



MINNESOTA ADULT TOBACCO SURVEY

Tobacco Use in Minnesota: 1999 to 2010

Minnesota Adult Tobacco Survey
MATS 2010 Methodology Report

February 2011





Minnesota Adult Tobacco Survey MATs 2010 Methodology Report

ClearWay MinnesotaSM

Minnesota Department of Health

Prepared by:

Westat
Rockville, MD

February 2011



Table of Contents

| <u>Chapter</u> | | <u>Page</u> |
|----------------|---|-------------|
| 1 | Introduction..... | 1-1 |
| | 1.2 Orientation to the Methods Report..... | 1-5 |
| | 1.3 Data Collection and Data Processing Timeline..... | 1-5 |
| 2 | Sampling..... | 2-1 |
| | 2.1 Random-Digit Dialing (RDD) Landline Sample..... | 2-4 |
| | 2.1.1 MATS 2010 Refinement of Basic RDD Landline Sampling Methodology: Refusal Conversion Subsample | 2-5 |
| | 2.1.2 Landline Sample Telephone Number Duplicates of MATS Cohort Sample Telephone Numbers | 2-5 |
| | 2.1.3 Within-Household Random Selection of Adult..... | 2-6 |
| | 2.1.4 RDD Landline Sample: Projected vs. Actual Counts..... | 2-6 |
| | 2.2 RDD Cell Phone Sample..... | 2-7 |
| | 2.2.1 Screening for Cell-Only and Cell- Mostly Households..... | 2-7 |
| | 2.2.2 Within-“Household” Random Selection of Adult..... | 2-8 |
| | 2.2.3 RDD Cell Phone Sample: Projected vs. Actual Counts..... | 2-8 |

Table of Contents (continued)

| <u>Chapter</u> | | <u>Page</u> |
|----------------|---|-------------|
| 3 | Data Collection Methodology..... | 3-1 |
| | 3.1 Questionnaire Development..... | 3-1 |
| | 3.1.1 MATS 2010 Questionnaire | 3-1 |
| | 3.1.2 MATS 2010 RDD Household Screeners..... | 3-7 |
| | 3.1.3 CATI Questionnaire Programming and Testing..... | 3-7 |
| | 3.2 Pilot Test | 3-8 |
| | 3.2.1 Background..... | 3-8 |
| | 3.2.2 Pilot Test Operations | 3-8 |
| | 3.2.3 Pilot Test Interview Monitoring and Interviewer Debriefing..... | 3-9 |
| | 3.2.4 Pilot Test Findings | 3-11 |
| | 3.3 Interviewer Recruitment and Training | 3-11 |
| | 3.4 Communications with Sample Members..... | 3-14 |
| | 3.4.1 Letters | 3-14 |
| | 3.4.2 MATS 2010 Informational Website | 3-15 |
| | 3.4.3 Toll-free Numbers and Contacts Provided by Westat and the Minnesota Department of Health..... | 3-16 |
| | 3.5 Data Collection Confidentiality Procedures and Protection of Human Subjects..... | 3-16 |
| | 3.6 Data Collection Operations..... | 3-18 |
| | 3.6.1 Calling Procedures..... | 3-18 |
| | 3.6.2 Home-based Interviewers..... | 3-21 |

Table of Contents (continued)

| <u>Chapter</u> | <u>Page</u> |
|----------------|--|
| 3.6.3 | Cell Phone Screener Incentive Payments 3-21 |
| 3.6.4 | Refusal Conversion..... 3-22 |
| 3.7 | Data Quality Assurance..... 3-23 |
| 3.7.1 | Interviewer Monitoring..... 3-23 |
| 3.7.2 | Data Cleaning and Editing During Data Collection 3-24 |
| 3.7.3 | Data Cleaning and Editing Following Data Collection 3-25 |
| 3.7.4 | Sample Performance Monitoring During Data Collection 3-26 |
| 4 | Data Collection Results..... 4-1 |
| 4.1 | Completed interviews..... 4-1 |
| 4.2 | Telephone Interviewing Results..... 4-1 |
| 4.2.1 | Landline Sample Call Dispositions 4-1 |
| 4.2.2 | Cell Phone Sample Call Dispositions 4-2 |
| 4.2.3 | Combined Sample Call Dispositions..... 4-2 |
| 4.3 | MATs 2010 Response Rates 4-10 |
| 4.3.1 | Methodology..... 4-10 |
| 4.3.2 | Screener Response Rates 4-13 |
| 4.3.3 | Extended Response Rates and Overall Response Rates 4-16 |
| 4.4 | Selected Operational Statistics..... 4-18 |
| 4.4.1 | Principal Sources of Non-response 4-18 |

Table of Contents (continued)

| <u>Chapter</u> | <u>Page</u> |
|---|-------------|
| 4.4.2 Refusal Conversion Results | 4-22 |
| 4.5 Interview Timing | 4-24 |
| 5 Weighting | 5-1 |
| 5.1 Household-Level Weights | 5-2 |
| 5.1.1 Household Base Weights | 5-2 |
| 5.1.2 Adjustment for Screener Non- response | 5-3 |
| 5.1.3 Adjustment for Multiple Telephone Numbers in Household | 5-3 |
| 5.2 Person-Level Weights | 5-4 |
| 5.2.1 Adjustment for Within-Household Selection | 5-5 |
| 5.2.2 Adjustment for Extended Interview Non-response | 5-5 |
| 5.3 Composite Weights | 5-6 |
| 5.4 Calibration to External Population Totals | 5-7 |
| 6 Comparability with Previous MATS Surveys and Limitations of the Data | 6-1 |
| 6.1 Comparability Issues | 6-1 |
| 6.1.1 Sampling | 6-2 |
| 6.1.2 Weighting | 6-3 |

Table of Contents (continued)

| <u>Chapter</u> | <u>Page</u> |
|---|-------------|
| 6.1.3 Data Collection | 6-4 |
| 6.1.4 Questionnaire Design and Specification | 6-5 |
| 6.2 Potential Limitations of the Data | 6-8 |
| 6.2.1 Sampling Error | 6-8 |
| 6.2.2 Potential Sources of Nonsampling Error | 6-10 |

Appendixes

- A. MATS 2010 Questionnaire
- B. MATS 2010 Questionnaire Skip Pattern Table
- C. MATS 2010 Landline Screener Questionnaire
- D. MATS 2010 Cell Phone Screener Questionnaire
- E. MATS 2010 Letters
- F. MATS 2010 Web Page Content
- G. Significant Modifications of MATS 2007 Questionnaire Items for MATS
2010

Table of Contents (continued)

| <u>Tables</u> | <u>Page</u> |
|---------------|--|
| 1-1 | Timeline of MATS 2010 development, data collection and data preparation..... 1-6 |
| 2-1 | MATS 2010 RDD landline sample – design projections and actual 2-7 |
| 2-2 | MATS 2010 RDD cell phone sample – design projections and actual 2-9 |
| 4-1 | MATS 2010 completed interviews, by sample type 4-1 |
| 4-2 | MATS 2010 landline telephone number sample dispositions 4-2 |
| 4-3 | MATS 2010 landline extended interview sample dispositions 4-4 |
| 4-4 | MATS 2010 cell phone telephone number sample dispositions 4-5 |
| 4-5 | MATS 2010 cell phone extended interview sample dispositions 4-6 |
| 4-6 | MATS 2010 combined telephone number sample dispositions 4-7 |
| 4-7 | MATS 2010 combined extended interview sample dispositions 4-9 |
| 4-8 | AAPOR response rate formula categories 4-11 |
| 4-9 | Unweighted counts of phone numbers by screener response rate formula category 4-15 |

Table of Contents (continued)

| <u>Tables</u> | <u>Page</u> |
|---------------|---|
| 4-10 | Estimated residential rates (e) among different types of non-contact cases 4-16 |
| 4-11 | Unweighted counts of sampled persons by extended response rate formula category 4-17 |
| 4-12 | Weighted response rates for landline and cell phone samples..... 4-18 |
| 4-13 | Primary sources of non-response in landline sample, collapsed across screening and extended interview stages 4-19 |
| 4-14 | Primary sources of non-response in landline extended interview sample 4-20 |
| 4-15 | Primary sources of non-response in cell phone sample, collapsed across screening and extended interview stages 4-20 |
| 4-16 | Primary sources of non-response in the cell phone extended interview sample 4-21 |
| 4-17 | MATS 2010 screener refusal conversion 4-22 |
| 4-18 | MATS 2010 extended interview refusal conversion..... 4-23 |
| 4-19 | Screener lengths (minutes)..... 4-24 |
| 4-20 | Extended interview lengths (minutes) 4-24 |
| 5-1 | Description of variables used to define MATS 2010 RDD raking cells..... 5-9 |



1. Introduction

The Minnesota Adult Tobacco Survey (MATS) is a comprehensive surveillance initiative designed to monitor progress toward meeting the goal of reducing tobacco use among Minnesotans. The major objective of MATS is to collect in-depth public health surveillance data on the adult population of Minnesota, focusing on tobacco use and cigarettes in particular. MATS is the most comprehensive source of information about smoking prevalence, behaviors, attitudes and beliefs in the adult Minnesota population; further, MATS provides valid scientific data that track the impact of comprehensive tobacco control efforts in Minnesota. MATS 2010 is the fourth survey in this ongoing surveillance initiative.

