in the refinement of the survey. Most concerns with the questionnaire were related to the assumptions or parameters of the scenarios. Expanding answer choices and including some definitions (in particular, a better definition of the National Capital Region, NCR) would also help to clarify the intent of some questions. One very significant and unexpected finding was the reaction to the phrase "National Capital Region." We were advised to change it because the phrase was not in common use. Members settled on Washington Metro Area, and that phrase was used in the last group in Maryland, in the telephone pretest and in the final production survey.

Telephone Interviewing Procedures

CSR conducted the telephone interviews from its Computer-Assisted Telephone Interviewing (CATI) Laboratory at the University of Virginia. CATI is a system in which computers are employed to increase the efficiency, accuracy, and flexibility of telephone surveys conducted by trained interviewers. Questions appear on the computer screen in programmed sequence as the interviewer presses the keys on the keyboard to record the respondent's answers. Accurate, instantaneous data entry is assured by the system. The computer system stores the database of telephone numbers and is used to control the sampling process, dial each sampled number, schedule call-backs, and record the disposition of each attempted call. CSR's CATI laboratory also allows for audio-visual monitoring of calls by lab supervisors.

Pretest of the Survey Instrument

To test the clarity of the new questions and to verify the length of the instrument, CSR conducted a telephone pre-test. Pretests are intended to follow the protocol for the data collection as closely as possible to reveal any unforeseen problems. Findings may indicate that an adjustment to procedures is necessary to fulfill expectations. It is also possible at this time to recommend altering either the projected scope or the project budget.

The initial pretest, from July 26-28, 2009 confirmed that the instrument was too long at just over 37 minutes. The target length was 25 minutes. The projected number of completions per hour was also lower than the budget target rate of .83 per hour. Interviewers were able to complete only about 2/3 of an interview in an hour, a number that would be at least partly explained by the increased time it took to conduct an interview. Interviewers offered constructive suggestions, wording changes and posed thoughtful questions based on their experience. Researchers were able to streamline some areas of the script, offer helpful prompts and make some cuts to the script based on this debriefing and the findings from analysis of the collected data. Finally, "question rationing" was implemented so that certain questions would be randomly asked of fewer than all respondents in order to ask a larger number of questions overall and still obtain a sufficiently large sample of responses to each question without making the survey substantially longer for any individual respondent. This amended script was reviewed by the project's survey committee and returned to active data collection for a second pretest from August 14-16. Although the completions per hour remained lower than the budgeted numbers, the length of the interview was judged to be acceptable and this version of the survey, with minimal edits, was sent to production interviewing in late August.

Production Calling

Production calling for the survey was carried out from August 28 through December 17, 2009. The majority of the telephone calls for the study (90%) were made from the CSR CATI laboratory under the direct supervision of CSR staff. Ten percent of the calls were subcontracted out to NSØN, Inc. in Salt Lake City, Utah which has a CATI lab running similar software to that at CSR. NSØN followed protocols for the survey established by CSR. Numbers were dialed automatically by the WinCATI computer system except for cell phones, which were manually dialed to conform with Federal telecommunications regulations. Calling was done regularly on Sunday through Friday evenings, Sunday afternoons, and occasionally on Saturday and weekday afternoons. The interviewers received at least six hours of training prior to production interviewing. Many had prior interviewing experience on similar

studies. Each phone number was given a maximum of 12 call attempts before it was treated as a "no answer" or "busy" number. Residential phones answered by automatic answering machines were treated the same as "no answer" calls (although counted separately).

With respect to landline phones, CSR interviewers did not leave messages on the answering machines of potential respondents but simply returned the phone number to the sample pool for another calling attempt at a later time. However, cell phone users were left a message on the first attempt so that they would understand the purpose of the call and could arrange for a callback at a convenient time. Answering machine announcements that identified the phone number as a place of business, however, were recorded as such and not re-attempted on the landline sample. Because many cell phones are used for business and personal reasons, business messaged-phones were reattempted three times without leaving a message and then automatically finalized as a business. As with landlines, cell phones identified by the person answering as a business were not surveyed.