Historically, the MATS surveillance initiative and the first three surveys—1999, 2003 and 2007—were directed by three partner organizations who lead comprehensive tobacco control efforts in the state of Minnesota: ClearWay MinnesotaSM, Blue Cross and Blue Shield of Minnesota (Blue Cross) and the Minnesota Department of Health (MDH). For MATS 2010, the partners were ClearWay Minnesota and MDH, who formed the MATS 2010 Advisory Panel that selected Westat as the survey vendor for MATS 2010, made key decisions about survey design and provided oversight for the instrumentation, data collection, analyses and reporting of findings.

ClearWay MinnesotaSM is a nonprofit organization that strives to enhance life for all Minnesotans by reducing tobacco use and exposure to secondhand smoke through research, action and collaboration. ClearWay Minnesota serves Minnesota through its grant-making program, QUITPLAN® Services to help people quit smoking and statewide outreach activities. QUITPLAN Services helped more than 18,000 adult Minnesotans successfully quit tobacco use. ClearWay Minnesota designs and develops innovative statewide multimedia campaigns to inform the public of QUITPLAN Services and raise the awareness of the harm of secondhand smoke exposure. ClearWay Minnesota also works to build capacity and engage priority populations in reducing the harm that tobacco causes their communities. ClearWay Minnesota was created in 1998 when the state received \$6.1 billion from its

settlement with the tobacco industry and 3 percent, or \$202 million, was dedicated by the Ramsey County District Court to establish the independent nonprofit organization.

The Minnesota Department of Health launched the first state-funded tobacco control program in the nation in 1985 with a portion of the proceeds from a cigarette tax. Since then, MDH has undertaken a number of tobacco control initiatives including participating as one of 17 American Stop-Smoking Intervention Study demonstration states, a national-level comprehensive tobacco control program sponsored by the National Cancer Institute. Funds from an endowment from the state's 1998 settlement with the tobacco industry were available to the department from 2000 through 2003 and were used to launch a comprehensive youth prevention initiative during that period. Currently, MDH works to reduce smoking through grants to reduce youth exposure to pro-tobacco influences, to create tobacco-free environments and to reduce tobacco related health disparities.

Through a competitive process, the Advisory Panel selected **Westat**, a leading health and social science research organization based in Rockville, MD, as the survey vendor for MATS 2010. Westat was also the survey vendor for MATS 2007. Westat contributed technical expertise in sampling, weighting, and survey and analytical methods. With detailed direction from ClearWay Minnesota and MDH, Westat designed MATS 2010 and collected, analyzed and reported on MATS 2010 data. As a full-service vendor, Westat made recommendations to the Advisory Panel for adjustments to the previous MATS methodology based on the most up-to-date developments in survey research and study design.

The main components of MATS 2010 were as follows:

- **Sampling:** developing and drawing statistical survey samples that are representative of the Minnesota adult population. The sample design called for a random-digit dialing (RDD) sample of the adult Minnesota population, using samples drawn from landline and cell phone sampling frames.
- **Questionnaire Development and Data Collection:** developing and administering a survey questionnaire that would obtain from the survey samples all the data items needed to support the larger health and tobacco-

related missions of the sponsoring organizations. The questionnaire covered domains such as general physical and mental health, alcohol use, cigarette smoking and other tobacco use (including new tobacco products created by the tobacco industry), smoking cessation, experience with health care provider smoking interventions, attitudes towards smoking, situational exposure to secondhand smoke in various settings, the effects of public and private policies and rules on smoking behaviors and perceptions, and demographic information. The questionnaire was administered using a computer-assisted telephone interviewing (CATI) system.

- **Survey Operations:** developing various operational procedures to support the administration of the questionnaire. These included telephone contacting rules and procedures that met or exceeded the standard requirements for the U.S. Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System surveys (BRFSS), along with supporting measures such as contact letters and an informational website.
- **Sample Weighting:** designing and creating sets of survey sample weights that can be used in analysis and reporting to make the sampled respondents' data statistically representative of the entire population they were designed to represent. Weights were based on the probability of selection into the sample as adjusted to selected available characteristics and counts of the adult Minnesota population. Survey weights were developed for the combination of landline and cell phone samples that is to be used for analysis and reporting on the overall Minnesota population.
- **Tabulation and Analysis:** designing the various in-depth analyses of the survey data needed to support the sponsors' current and future programmatic, advocacy, public health, and tobacco-related health care delivery activities, as well as developing the detailed analytical tools and specifications for tabulating and analyzing the data.
- **Reporting:** preparing an in-depth report profiling the adult Minnesota population in regard to the use, knowledge, attitudes, experiences, and plans surrounding tobacco/cigarette use, tobacco cessation, exposure to secondhand smoke, and other tobacco and health-related areas.

MATS 2010 data are to be used both to report the prevalence of tobacco use, exposure to secondhand smoke and related factors as of 2010 as well as to measure



changes in these variables over time since MATS 2007, as well as monitor general trends from 1999 to 2010. In this context, a critical objective for MATS 2010 was to maintain continuity with the previous MATS surveys. This continuity served primarily to support reliable tracking of population trends over time and to support inferential statements that observed significant changes over time reflect actual changes in the population and are not artifacts of differences in the survey design. Some changes in design are inevitable or unavoidable in large-scale surveys repeated over long time periods. Maintaining continuity in MATS 2010 was, therefore, a balancing act between adhering to past MATS instruments and methods and making desired or necessary improvements.

Comparability was also an objective for the design of the survey weights. Investigators from the University of Minnesota weighted MATS 1999 and MATS 2003 in accordance with generally accepted practices, such as CDC's BRFSS and other statewide tobacco surveys. These surveys generally post-stratify only on age and gender. In recent years, however, concern has grown among the research community regarding the representativeness of telephone survey samples, particularly in terms of educational attainment. Telephone surveys increasingly appear to be more likely to reach individuals with higher education attainment (e.g., those with college degrees) than those with less education (e.g., those with a high school diploma or those who did not complete high school). While this phenomenon is not altogether new to survey research, the magnitude of the problem seems to have increased rapidly in the recent past. Because smoking and education status are inversely associated, the MATS Advisory Panel and Westat chose to include education as an adjustment factor for MATS 2007 and MATS 2010. To facilitate the most accurate comparisons between years of MATS administrations, the data from MATS 1999 and MATS 2003 were reweighted in 2007 to include educational attainment and race. Therefore, estimates from MATS 1999 and 2003 presented in MATS 2010 reports may vary slightly from estimates reported in publications prior to 2007.

1.2 Orientation to the Methods Report

This report constitutes the public documentation of general technical aspects of the MATS 2010 survey. It covers the sampling (Chapter 2), questionnaire development and data collection methodology (Chapter 3), the operational results of the data collection (Chapter 4), and the sample weighting (Chapter 5). Appendices include the MATS 2010 questionnaire, letters used in contacting the survey sample members, and the contents of a website that was created to provide information about the survey to potential respondents. The MATS 2010 analysis and reporting components are outside of the scope of this survey methods and appear as a separate, in-depth analytical report, titled *Tobacco Use in Minnesota: 2010 Update* (Minneapolis, MN: ClearWay MinnesotaSM and Minnesota Department of Health; February 2011). This report can be found at www.mnadulttobaccosurvey.org.

The day-to-day development and conduct of MATS 2010 required many detailed, internal design, specification, and reporting documents and tools that are beyond the scope of this report. The Minnesota Adult Tobacco Survey 2010 Comparability Report¹ (Westat, March 2010) itemized the areas where MATS 2010 conformed to or diverged from MATS 2007, along with descriptions of the rationale for any differences and their potential impact on comparability as they might affect the findings of trends over time from MATS 2007 to MATS 2010.

1.3 Data Collection and Data Processing Timeline

Westat's work on the development of the MATS 2010 survey began in September 2009. Data collection concluded in June 2010 and final quality assurance checks of the data and post-coding of open-ended responses were completed by August 2010. Table 1-1 shows the timeline for the major activities of MATS 2010 from survey design through creation of the weighted data sets. Analysis and reporting activities are not included in this timeline.

¹ Readers interested in more information about this report may contact Ann St. Claire, ClearWay Minnesota, at (952) 767-1416 or astclaire@clearwaymn.org.

**Table 1-1. Timeline of MATS 2010 development, data collection and data preparation**

| Date | Task |
|-----------------------|--|
| Summer 2009 - 9/29/09 | Select Westat as survey vendor, begin survey development (MATS Advisory Panel) |
| 9/29/09 | Hold kick-off meeting with Westat and Advisory Panel |
| 9/29/09-1/31/10 | Design, program and internally test MATS 2010 CATI questionnaire |
| 9/29/09-1/31/10 | Develop data collection protocols and supporting materials |
| 9/29/09-1/31/10 | Design samples, create sampling frames, draw and process sample for data collection (RDD & Blue Cross) |
| 2/1/10-2/5/10 | Conduct RDD pilot test and revise questionnaire |
| 2/12/10-2/17/10 | Telephone interviewer training |
| 2/19/10-5/30/10 | Telephone data collection |
| 5/31/10-8/26/10 | Final data quality assurance, post-coding, and weights for review & acceptance |

2. Sampling

The MATS 2010 sample design utilized a random-digit dialing (RDD) sampling method from two sample frames with the goal of conducting 7,000 interviews. The sample frames were 1) a statewide frame of all possible landline telephone numbers and 2) a statewide frame of all possible cell phone numbers. The targets were 5,950 interviews from the landline frame and 1,050 interviews from the cell phone frame. In terms of the reliability of the Minnesota adult smoking prevalence rate to be estimated from the combined dataset, MATS 2010 was designed to detect a 2.5 percentage point difference between two point estimates, one for 2007 and one for 2010, with 80 percent probability (power) at the 95 percent confidence level, based on a one-tailed significance test.² A single CATI questionnaire was used for both the landline and cell phone samples.

MATS 2010 Dual Frame Design

Landline Households

The MATS 2010 sample design included an RDD landline sample, as did the MATS 2007 sample design. However, unlike MATS 2007, there was no oversampling of phone numbers from telephone exchanges with higher proportions of African Americans, and no oversampling of young adults from households with both young adults and older adults.

Cell Phone-only (“Cell-only”) Households

It is well established that the cell phone-only (“cell-only”) households are undercovered in landline-based RDD surveys. The numbers of households and persons in the United States who have cell phones have greatly increased in the last few years with estimates of cell-only adults reaching 14.5 percent nationally for the

² The MATS 2007 sample was designed in part to be large enough to detect a 2.5 percentage point difference based on two-tailed significance tests. Detecting the same difference using a two-tailed test requires a larger sample size than using a one-tailed test. See the *Tobacco Use in Minnesota: 2010 Update* for more information about the MATS significance tests.

last 6 months of 2007 (Bloomberg and Luke, 2008). The nature of telephone surveys has evolved to deal with these changes.

Studies of cell phone users have also found that the characteristics of the persons in cell-only households are different from those in households with landlines. For example, cell-only adults are much less likely to have health care insurance than adults in households with landlines. Demographics such as age and gender are cited as associated with cell-only households. Since people with certain characteristics are undercovered in varying degrees by landline RDD surveys, there is increasing concern about the quality of estimates derived from them. For example, given the high prevalence of young adults in cell-only households, some observed decreases in the prevalence of certain health-risk behaviors may be artifacts of young adult undercoverage.³ Such findings suggested that bias due to the failure to cover these households is possible.

In response to the concern that, by 2010, 20 percent or more of households may be cell-only, the MATS 2010 sample design included a cell frame sample component designed to improve the coverage of the Minnesota population, in particular those living in households without a landline. The coverage concerns were even greater for younger adults: nearly half of adults aged 25-29 years (45.8 percent), more than one-third of adults aged 18-24 (37.6 percent) and approximately one-third of adults aged 30-34 (33.5 percent) lived in households with only wireless telephones as of the January-June 2009 period, the most recent period for which these data were available when MATS 2010 was being planned.⁴

Cell Phone-mostly (“Cell-mostly”) Households

The cell phone frame was comprised of more than just cell-only persons. The majority of cell frame numbers are associated with persons living in households with a landline number as well. To conduct interviews of these dual-frame persons

³ Delnevo, C, Gundersen DA, Hagman, BT. Declining Estimated Prevalence of Alcohol Drinking and Smoking among Young Adults Nationally: Artifacts of Sample Undercoverage? *Am. J. Epidemiol.* (2008) 167 (1): 15-19.

⁴ Blumberg, SJ, Luke, JV. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January - June 2009. Released 12/16/2009.



from the cell frame would have been in essence to oversample such persons. Oversampling of all such persons would have caused the sample design to be less efficient. However, previous research of dual-frame persons from the cell frame had indicated that the coverage of such persons was differential by their landline/cell usage patterns. In particular, those who received most or all of the calls through their cell phone were shown to have better coverage through the cell frame than through the landline frame. Additionally, from previous research, patterns of under-identification of young adults have been observed within “mixed-aged” households (those with both young adults and older adults) from traditional landline samples. It seemed plausible that the inclusion of cell-mostly households from the cell frame could potentially address these issues to some degree. So, for this segment of the dual-frame persons, it was felt that the relative coverage improvement outweighed the relative inefficiency of oversampling these persons. For the remaining dual-frame segment, where most of their calls were not being received on their cell phone, it was felt to be less efficient to include them in the sample.

Accordingly, the MATS 2010 cell phone RDD screener asked questions about the use of both landline and cell phones; cell phone screener respondents whose households were cell-only or cell-mostly were included in the cell sample, while those households that were landline-mostly were dropped from the sample; however, this group still had the chance of being sampled through the landline frame. Both the cell-mostly and the landline-mostly are dual-frame populations, that is, they exist on both frames and can be sampled through either one. As discussed in Chapter 5, the weighting of the combined landline and cell phone frames to produce the single, final MATS 2010 data file had to make adjustments for the dual probability of selection of any dual-frame cases included in the file. One goal of the weighting process is to reduce the sampling variance resulting from the complex sample design. The relative mix of landline-only, landline-some/cell-some, cell-mostly, and cell-only households in the overall population led to a design where retaining the cell-mostly households but dropping the landline-mostly households identified when screening the cell sample produced the best trade-off of reduced sampling variance and data collection cost across the entire combined sample.



Sample Allocation

The expected proportion of adults in cell-only and cell-mostly households was about 30 percent. However, it is well established that the cost of obtaining a completed cell phone interview is substantially greater than for a landline interview. With this in mind, MATS 2010 employed an undersample of cell phone numbers in an effort to make the survey design more cost-efficient, that is, achieve the desired degree of sample reliability and improved coverage at a lesser cost than would be required for a totally proportional cell-landline sample allocation.

With the cell-to-landline interview cost ratio estimated to be around 4 to 1 under the cell-only/cell-mostly design, the most efficient sample design was a reduced allocation of 15 percent of the total completed interview sample to the MATS 2010 cell frame sample.

This resulting 85/15 percent landline/cell frame allocation of 7,000 completed interviews corresponded to a design of 5,950 landline and 1,050 cell completed interviews.