In order to reduce non-response bias, CSR conducted "conversion calling." Non-response bias in surveys results when qualified respondents do not complete a survey, usually because they refuse to cooperate. In conversion calling, CSR's most highly trained interviewers call back households in which a respondent had previously refused to take the survey. First, CSR kept track of the "tone" of initial refusals. "Hard" refusals, those in which people explicitly asked not to be called again or were noticeably agitated or upset about receiving a phone call, were not called back at all. "Soft" refusals, those for which it seemed that interviewers only caught someone at a bad time, were called back and contacted again after an interval of at least three days. Respondents were removed from calling after three "soft" refusals.

Productivity and Response Rates

The study statistics reported here combine results from calling efforts by CSR and by NSØN. A total of 30,824 phone numbers were attempted in the course of the survey, resulting in 2,533 completions and 182 partial cases, of which 124 were sufficiently complete to be used for analysis in the final dataset for a total of 2,657 interviews. The interviews took an average of 29.1 minutes to complete once a qualified respondent was identified through final respondent comments, with a median time of 27.7 minutes. Interviews completed in the Spanish language took 32.8 minutes on average (median of 32.6 minutes) to complete once a qualified respondent was identified, compared to 29.0 minutes in English (27.6 median time). Some of this difference may be accounted for because a higher proportion of Spanish language surveys were conducted by cell phone (58.8%) than was the case in English (16.1%). Cell phone surveys tend to be slightly shorter at the beginning because of the simpler selection process but longer at the end because of the need to obtain information for providing the incentive.

Landline surveys have a more complex selection process (discussed above) aimed at randomizing participant selection within a household. For the cell phone it was assumed that the person answering the phone was the primary user unless stated otherwise by the respondent. Nevertheless, overall on this survey, there was no significant difference between the cell phone and the landline surveys in the duration of the respondent selection stage of the survey, as is normally the case. However, there was a difference in duration at the end. The average length from greeting to goodbye on a landline interview was 32.8 minutes (same mean for both listed and RDD) whereas for the cell phone it was 34.5 minutes. If we look at the time from the point at which a qualified respondent was selected until hang-up, the cellular telephone survey took 31.5 minutes on average compared to 30.1 minutes for the landline.[5]

The final disposition of each of the attempted phone numbers is shown in two tables at the end of this Appendix. The disposition report is presented in a format that has been recommended as an industry standard by the American Association for Public Opinion Research. The AAPOR rate was calculated by a custom analysis of the complete call history of each attempted number, using a program written in SPSS by CSR technical staff. CSR completed a total of 2715 complete and partial interviews (including those completed in the conversion phase of calling and some cases that had to be excluded from the final

[5] The median times from hello to hang-up was 32.3 minutes for cellular, 31.1 for landline RDD, and 31.0 for listed.

dataset), for an overall response rate of 15.4%. Thirty-five interviews were conducted in Spanish. The response rate is a way of expressing the proportion of completed interviews against the number of eligible possible contacts. It does not include failures to interview at the number dialed because the number has not been assigned to an eligible participant or is not working. It does include eligible respondents who refuse cooperation and other numbers whose eligibility cannot be determined (busy or never answered numbers).

The true response rate depends on how one estimates the percentage of working residential phones that exist among the many numbers that never answered our many call attempts. An estimate of 15.6% for the landline only RR4 response rate (not shown in the table) is based on the most conservative assumption (equivalent to the CASRO rate) that the percentage of residential households among unreachable numbers is the same as the percentage among those we reached, i.e., 72.7%.[6] However, because CSR completed multiple attempts to nearly all of the no-answer numbers and based upon prior experimentation with listed and RDD samples in Virginia, we estimate that the residency rate is around 20% of no-answer numbers and that our true response rate (adjusted RR3) for landlines is closer to 16.8%. Within the landline sample the adjusted RR4 for RDD production was 16.0% and the unadjusted RR4 for listed production was 18.6%. For the cell phone portion of the sample, the estimated response rate is 12.2% and as with directory-listed sample the adjustment is not used.