2.1 Random-Digit Dialing (RDD) Landline Sample

While some modifications were made, the basic RDD landline sample design for MATS 2010 followed the CDC BRFSS protocols, this mainly being that the sample of randomly generated telephone numbers was selected from banks of numbers that contain at least one “listed” telephone number. (A bank is a series of 100 consecutive telephone numbers with the same area code, exchange, and first two digits of the line number.) No sample was selected from banks with zero listed numbers. The sampling frame consisted of area code/exchange combinations located in the state of Minnesota.

For purposes of the RDD telephone screening process, the cases eligible for MATS 2010 were defined as sampled phone numbers associated with a residence located in the state of Minnesota. Non-residential phone numbers or those associated with a residence outside of Minnesota were dropped as ineligible. Even though the sample frame was limited to Minnesota area codes/exchanges, it was possible for some

numbers to be located out of state, due to some overlap at state borders or other circumstances in the assignment of phone numbers by telecommunications companies serving Minnesota residences.

A table with the landline sample design projections and actual sample counts is provided in Table 2-1 in Section 2.1.4.

2.1.1 MATS 2010 Refinement of Basic RDD Landline Sampling Methodology: Refusal Conversion Subsample

The initial sampling plan for MATS 2010 used a disproportionate stratification technique that is based on refusals. For this procedure, Westat drew a larger sample of landline telephone numbers than would otherwise be selected to achieve the target number of completed interviews and then randomly pre-designated each telephone number to be assigned for household screener refusal conversion or not, if the screener was refused when the number was called. MATS 2010 used a subsampling rate of 60 percent to flag landline cases for which refusal conversion would be attempted, if the household member refused to answer the screener questions. The rate of 60 percent was based on projected trade-off between cost efficiency and sampling variance. This same technique was used in MATS 2007. For reasons described in section 3.7.4.1, this subsampling was eliminated during data collection and *all* landline screener refusals were ultimately subject to refusal conversion attempts. (As called for in the original design, all cell phone screener refusals were subject to refusal conversion attempts.) See sections 3.4.1 and 3.6.4 for more information about the refusal conversion process.

2.1.2 Landline Sample Telephone Number Duplicates of MATS Cohort Sample Telephone Numbers

ClearWay Minnesota enrolled a cohort of current and former smokers identified in MATS 2007 into the MATS Cohort Study, a longitudinal survey that collected data from these individuals in 2008, 2009, and 2010. Before being assigned to data collection, the sampled landline numbers for MATS 2010 were compared to the current phone numbers of record for the cohort study sample members. To avoid possible burden and confusion, 15 landline sample numbers that matched a number belonging to the MATS Cohort sample were dropped from the MATS 2010 sample.

These 15 cases appear as the “Overlap telephone numbers with MATS Cohort sample” in Table 2-1.

2.1.3 Within-Household Random Selection of Adult

The MATS 2010 sample design called for one adult at least 18 years old to be selected at random from each household that was identified through the RDD screening process. To select an adult from within a household, the Rizzo method⁵ was used to select an adult from within a sampled household.

2.1.4 RDD Landline Sample: Projected vs. Actual Counts

A random sample of landline telephone numbers was generated using the GENESYS Sampling System. These sample numbers were randomly formed into a number of “release” groups. Release groups allow for the controlled, random release of sampled numbers, so that yields of completed interviews can be closely monitored and additional groups released to achieve the desired number of completed interviews, once the yield patterns become established. After an initial set of release groups was assigned to the MATS 2010 telephone interviewing operation, additional groups were released, as needed, to ensure that the goal of 5,950 completed landline interviews was met.

During data collection, the yields of completed interviews from the cell phone sample considerably exceeded the original projections; consequently, the allocation of landline cases in the 7,000 completed interviews was reduced and that of the cell phone cases increased. See Table 2-1 for a breakdown of projected and actual RDD landline sample yields, which resulted in 5,555 completed landline interviews. Full details on the RDD data collection operational results appear in Chapter 4.

See Chapter 4 for more detail on the outcomes of the RDD landline telephone interviewing operations.

⁵ Rizzo L, Brick JM, Park I. A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys. *Public Opinion Quarterly* 2004 68(2):267-274.

Table 2-1. MATS 2010 RDD landline sample – design projections and actual

| Sample outcomes | Design | Actual |
|--|---------------|---------------|
| Total sample drawn (including reserve) | 104,533 | 105,000 |
| Total sample needed | 69,689 | 61,702 |
| Electronically purged numbers (non-working & business) | 38,329 | 35,971 |
| Duplicates of MATS Cohort sample numbers | - | 15 |
| Released to telephone interviewing | 31,360 | 25,731 |
| Non-residential/ineligible | 10,757 | 9,068 |
| Non-response | 6,814 | 2,860 |
| Unknown eligibility | 5,748 | 5,987 |
| Households screened | 8,041 | 7,816 |
| Extended interview non-response | 2,091 | 2,261 |
| Total extended interviews | 5,950 | 5,555 |

2.2 RDD Cell Phone Sample

The RDD cell sample design for MATS 2010 called for a sample of randomly generated telephone numbers that are contained within the universe of telephone number classified as cell phone numbers.

A table with the cell sample design projections and actual sample counts is provided in Table 2-2 in Section 2.2.3. It should be noted that the actual “Released to telephone interviewing” count of 18,629 included 178 cell phone numbers that were actually selected in the landline sampling process, but were later identified as cell phone numbers during the process of purging non-working and non-residential telephone numbers from the landline sample after sample selection.

2.2.1 Screening for Cell-Only and Cell-Mostly Households

As with the landline sample, cell phone numbers not associated with a residence in the state of Minnesota were screened out. However, additional screening of the cell phone sample was required. The cell phone sample screening also screened out cell phone sample cases that also had a landline number, but that did not use their cell phone most of the time (the landline-some/cell-some cases), as described above at the beginning of this chapter.



2.2.2 Within-“Household” Random Selection of Adult

As with the landline sample, the cell sample design called for one adult to be selected at random from each household that was identified through the RDD screening process. While researchers hold different opinions on whether a cell phone is a person-level or household-level device, studies have shown that a non-negligible proportion of cell phone users share their phones with other household members, a non-rare phenomenon even among cell phone-only and cell-mostly households. So, in MATS 2010, cell phones were treated as household devices. During the screener interview, the respondent was asked whether other household members shared the cell phone, and if the answer was yes, the Rizzo method was employed to conduct the within-household sampling before the extended interview. If the answer was no (no one else shared the phone), then the person answering the phone was by default the respondent, unless the phone user was under 18 years old and therefore ineligible for the survey. Further, it is common for minors to use cell phones, either their own personal phone that only they use or one that they share with other household members. If a minor answered a cell phone screener call, the MATS 2010 screening protocol determined if he or she was the sole user of the phone. If they were the sole user, the phone number was considered as ineligible for MATS 2010. If they shared the phone with other household members, the protocol determined whether any of these were age-eligible adults and, if so, sought to conduct the screening interview with an adult household member (since minors are not eligible to serve as RDD screener respondents).

2.2.3 RDD Cell Phone Sample: Projected vs. Actual Counts

Table 2-2 presents a breakdown of projected and actual cell phone sample yields, which resulted in 1,502 completed cell phone interviews. Full details on the cell phone data collection operational results appear in Chapter 4.

**Table 2-2. MATS 2010 RDD cell phone sample – design projections and actual**

| Sample outcomes | Design | Actual |
|--|---------------|---------------|
| Total sample drawn (including reserve) | 73,341 | 74,000 |
| Total sample needed | 36,670 | 18,629 |
| Released to telephone interviewing | 36,670 | 18,629 |
| Non-residential/ineligible | 22,668 | 10,558 |
| Non-response | 4,910 | 2,751 |
| Unknown eligibility | 7,425 | 3,459 |
| Households screened | 1,667 | 1,861 |
| Extended interview non-response | 617 | 359 |
| Total extended interviews | 1,050 | 1,502 |

See Chapter 4 for more detail on the outcomes of the RDD cell telephone interviewing operations.



3. Data Collection Methodology

3.1 Questionnaire Development

MATS 2010 required developing two questionnaires. The main questionnaire to be developed was the substantive survey instrument containing all of the questions for the MATS 2010 adult tobacco survey interview. In addition, the RDD sample needed a household screening questionnaire, some form of which is used in every RDD survey to identify households and then identify and sample people within the households. For brevity, the household screening questionnaire is generally referred to as the “screener” and the MATS 2010 adult tobacco survey questionnaire as the “extended” questionnaire; the latter term is used by many survey researchers because this stage of an RDD interview “extends” from the household screening interview. The same extended questionnaire was used for both the landline and cell phone samples. The screener for the cell phone sample required some questions not contained in the landline sample screener, to obtain the information related to the issues discussed above in Sections 2.2.1-2.2.2, such as ownership of the cell phone, usage of the phone, and state of residence of the phone’s user.

3.1.1 MATS 2010 Questionnaire

The MATS 2010 Advisory Panel began the process of designing the MATS 2010 instrument in the summer of 2009 by reviewing the MATS 2007 instrument and proposing items to be added, eliminated or reworded. Applying an iterative, consensus approach, the Panel worked through various versions, adding items to address new research questions or provide further information about previous research questions. The proposed changes reflected the current research agenda of MATS 2010, the experience with the utility of MATS 2007 data, the need to eliminate some items to accommodate new items, and the desire to somewhat reduce the overall length of the interview. The final decision to eliminate a question or panel of questions usually reflected a general consensus that the eliminated items were of interest in the past but not in the present, or were of lesser importance, given the need to obtain different information in MATS 2010. Westat began working with the Advisory Panel to refine and finalize the design of the

questionnaire in September 2009. While some items remained to be added or eliminated in mutual discussions, Westat focused on working with Advisory Panel representatives on question wording, response category selection and wording, question flow and ordering, and optimizing the design for telephone interviewing.

As noted in Section 1.1, maintaining continuity with the previous MATS surveys was a critical objective. However, changes in questionnaire design are desirable or unavoidable in large-scale surveys repeated over long time periods, due to the emergence of new issues or phenomena in regard to tobacco control, epidemiology, treatment, and education; scientific advances; altered focus on the part of the researchers, administrators, and practitioners who use the time series data; and the impacts of real world occurrences, such as political forces, actions of the tobacco industry, funding limitations, or social factors. Maintaining continuity in the MATS 2010 questionnaire was a balancing act between absolute conformity and making desired or necessary improvements.

Examples of substantial changes made in the MATS 2010 instrument include the elimination of questions about:

- Social settings in which people smoke;
- Source of payment for smoking cessation aids;
- Awareness of ads or commercials encouraging smokers to quit or about the dangers of secondhand smoke; and
- Knowledge about various harmful effects of smoking.

In addition, various questions about second-hand smoke policies and opinions were eliminated (such as the smoking ban in bars and restaurants) and replaced with others that are of current interest or relevance (see below).

The elimination of these questions is consistent with the changing relative importance of the research questions, the relevance of the question to the current circumstances in 2010, or the availability of the information from other sources. For example, most people, smokers and non-smokers, are now highly aware of the harmful effects of smoking, on the smoker and others.

Major additions to the MATS 2010 instrument allow exploration of new research questions about:

- People’s behavior to save money on cigarettes (to assess compensating for tobacco tax increases, tobacco industry price increases, and problems in the economy and employment situations);
- Use of various, newly emerging alternative tobacco products;
- Emerging issues about smoke-free policies in specific public spaces, particularly casinos, and in private vehicles when children are present; and
- Mental health and sense of financial well-being in relation to smoking.

Like the questions that are eliminated, the added questions are consistent with the changing relative importance of previous research questions to ClearWay and with the new MATS 2010 research questions.

In MATS 2010, questions about the use of less common, alternative tobacco products (such as hookah) were expanded and preceded the questions about traditional non-cigarette products (such as cigars, pipes, and smokeless), rather following them, as in MATS 2007. Thus, the following questions now preceded the traditional non-cigarette use questions:

Have you ever used any of the following products?

- a. A hookah water pipe?
- b. Electronic cigarettes, such as “Smoking Everywhere” or “NJoy”?
- c. Snus, such as “Camel Snus” or “Tourney Snus”?
- d. Any tobacco product that dissolves in the mouth, such as tobacco tablets, sticks, or strips?

Asking about smokeless tobacco use after this sequence, alerting respondents to some additional varieties of smokeless tobacco, might marginally improve the validity of the lifetime incidence/point prevalence estimates obtained for the overall smokeless tobacco use question in 2010.



Similarly, the questions about the harmfulness of various tobacco products were expanded. The MATS 2007 questions were:

In your opinion, are any of the following products less harmful, more harmful, or just as harmful as smoking cigarettes?

- a. Smoking tobacco in a hookah pipe?
- b. Smokeless tobacco such as snuff and chewing tobacco?
- c. Light or ultra light cigarettes?
- d. Natural cigarettes like Native Spirit cigarettes?
- e. Roll-your-own cigarettes?

MATS 2010 added two items and moved old item b (smokeless tobacco) to the end of the list, as item g:

In your opinion, are any of the following products less harmful, more harmful, or just as harmful as smoking cigarettes?

- a. Smoking tobacco in a hookah pipe?
- b. Light or ultra light cigarettes?
- c. Natural cigarettes like Native Spirit cigarettes?
- d. Roll-your-own cigarettes?
- e. Electronic cigarettes?
- f. Snus, a new smokeless, moist, pouch tobacco product, such as Camel Snus?
- g. Other smokeless tobacco, such as snuff and chewing tobacco?

Snus is a form of smokeless tobacco. MATS 2010 asked explicitly about the perceived harmfulness of snus. Adding this item made it necessary to cue the respondent to exclude snus from consideration in response to the general smokeless tobacco item, to avoid double-counting snus in both items f and g. To accomplish

this, MATS 2010 moved the general smokeless tobacco question to the position immediately following the snus question, and referred to “other” smokeless tobacco, in order to decouple snus from the response.

Three constructs were retained in MATS 2010, but with alterations that may somewhat reduce comparability with MATS 2007.

First, the question about binge drinking in MATS 2007 used the standard of 5 drinks on a single occasion, for all respondents; in 2010, the question is gender-specific: 5 drinks for males and 4 drinks for females. This is consistent with the practice in BRFSS and is based on the different average body masses of males and females and the relationship of body mass to alcohol absorption. This question change may marginally increase the reporting of binge drinking by females compared to what would have been observed using the previous standard.

Second, MATS 2010 retained the questions about people’s use of health care providers and their experiences with them in regard to smoking. However, MATS 2007 asked separately about four types of providers (doctors, nurses, dentists, and pharmacists), while MATS 2010 asked about providers overall. This change was made to reduce interview length, respondent burden, and questionnaire complexity; however, people’s general experience with the health care system (as compared to specific provider types) was the major aspect of this research question in 2007, and the 2010 revisions still collected the information at this level. To minimize the effect of asking the single, global form of the questions about health care providers, MATS 2010 specifically named the four types of providers of interest, and each follow-on question triggered the respondent anew to think about the entire list by referring back to “these health professionals you saw...”

Third, MATS 2010 revised the wording about the use of smokeless tobacco from

Have you used smokeless tobacco such as chewing tobacco or snuff at least 20 times in your life?

to

Have you used any kind of smokeless tobacco such as chewing tobacco, snuff, or snus at least 20 times in your life?

Since snus is a form of smokeless tobacco and is being heavily test-marketed and promoted by the tobacco industry, it was decided to include it specifically in the list of examples. Further, since dissolvables or other new products are also emerging in the smokeless category, it was decided to cue the respondents to think broadly about smokeless tobacco by inserting the words “any kind of” before “smokeless tobacco.”

The final MATS 2010 questionnaire appears as Appendix A of this methods report. In addition to developing this interview script format of the questionnaire during the design process, Westat also developed two tabular formats for documentation and quality control. One was a detailed table showing the skip patterns for every question for every smoking status or other criteria that affected skip patterns, which appears as Appendix B. The second was a detailed table crosswalking and documenting every question or response category added, deleted, or changed from MATS 2007 to MATS 2010, along with an assessment of its possible impact on data comparability between MATS 2007 and MATS 2010. This table is incorporated in the Minnesota Adult Tobacco Survey 2010 Comparability Report, which provided further details of the changes and additions that resulted in the MATS 2010 questionnaire.