Finally, the efficiency of the calling can be expressed in terms of number of completions per hour of calling (CPH). The overall interview production rate was 0.62 interviews per hour. It is noteworthy that cell phone and landline RDD (random B) productivity was very similar in this survey. When fielding the survey, we split the RDD sample into its listed back-matched and unable to be back-matched components. The back-matched sample had a productivity of .68 in contrast to the sample that could not be back-matched which proved even less productive than the cell phone sample at .35 (compare to figures in Table B-2 below). In addition to the normal loss of productivity associated with using a higher proportion of cell phones in the total sample along with the declining rates of RDD productivity nationwide, the lower response rate and productivity for this survey can in large measure be accounted for by the length of the questionnaire and the nature of the topic. These led to higher rates of refusal than is normally the case.

Table B-2: Completion Rates and Completions per Hour by Sample Type

	Completions + Partials	Interview Length*	Rate/hr	Response Rate (rr4)
PR1: RDD	1350		0.55	16.0
PR2 Listed	912		0.84	18.6
PR3: Cell	453	34.5	0.54	12.2
Landline only	2262	32.8	0.64	16.8
Overall	**2715**	**33.1**	**0.62**	**15.4**

*Interview length: average number of minutes from 'hello to hang-up'

[6] Calculated according to AAPOR suggested formula RR4 adjusted by comparison of listed and RDD telephone no-answer rates used to estimate the proportion of RDD no-answer numbers that are actually non-working or unassigned numbers, with *e1*=.167 and *e2*=.837. We estimated *e2* by dividing households determined to be eligible by the N of households overall (including cell phone numbers for the overall estimate where the actual rates of eligibility are unknown but likely to be much lower than our estimate). The estimated *e2* was also applied to housing units where eligibility could not be determined. We derived *e1* by taking the product of *e2* and the estimated residency rates calculated by the CASRO method. Partial interviews are not counted in the numerator of the RR3 formula but are counted in RR4 formula (see Table at the end of this appendix for details).

Geographic Representation

When RDD and cell phone sampling is employed, the surveying organization does not have any exact prior information on the location of the household. To protect respondent confidentiality and preserve a sense of privacy in the interview, CSR does not usually ask respondents to supply their address except for those randomly asked to participate in a follow-up study and cell phone participants who were asked what address to send the $10 gift card. However, this address data is kept separate from their questionnaire responses. Instead, interviewers asked respondents to identify which city or county they lived in. Respondents who indicated they were from specific locations in Virginia (Arlington County, Alexandria City, Fairfax City, Fairfax County, Falls Church City, Loudoun County, Manassas City, Manassas Park City, Prince William County), Maryland (Montgomery County and Prince George's County) as well as the District of Columbia were included in the survey. Later in the interview, respondent inclusion was confirmed by requesting their zip code information. Respondents who could not supply an area name, or who gave a name not on the list were excluded from the sample. When a respondent was uncertain of whether their town was located in one of the counties accepted for the survey the interviewer would ask the calling lab supervisor to look it up on the internet before proceeding. This procedure allowed CSR to code respondents into one of twelve geographic areas that are included in the analysis of the National Capital Region. This information was then used to group respondents into eight predefined geographic regions (see Table B-4) that would serve as a demographic variable in analyzing responses. This method helps inform our understanding of the representation of response from different areas of the NCR**.**

While this procedure has a satisfactory degree of accuracy for the purposes of comparison, it includes some inevitable inaccuracies due to errors on the part of respondents or interviewers. Apart from the District of Columbia and Prince George's County, MD, the geographic distribution was very close to population estimates for the region and thus geographic weighting was fairly minimal.

Table B-3: Respondents by Sample Type and Area - Unweighted

Area	Random Digit Dialing		Directory Listed		Cell phone		State Totals	
District of Columbia	154	50.0%	101	32.8%	53	17.2%	**308**	**11.6%**
Virginia	694	48.7%	495	34.7%	236	16.6%	**1425**	**53.7%**
Maryland	468	50.8%	301	32.7%	152	16.5%	**921**	**34.7%**
Sample Totals	**1316**	**49.6%**	**897**	**33.8%**	**441**	**16.6%**	**2654**	

Weighting

The inclusion of cell phones in the 2009 sample required a disproportionate sampling scheme; therefore the 2009 data are weighted on the type of sample from which each case was selected. Also, as is usual when using telephone survey methods, the sample composition did not exactly match the composition of the entire population of households in the NCR. Accordingly, statistical weighting of the survey results was designed to accomplish three objectives: (1) to correctly represent the geographic areas, (2) to properly represent certain demographic characteristics of the population, and (3) to properly represent different types of phone service in the National Capital Region population (cell phone-only cases, landline-only cases, and those with both kinds of telephone service), as well as the correct proportion of unlisted landline telephones.