The final MATS 2010 questionnaire covered domains such as general physical and mental health, alcohol use, cigarette smoking and other tobacco use, smoking cessation, experience with health care provider smoking interventions, attitudes towards smoking, situational exposure to secondhand smoke in various settings, the effects of public and private policies and rules on smoking behaviors and perceptions, and demographic information.

Westat developed detailed specifications to program the MATS 2010 questionnaire as a CATI survey instrument. The programming specifications are embedded in the MATS 2010 instrument included as Appendix A.

3.1.2 MATS 2010 RDD Household Screeners

Appendix C contains the MATS 2010 RDD landline household screener instrument and Appendix D contains the RDD cell phone screener instrument. The landline screener was a standard RDD screener, as adapted by Westat to implement the Rizzo method for RDD respondent selection. The cell phone screener also used the Rizzo method when necessary (i.e., when the cell phone number was used by more than one person to receive calls). The cell phone screener also incorporated various questions needed to determine sampling eligibility as described above in Section 2.2 and to select the respondent for the interview. These included questions to:

- Confirm that the phone number is a cell phone number;
- Confirm the cell phone number belongs to a Minnesota resident;
- Exclude cell phone numbers used exclusively by minors;
- Determine the owner or primary user of the cell phone;
- Determine the degree to which the household receives its calls by landline, cell, or both;
- Determine which adults in the household receive calls on the sampled cell phone number; and
- Collect mailing addresses for purposes of paying a \$5 cash incentive to individuals who completed the cell phone screener (see section 3.6.3 below for details about this incentive).

3.1.3 CATI Questionnaire Programming and Testing

Programming of the CATI questionnaire was carried out by Westat's CATI programming team, led by a senior CATI systems analyst. Testing of the programmed instrument was performed by the programmers, by an independent testing department at Westat, and by questionnaire designers from Westat and the MATS 2010 Advisory Panel. The several levels of testing revealed a few items that required correction and a few items that resulted in minor changes to the instrument design and specification.

3.2 Pilot Test

3.2.1 Background

Between February 1 and February 5, 2010, Westat conducted a pilot test of the RDD survey, including the MATS 2010 questionnaire, the landline and cell phone screeners, the within-household sampling procedures, the interviewer scripts and telephone contact procedures, and the handling of the cases in the CATI system's automated scheduling and case management system. The pilot test objectives were live field testing of the:

1. Programming of the CATI questionnaires;
2. MATS 2010 questionnaire's suitability for administration by interviewers;
3. Respondents' comprehension of the questions and their ability to provide answers;
4. Screening questionnaires, screening rules and procedures, and respondent selection; and
5. Assumptions for RDD landline and cell phone number sample yields, screener and individual cooperation rates.

3.2.2 Pilot Test Operations

The pilot test had a goal of 100 completed interviews, 70 from the landline sample and 30 from the cell sample. An initial sample of 838 landline numbers and 1,202 cell phone numbers was assigned to data collection. The pilot data collection employed substantially all of the data collection procedures to be implemented in the full survey, with two planned exceptions. Because the data collection period of the pilot test was brief and the primary objective was to test the instrument, the pilot test purposely did not carry out the refusal conversion protocol for either the household screener or the extended interview, as planned for the full survey. The pilot test sample also did not receive any of the supporting letters (advance letter, non-contact letter, and refusal conversion letter) that the actual survey sample received (see Section 3.4.1 for a full description of these letters).

3.2.3 Pilot Test Interview Monitoring and Interviewer Debriefing

During the pilot test, Westat's telephone supervisory staff conducted live monitoring of the interviews. Monitors could hear both sides of the conversations and see on their computer screens a live, mirrored version of the interviewer's actual CATI screen. During regular survey operations, the monitoring is conducted as a quality assurance measure of the interviewer's following of the data collection protocols, correct reading of the questionnaire text, handling of questions and problems, and entry of responses. Since experienced interviewers were assigned to the pilot test, the monitors were able to focus on the aspects of the data collection design that were being subject to testing: whether the procedures worked as planned, whether the questionnaire wording and flow supported clear administration by the interviewers, and whether the respondents had any general difficulties in understanding the questions or formulating an answer. The monitoring produced no reports of general problems along these lines. Interviewers were able to handle questions that some respondents asked about the survey or specific questions, based on their training and utilizing the set of Frequently Asked Questions developed for them by the survey managers.

When the interviews were completed, Westat's telephone operations manager, two MATS 2010 project managers, and MATS Advisory Panel members from ClearWay and MDH held a focused one-hour debriefing session with 9 data collectors and 2 supervisors. The debriefing session was conducted in person and by telephone. It consisted of a discussion in response to a list of questions about the interview designed to elicit both respondent and data collectors' reactions to the questionnaire design and the interviewing experience. The questions were both general and structured and focused on specific questions that were new to the instrument.

The data collection staff were positive about the questionnaire in terms of their ability to administer it clearly and of respondents' ability to provide answers to the questions with little difficulty. Data collectors noted without prompting that respondents were very receptive to the study when they learned that the sponsor was the Minnesota Department of Health; the data collectors, many of whom had previous experience with cell phone surveys, reported that MDH sponsorship was

especially helpful in gaining cooperation on the cell phones. As a result of the data collector comments, a few additional instructions or definitions were added to the CATI screens for the full study.

The pilot test revealed no significant problems with questions new to MATS 2010. Some minor wording changes were made to a few questions to make them clearer to the respondents or to assist them in providing answers consistent with the intent of the question.

The statisticians examined the various yields of the pilot test, to the extent that they were predictive of yields in the full study. The pilot test was not designed to predict yields, given the unknown effect of deliberate pilot test plans, such as stopping the test when the desired number of completed interviews was reached, not using the full study mailing protocol, and not implementing refusal conversion (see Sections 3.4.1 and 3.6.4 below). However, absent any strong negative indicators, it seemed likely that the household screener and adult tobacco survey extended interview would achieve the response rates and cooperation rates underlying the sample plan for both the landline and cell phone samples. In fact, even with the truncated calling protocols, the yield rate associated with the completion of 54 extended cell phone interviews was somewhat greater than expected. While the rates of cell phone cases that were ineligible according to the several MATS 2010 eligibility criteria – cell phone users less than 18 years old, living outside of Minnesota, or not “Cell-Only” or Cell-Mainly” – were fairly close to expectations, the percent of cell phone numbers that were non-residential/not working (35 percent) was smaller than had been conservatively estimated (55 percent). Based on this positive pilot test result, the size of the initial cell phone sample release for the main study was scaled back.

The pilot interview length timings were very close to the desired overall average. The MATS 2010 pilot test revealed that the design of the MATS 2010 questionnaire successfully achieved its objectives, in terms of obtaining the desired information, being clear and unburdensome to respondents, and readily administered by interviewers. The live test also confirmed that the CATI instrument performed correctly as to flow and data capture, as intended by the design and as previously verified by beta testing. Perhaps most importantly, the test showed that the overall design of MATS 2010 – interviewer training, calling procedures, the RDD screening,

explaining the purpose of MATS 2010, identifying and obtaining cooperation from selected individuals, and successfully taking them through the MATS 2010 questionnaire – was well considered and feasible in a real world setting.

3.2.4 Pilot Test Findings

As a result of the pilot test, ClearWay, MDH, and Westat agreed to a few minor changes to the survey instrument to improve specific areas. As described more fully in the Minnesota Adult Tobacco Survey 2010 Pilot Test Report⁶ (Westat, 2010), the principal questionnaire changes were:

Questionnaire

Addition of various explanations, definitions, and descriptions for data collectors to use as needed.

Sampling

The percent of cell phone numbers that were non-residential/not working (35 percent) was smaller than had conservatively been estimated. Based on this favorable result, the size of the initial cell phone sample release for the main study was scaled back.

Westat presented the specific findings and additional details about the pilot test in the MATS 2010 Pilot Test Report.

3.3 Interviewer Recruitment and Training

Westat assigned interviewers from its current staff of interviewers and recruited additional staff as needed to supplement current staff. The additional staff were located either in call centers or worked from their homes. All interviewers received two waves of training: general interviewer training and MATS-specific training. (Current interviewers had previously received the general interviewer training; newly recruited interviewers received both). The training was self-paced and self-

⁶ Readers interested in more information about this report may contact Ann St. Claire, ClearWay Minnesota, at (952) 767-1416 or astclaire@clearwaymn.org.



administered, with the interviewers working through Web-based self-study modules. Before an interviewer could progress to the next module, he/she had to pass an assessment on the module just completed, with a 100 percent correct score. Interviewers could review content until they were able to attain 100 percent correct on all assessments.