Demographic weighting

This is the most common type of weighting that is done with a dataset. Random sampling error, systematic differences in rates of refusal between different groups, and differences among households in the amount of time that someone is home to answer the phone can result in a sample that somewhat over-

represents females, under-represents homeowners, and under-represents African Americans. To correct these imbalances, CSR weighted the sample data on demographic characteristics. Statistical weighting is larger for those respondents who are in underrepresented groups, and smaller for those who are in overrepresented groups, so that the aggregate result is what we would have obtained from a fully uniform, random sample of the whole population.

In order to calculate the correct weights, CSR drew upon information from the 2008 ACS (American Community Survey) in order to get the correct proportions of the adult population in the National Capital Region. Because the ACS provides information on homeownership by race, CSR combined the two demographic variables to create one weight based on homeownership and race together. The proportion of male/females from the adult population of the National Capital Region was the basis for the weight for gender and the proportion of homeowners/renters, cross-tabulated by race, served to determine the weight based on homeownership and race. Note that we did not weight the data with respect to age.

Geographic weighting

In addition to demographics, statistical weighting was applied to correct geographic representation within the National Capital Region. Table B-4 shows the percentage of the area in the population taken from the 2008 American Community Survey compared to its percentage in the sample. The geographic weight is the amount each case would need to be multiplied by in order to have the sample percentage for each area be equal to its actual population proportion. In practice, the geographic weight is combined with the other weights through an iterative process called "raking."

Table B-4 Distribution of National Capital Region, and Geographic Weight Values

NCR areas	Population of Households, ACS 2008		2009 Unweighted Sample		Geographic Weight
	(count)	(%)	(count)	(%)	
District of Columbia	284,164	18.88%	308	11.61%	1.627
Arlington County, VA	100,876	6.70%	138	5.20%	1.289
Alexandria County, VA	70,058	4.66%	123	4.63%	0.005
Fairfax & Falls Church, VA	400,769	26.63%	731	27.54%	0.967
Loudoun County, VA	101,396	6.74%	183	6.90%	1.977
Manassas & Prince William, VA	149,796	9.95%	250	9.42%	1.057
Montgomery County, MD	362,269	24.07%	589	22.19%	1.085
Prince George's County, MD	319,620	21.24%	332	12.51%	1.698
Total	1,504,784	100%	2,654	100%	

Cell phone weighting

Current research on cell phone interviewing is still in its infancy, and there are no standard, accepted methods for weighting the results of a 'dual frame' sample that combines completed interviews from landline samples with completed interviews from cell phone samples. Prof. Guterbock has been working on the development of appropriate methods, and our approach to the current study applies his latest research to the available local data. Here we treat RDD and listed samples as one "landline" sample, thus treating our triple-frame design as a dual-frame sample (cell phone and landline sampling frames).

The heart of the weighting problem is simple: there is no available external source that will tell us the percentage of the area population that has cell phone-only service, landline only, or both. Authoritative data are collected at the national level by the Centers for Disease Control in the National Health Interview

Survey, a very large, continuous, in-person data collection focused on health issues.[7] That survey determines the phone-service status of each household in a representative national sample, and results from as recently as the first half of 2009 are currently available. However, these data are available only at the national or broad regional level. It is doubtful that these broad averages across regions are directly applicable to the National Capital Region.

The estimation problem is made somewhat more difficult by the fact that rates of survey response are not even across different phone-use segments. That is, cell phone-only adults are much more likely to answer their cell phones than are those who have both kinds of phones. This is understood to reflect differences in telephone behavior between cell phone-onlies and dual-phone users. Cell phone-onlies are presumably more likely to have their phones with them, to have their phones turned on, and to accept calls from unknown numbers than are those who continue to rely primarily on landline phones. For these reasons, the percentage of cell phone-only cases encountered in actual cell phone surveys is much higher than their actual share among all cell phone users. It is probably also the case that landline-only households are somewhat overrepresented within landline samples, as compared to those who have both kinds of phone. The latter group is referred to below as the *overlap sample*, because the households having both landline and cell phones lie at the intersection of the cell phone frame and the landline frame.

In order to estimate the degree of under-representation of the overlap sample segment in the cell phone sample and in the landline sample, we compared results from the 2007 California Health Interview Survey (a telephone survey combining RDD sample with cell phone-only households) with the results from NHIS for the Western Region of the United States (second-half 2007 results).[8] Using algebraic formulas developed by Prof. Guterbock, we were able to determine the values for two *response rate ratios:* r1, the ratio of the response rate to cell phone calling in the overlap sample compared to the response rate of cell phone-onlies, and r2, the ratio of the response rate to landline calling in the overlap sample to the response rate of landline-onlies. The NHIS for the Western region reports that the phone-service proportions in the Western region were: 13.2% cell phone-only, 67.9% dual-phone (overlap), and 18.9% landline only. If response rates were equal (r1 = r2 = 1.0), and if California's phone usage is the same as that of the Western region, then the CHIS 2007 would have found 16.3% of the cell phone completions to be cell phone-onlies. Instead, CHIS 2007 reports 34.6% percent cell phone-onlies. CHIS should have found 21.7% landline-onlies in the landline sample, but actually had 32.7% landline-onlies in its landline RDD sample. Applying Guterbock's formulas to these data results in an estimate of r1 = .368 and r2 = .598.

Because final results of the survey were not available at the time when decisions had to be made about the sample weights, the basic weights were determined using near-final survey data as shown in Table B-5. The "estimated true" values are derived by application of the values for r1 and r2 estimated above to the data from our 2009 survey completions in the NCR.

Table B-5: Initial estimates of the phone-service segments in the NCR 2009

	Cell phone sample		Landline sample		Combined samples		Est. true	Weight	Weighted N	
Cell Only	132	31.4%	6	0.3%	138	5.8%	13.65%	2.355	325	13.6%
Overlap (Both)	288	68.4%	1743	88.9%	2031	85.3%	80.35%	0.942	1914	80.3%
LL Only	1	0.2%	212	10.8%	213	8.9%	8.9%	0.672	143	6.0%
	421		**1961**		**2382**		**100%**		**2382**	

[7]Steven J. Blumberg and J.V. Luke. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2007." National Center for Health Statistics, May 13, 2008. Available at: http://www.cdc.gov/nchs/nhis.htm.

[8]Thanks to Michael Brick of Westat for sharing some of the preliminary results from CHIS 2007 for this purpose.

Once these estimates were made, a further decision needed to be made about weighting the overlap sample. By design, we did not complete a very large number of cell phone cases because of their greater expense. In theory, if all phones in the area had been called with equal likelihood, we would have reached one half of the overlap sample through their cell phone and one half through their landline. This would call for weighting the portion of the overlap sample reached through cell phone up by a very large weight to bring their share of the overlap to 50%, which could potentially have distorted the results and also increased the 'design effect' in the study, reducing the precision of the estimates. We decided to apply a weight of 2.0 to the cell phone cases in our overlap sample, allowing the weight on the landline cases in the overlap sample to take a value that would result in an overall overlap percentage in the weighted sample of 80.35%. Table B-6 shows these weights as applied to the completions in the near-final sample. When data were subjected to final cleaning and the last few interviews were completed, the final number of usable cases increased slightly, but the weights shown below were applied to all cases in each phone-usage segment.

Table B-6: Final estimates of the phone-service segments in the NCR 2009

	Cell phone sample		Landline sample		Combined samples		Est. true	Weight	Weighted N	
Cell only	132	31.4%	6	0.3%	138	5.8%	13.65%	2.3554	325	13.6%
Overlap: Cell	288	68.4%	0		288	12.1%	24.18%	2.0000	576	24.2%
Overlap: LL	0		1743	88.9%	1743	73.2%	56.16%	0.7675	1338	56.2%
LL only	1	0.2%	212	10.8%	213	8.9%	6.01%	0.6720	143	6.0%
	421		**1961**		**2382**	**100%**	**100%**		**2382**	**100%**

Listed status weighting

We also weighted the results to accurately represent unlisted landline cases. These are somewhat underrepresented because the directory-listed sample has only a small percentage of unlisted households. To correct for this, we weighted all unlisted landline households reached on either the RDD or EWP (listed) samples so that, in total, they represent 17.2 percent of the landline completions.