The general interviewer training (GIT) modules covered topics such as:

- The concept of data and social science research, and the role of the interviewer in this research process
- Principles and tenets of standardized interviewing and the use of the CATI system
- Concept of a scientific sample and the importance of probability sampling; the various ways data can be collected: in-person, telephone, Web, observation, medical measurement, etc.
- Survey design, administration, and respondent contact procedures
- Standard call disposition codes
- Importance of interviewer neutrality, verbatim question delivery, and exact recording of responses as central to standardized interviewing
- Projecting professionalism and expert knowledge of the survey as key characteristics in securing respondent cooperation
- Listening skills and speaking skills
- Voluntary nature of survey participation, informed consent, and confidentiality

Training in use of the CATI system employed an interactive, self-administered, computer-based tutorial. Each interviewer moved through a series of topics, such as instruction on logging onto the CATI system, using the keyboard, the mouse, and special CATI commands. At this point interviewers learned the keys and commands for entering data and handling situations outside of the automated flow of the CATI questionnaire. Also included in this session was practice in the coding

of contact results. Trainees experienced recorded replications of common contact situations and learned the proper coding techniques through presentation and practice. A follow-up test was administered to evaluate mastery of the contact procedures. The interviewers who were considered for the MATS assignment and to receive the MATS-specific training were limited to those who achieved a perfect score (100 percent) on this test.

Westat based the MATS-specific training on the BRFSS training protocol as described and demonstrated at the BRFSS website

<http://www.cdc.gov/brfss/training/interviewer/index.htm>. This protocol emphasizes the presentation of questions, instructions to interviewers on administering the questions, and the rationale for the questions, so that the interviewers understand the importance of following the protocol.

The MATS-specific training was conducted in the same way as the general interviewer training, including the requirement to score 100 percent correct on each module's evaluation assessment before proceeding to the next training module. The modules focused on:

- MATS questionnaire items and the flow of the MATS questionnaire, including terms and definitions
- RDD screening process
- Contact scripts
- Handling of problem situations and the use of the prepared, standard responses to frequently asked questions

Training instructors and team leaders were available to assist interviewers should they encounter difficulty with a particular training concept or module. MATS trainers communicated with interviewers through instant messaging, training-based electronic bulletin boards, email, and telephone calls. Following the self-administered training modules, interviewers attended a live, web-based training session with a live trainer. The trainer reinforced concepts learned through the self-paced trainings, moving through practice versions of the questionnaire and allowing the trainees to practice administering the questionnaire. After interviewers

completed all of their training modules including the live web-based session, they were teamed with a partner to conduct practice interviews with each other (role plays). During the role play sessions each interviewer was monitored and coached to assess and enhance their interviewing skills. Once the supervisor determined the interviewer demonstrated the appropriate command of the interview and study materials, the interviewer was permitted to conduct actual MATS interviews.

3.4 Communications with Sample Members

There were a variety of methods used to communicate with the MATS 2010 sample prior to and during data collection. These included a variety of letters, an informational website, and several contact numbers that potential respondents could call for information or other purposes. These tools were designed to improve response rates and provide information to sample members or to the general public about the survey.

3.4.1 Letters

MATS 2010 developed three different types of letters sent to members of the RDD landline sample. These were an advance letter, a non-contact letter and a refusal conversion letter. All RDD letters were printed and mailed using letterhead of the Minnesota Department of Health and signed by the Assistant Director of the MDH Division of Health Policy. Because it is not possible to reverse match cell phone numbers to addresses, MATS 2010 did not implement any mailings for the cell phone sample.

- The **advance letter** was mailed to each address that could be associated through a listed number database with a sampled RDD landline number that was in the released sample. It was addressed generically to “The Household at...” the matched address. Its purpose was to inform the household of their possible inclusion in the survey, give them information about it, stress the voluntary and confidential nature of the interview, urge their participation if selected, provide them a reference to the MATS 2010 website (described in Section 3.4.2) and the phone number at MDH that they could call for more information.



- The **non-contact letter** was a variant of the advance letter. It was mailed when, after repeated calls to an RDD landline number, no contact had been made that would allow determination of whether or not the phone number belonged to a residence. Like the advance letter, the non-contact letter could be sent only for phone numbers that had been matched to an address. The non-contact letter was designed to get through to people who may have been screening calls through Caller ID, who may have had a phone line to which no phone was connected, or similar reason why contact could not be made. The letter stressed the importance and legitimacy of the survey and urged the recipient to respond to calls from Westat or to call Westat's toll-free number.
- The **refusal conversion** letter was mailed when a phone number had been established as belonging to a household but the members refused to participate in the household screening interview. Like the advance letter, the refusal conversion letter could be sent only for phone numbers that had been matched to an address. The refusal conversion letter was designed to persuade the household to participate in the screening and then in the extended interview. It contained much of the information included in the advance letter, with additional emphasis on the importance of the recipient's participation.

Copies of the three letters appear in Appendix E.

3.4.2 MATS 2010 Informational Website

The MATS 2010 Advisory Panel and Westat web designers developed an informational website to provide sample members and potential respondents with a set of brief, simple, and clear informational points about the survey. Its purpose was to encourage participation among selected respondents, enhance the perceived legitimacy of the survey, and answer questions potential participants might have. Legitimacy was enhanced by the visibility of the web page on the Minnesota Department of Health's official website, at www.health.state.mn.us/2010HealthStudy.



The main web page provided a brief overview of the survey with menu links to four subpages covering the topics:

- How participants are selected
- How the survey works
- Frequently asked questions
- Sponsoring agencies and contact information

In the letters and web pages, MATS was characterized as a health study with an emphasis on tobacco rather than exclusively as a tobacco survey. This was designed to avert non-smokers from a disinclination to participate due to a perceived lack of relevance to them and to mitigate smokers' possible perceptions of persistent focus on them by media, government, and the health care community.

The contents of the MATS 2010 informational website appear in Appendix F.

3.4.3 Toll-free Numbers and Contacts Provided by Westat and the Minnesota Department of Health

Westat operated a toll-free number that MATS 2010 sample members could call to obtain information about the survey. Westat's inbound call center answered the toll-free line and either responded with the requested information or referred the caller to an assigned contact person at MDH. MDH provided the name and direct line of a contact person. These numbers and contacts were printed in the appropriate letters, were available on the website, and were provided upon request by the telephone interviewers.

3.5 Data Collection Confidentiality Procedures and Protection of Human Subjects

All Westat staff are bound by strict confidentiality and privacy rules and procedures that are designed to prevent deliberate or inadvertent disclosure of the identity or survey data of anyone belonging to a data collection sample. All Westat staff are trained in the relevant protocols, covering oral, electronic, or printed

disclosure, and in the techniques to safeguard such information in all of these forms. As a condition of employment, they are required to sign a pledge of confidentiality laying out these requirements. They undergo required annual training in the issues of human subjects protection and information security.

These general rules and procedures apply equally to center-based and home-based interviewers; home-based interviewers are subject to further requirements, in terms of working from a segregated office space within their home environment and outside of the presence of anyone else in the household. Using the web-based interface, all data collected by the home-based interviewers were entered in real time into the central survey database maintained within Westat's physical facility and behind Westat's software firewall. Sample identifying information, questionnaire text, and response data were only visible on the interviewers' screens; no data could be copied or saved electronically or printed locally.

All sample and survey data were maintained on Westat's secure, password protected network, with access to MATS-related data limited to approved MATS 2010 project staff.

The MATS 2010 survey questionnaire, data collection, and data security plan were reviewed and approved by the Westat Institutional Review Board (IRB), a specially constituted review body established to protect the welfare of human subjects recruited to participate in biomedical and behavioral research. Westat's responsibilities are detailed in the regulations concerning human subjects protections and the Multiple Project Assurance (MPA) granted by the U.S. Department of Health and Human Services, Office for Protection from Research Risks, Division of Human Subject Protection.