Combining the weights

The final step in the weighting process was "raking," a statistical procedure used to produce combined weights for the five weighting factors: gender, race with home-ownership, geography, phone service type, and listed versus unlisted telephone status. The percentages for geographical areas in Table B-4 were used along with the weights for phone usage from Table B-6 in an iterative process that produced a final weight for each of the 512 design cells (2 genders $\times$ 4 homeownership-race categories $\times$ 4 phone-service segments $\times$ 8 areas $\times$ 2 listed statuses [unlisted landline versus all others]) that would best fit with the given marginal population distribution for each weighting factor. This procedure necessarily treats the distribution of phone-service segments as being equal across the geographic areas.

A more complete description of the cell phone estimation procedures used here, along with algebraic formulas needed to calculate and apply the response rate ratios, is available upon request.[9]

[9]Thomas M. Guterbock. "Estimating Phone Service and Usage Percentages: How to Weight the Data from a Local, Dual-Frame Sample Survey of Cellphone and Landline Telephone Users in the United

Sampling Error and Statistical Testing

Our final sample includes 2,657 respondents. If these cases had been drawn by simple random sample, the survey would have a margin of error of plus or minus 1.90 percent. However, in addition to sampling error there is a design effect that impacts the total margin of error which we calculate by introducing the weights derived by the "raking" process described above into the Complex Sampling module of SPSS statistical software. This tool allows calculation of a "design effect" for each question in the survey. The design effect shows how the variance of sample estimates is increased by the effect of post-stratification weighting. We base our estimate of the overall margin of error on a key survey question, in this case the question a respondent had to answer to be included in the final dataset (PETS). For that question, 2,654 respondents provided an answer and so the design effect is 1.449, meaning that the margin of error in our sample is equivalent (because of the weighting) to the margin of error we would have obtained from a simple random sample of 1,832 (2,654/1.449), the effective sample size for that question. The margin of error is increased by the square root of the design effect, a factor in this case of 1.20. The final margin of error is 2.29%. This means that in 95 out of 100 samples of this size drawn from the National Capital Region, the results obtained in the sample would fall in a range of ± 2.29 percentage points of what would have been obtained had every household in the area with a working landline or cellular telephone been interviewed. Larger sampling errors are present when analyzing subgroups of the sample or questions that were not asked of all respondents; smaller sampling errors are present when a lopsided majority gives the same answer (e.g., 80 percent of the sample are satisfied with a given service).

Statistical significance tests were used to verify the existence of differences among various subgroups. We used independent-sample t-tests for differences in means and the Pearson Chi-Square test of independence for differences in proportions. In chi-square tests of some items, the several response categories were collapsed into two. We report in these pages differences that yield a "p-value" of .05 or less. A level of .05 indicates that there is only a 5 percent chance that the difference we find is due to sampling error, rather than reflecting a real relationship within the study population. The statistics for evaluating statistical significance were calculated using the SPSS Complex Sampling module and hence take into account the "design effect." When the design effect is taken into account, tests of significance become more conservative, requiring a somewhat larger difference between groups to achieve significance at the 95% confidence level. The statistics for evaluating statistical significance do not measure error from sources other than random sampling error. Such error can occur in any poll or survey.

The following disposition reports detail the final resolution to every attempted phone number in the sample. The report form is arranged for calculation of AAPOR standard rates.

States." Paper presented at the Annual Meetings of the American Association for Public Opinion Research, Hollywood, Florida, May 14, 2009.

Table B-7: Sample Disposition Report

NCR 2009 – Disposition Listing for All Samples

[dispositions arranged for calculation of AAPOR standard rates]

Disposition Code	Disposition Description	All Samples Total	Random Digit Dialing	Directory Listed	Cellular (Wireless)
1100	Complete	2533	1259	856	418
1200	Partial	182	91	56	35
2110	Eligible: Refusal	3293	1687	975	631
2120	Eligible: Break-off	193	103	65	25
2210	Eligible: Resp Never Avail	1231	542	306	383
2221	Eligible: Ans Mach, No Mess	5330	3123	1885	322
2222	Eligible: Ans Mach, Message	2448	49	5	2394
2310	Eligible: Dead	0			
2320	Eligible: Phys/Mentally Unable	81	46	33	2
2330	Eligible: Language Unable	621	259	92	270
2340	Eligible: Misc. Unable	34	8	8	18
3120	Busy	332	189	29	114
3130	No Answer	2107	1752	43	312
3140	Ans Mach (Don't Know if HU)	959	245	98	616
3150	Technical Phone Problems	553	175	35	343
3210	HU, Unknown Eligible: No Scrnr	2481	1264	608	609
3220	HU, Unknown Eligible: Other	3	1		2
4100	Out of Sample	624	89	29	506
4200	Fax/Data Line	1168	1030	123	15
4310	Non-working Number	2532	656	268	1608
4320	Disconnected Number	1307	820	386	101
4410	Number Changed	243	171	19	53
4420	Cell Phone	N/A			
4430	Call Forwarding	2			2
4510	Business/Govt/Other Org	1743	1384	144	215
4520	Institution	0			
4530	Group Quarter	1	1		
4700	No Eligible Respondent	823	168	64	591
4800	Quota Filled	0			
	Total	**30824**	**15112**	**6127**	**9585**

Table B-8: Sample Disposition Report

NCR 2009 – AAPOR Standard Rates Calculation

[Dispositions summary for all Telephone Samples]

AAPOR Standard Rates and Dispositions Summary	OVERALL Ave	Random Digit Dialing	Directory Listed	LANDLINE Ave	Cellular (Wireless)
*Estimated Residency 1**	0.381	0.187	0.806	0.365	0.415
Estimated Residency 2	0.837	0.935	0.960	0.944	0.576
Response Rate 1	0.125	0.121	0.173	0.137	0.087
Response Rate 2	0.134	0.130	0.184	0.147	0.095
Response Rate 3 *	0.143	0.150	0.175	0.157	0.113
Response Rate 4 *	0.154	0.160	0.186	0.168	0.122
Response Rate 5	0.184	0.185	0.206	0.193	0.150
Response Rate 6	0.197	0.199	0.220	0.207	0.163
Cooperation Rate 1	0.365	0.364	0.411	0.382	0.299
Cooperation Rate 2	0.391	0.391	0.437	0.408	0.324
Cooperation Rate 3	0.408	0.401	0.439	0.415	0.377
Cooperation Rate 4	0.438	0.430	0.467	0.444	0.408
Refusal Rate 1	0.156	0.166	0.204	0.178	0.101
Refusal Rate 2 *	0.191	0.213	0.212	0.213	0.177
Refusal Rate 3	0.219	0.250	0.243	0.247	0.146
Contact Rate 1	0.310	0.320	0.409	0.349	0.215
Contact Rate 2 *	0.355	0.393	0.415	0.399	0.279
Contact Rate 3	0.435	0.482	0.487	0.484	0.258
Complete Interview	2532	1258	856	2114	418
Partial Interview	183	92	56	148	35
Refusal and Break-off	3486	1790	1040	2830	656
Non-contact	9009	3714	2196	5910	3099
Other eligible but unable	736	313	133	446	290
Unknown if household	3951	2361	205	2566	1385
Unknown if other	2484	1265	608	1873	611
Ineligible Numbers	8443	4319	1033	5352	3091
Total Dialed Attempts	160765	94480	34779	129259	31506
TOTAL	**30824**	**15112**	**6127**	**21239**	**9585**
% of Landline		71.2%	28.8%	100.0%	
% of Overall	100.0%	49.0%	19.9%	68.9%	31.1%

**Contains CSR adjustment rate for Virginia residency for RDD portion of the sample. Estimated residency rate for cellular (wireless) sample derives from Landline assumptions. No adjustment estimates available for cellular samples at this time.*