Westat's general confidentiality procedures are designed to comply with applicable requirements of state and federal law relating to Protected Health Information (PHI), including the Health Insurance Portability and Accountability Act of 1996 (HIPAA). PHI and HIPAA apply to health information contained in health records; collected survey data are covered by other federal statutes and is subject to the oversight of the Office of Protection from Research Risks.



3.6 Data Collection Operations

Data collection occurred between February 19, 2010 and May 30, 2010. Calling took place from 9 AM to 9 PM weekdays, 10 AM to 6 PM Saturdays, and 2 PM to 9 PM Sundays (all times Central time). Consistent with standard operational practice for personal telephone interviews of individuals in their homes, the majority of calls occurred between 6 PM and 9 PM weekdays and throughout the weekends, in order to optimize the amount of effort applied when people are most likely to be found at home.

3.6.1 Calling Procedures

3.6.1.1 General Case Handling and Contact Procedures

MATS 2010 telephone procedures applied a hierarchical approach to case management. This includes making cases available for call attempts based on the current status of the call. Those cases for which an exact appointment was scheduled had the highest priority, followed by those with a general call back time (e.g., information that “evenings are best” to reach the desired person), those that had been called previously without human contact and lastly cases that had never been called. Consistent with BRFSS protocol, at least 15 call attempts were made to each sampled number, unless the number needed fewer attempts to reach a natural final disposition. Until contact was established, the CATI scheduling system automatically spread out the calls across various times of day and various days, including weekdays and weekend days.

As described in section 3.6.4, a second effort was made to convert refusals to the RDD household screener or to the extended interview, except for those few characterized as “hostile” (or “adamant”) refusals.

3.6.1.2 Supplemental Calling Procedures

Once data collection was in progress, MATS 2010 implemented several measures designed to improve response rates and increase the number of completed interviews obtained from the released sample. The most significant of these was the decision to re-activate cases that had been coded as final non-response because they had reached the maximum number of call attempts according the BRFSS/MATS

protocol. Periodically during the second half of the data collection period, Westat re-activated such cases for another round of calls.

In order not to badger households, the MATS calling protocol allowed for only one answering machine message to be left at a phone number prior to the point when live contact was made. Towards the end of data collection, a second answering machine message was left at any phone number where live contact still had not been made.

3.6.1.3 Calling Rules Determined by Special Case Eligibility Rules

In the previous round of MATS, two issues emerged that required the MATS 2007 design team to develop handling rules that were consistent with the study's research objectives, statistical sampling practices, and data collection operational procedures. Their common feature was temporary residence out of state. Prior to the start of MATS 2010, these rules were reviewed and further refined.

3.6.1.3.1 *Snowbirds*

Discussions among the survey sponsors and Westat produced a protocol for handling "snowbirds," residents who leave the state for warmer climates in the winter. MATS 2010 considered them as valid Minnesota residents on an extended vacation. They were, therefore, eligible members of the sample, and MATS 2010 sought to interview them. However, because some MATS 2010 research questions address social and environmental factors and Minnesota policies and programs, the survey designers concluded that it was best to interview people only when they were physically within the state borders. If snowbirds currently dwelling out of state were identified during initial calls, arrangements were made to interview them upon their return to Minnesota, if they returned before the end of the data collection period. If they were not in the state during this period, they were not followed to their winter residence for an interview. Since they remained as eligible members of the sample, the latter group was classified as a form of survey non-response, rather than as ineligibles who could be dropped from the sample for response rate calculations and sample weighting.



3.6.1.3.2 *College Students Living Away from Home*

A common protocol for RDD surveys (based on fairly complex rules that the U.S. Census uses to define household membership) is that students who reside away from home are considered members of the household if they live in group quarters (such as a dormitory) but not if they reside in private or small common residential units (generally defined as those occupied by nine or fewer unrelated individuals).

The research issues that arose in regard to snowbirds similarly applied to the out-of-home student situation, and MATS slightly modified the commonly applied RDD rules for students.

Young adults who were found to be attending college in Minnesota were eligible to be interviewed, even if they were not currently residing in the household at the phone number of record (most likely their parents' house). The procedure was to call them at any phone number where they could be reached to conduct the interview, including calling back at the number of record if they would be available there before the end of data collection.

Young adults who were found to be attending college outside of Minnesota were classified as not eligible for the study, because they would be less exposed to the social and environmental factors and Minnesota policies and programs that were a focus of MATS 2010. They were dropped from the sample for response rate calculations and sample weighting.

The RDD screening protocol did not seek to explicitly identify students. The protocol was applied only in the situation where a student's status as residing temporarily away from home happened to emerge in conversation, whether with the sampled student or with someone else in the household. If this happened, the interviewer was instructed to ask if the student was attending school in state or out of state. If out of state, the interviewer recorded the case as a special problem with detailed comments for review by supervisory staff. If the supervisor concurred with the determination, the case was coded as ineligible; if not, it was reactivated with instructions to continue to pursue the interview with the student wherever he or she could be reached.

3.6.2 Home-based Interviewers

As noted above, MATS 2010 used traditional call center-based telephone interviewers and, mainly, telephone interviewers working from their homes. Over the past few years, distributed call center operations using home-based telephone staff networked through advanced web-based platforms have become standard practice in the survey research, customer support, and telephone counseling professions (such as tobacco telephone quitlines), among others.

Regardless of their location, the MATS 2010 center-based and home-based interviewers accessed the same CATI system with a common CATI instrument, survey database, sample management system, call scheduler, and autodialer, employing a secure, web-based interface. All interviewers received cases from a common sample management database, and the call scheduling/case management system transparently assigned cases to available interviewers. Using the web-based interface, all data collected by the home-based interviewers were entered in real time into the same central survey database accessed by the center-based interviewers, maintained within Westat's physical facility and behind Westat's software firewall. Sample identifying information, questionnaire text, and response data were only visible on the interviewers' screens; no data could be copied or saved electronically or printed locally.

Home-based interviewers were likewise monitored and received feedback using the same protocol, methods, and interface as for center-based interviewers. See section 3.7.1 for more information about interviewer monitoring.

3.6.3 Cell Phone Screener Incentive Payments

Many survey research companies, including Westat, interpret the TCPA (Telephone Consumer Protection Act of 1991) as prohibiting calls to cell phones unless the cost of the calls is reimbursed. Hence, the MATS 2010 offered the cell phone screener respondents a \$5 cash payment, designed to compensate them for any time usage costs incurred for responding to the interview. This promised payment may also have served as an incentive for the respondent to agree to the screener interview.

MATS 2010 offered this payment to all cell phone screener respondents, even if the screener determined the number was not eligible for MATS 2010, as described



above in section 2.2. This group included all eligible screeners, cell numbers determined to be non-residential (e.g., business use only), and cell numbers where the user was not a cell-only or cell-mostly user. In order to receive the cash payment, the respondents were requested to provide their mailing address information at the end of the screener. (For reasons of practicality and informed consent issues, MATS 2010 did not offer the payment or collect the address of any minor who was the sole user of a sampled cell phone number, which was therefore immediately screened out as an ineligible number.)

It is common for cell phone screener respondents to decline the payment offer, possibly because their cell phone calling plan allows unlimited calls or because they choose not to provide their mailing address. In MATS 2010 there were 3,897 screeners that were eligible to receive the payment; 2,452 (62.9 percent) of these chose to receive it and provided their mailing addresses.

3.6.4 Refusal Conversion

In scientific surveys, it is standard practice to recontact people who initially refuse to participate in an interview, in a second attempt to persuade them to participate. This refusal conversion process is designed to increase the sample size and response rate, and also to reduce bias associated with including in the sample only those who are most inclined to respond, i.e., those who respond immediately to the participation request. This section describes the MATS 2010 refusal conversion process for the RDD screeners and extended interviews. See Section 4.4.2. for the quantitative results of these refusal conversion efforts.

When interviewers encountered reluctant respondents, they first attempted to avoid the refusal by addressing any concerns expressed. When that attempt was unsuccessful, the interviewer coded the case a refusal and completed a non-interview report form. Included in this form was a brief description of the reason for and the strength of the refusal. The strength of the refusal was coded mild, firm or hostile (adamant). Mild and firm refusals were mostly determined at the discretion of the interviewer. Typically refusals were designated mild if the respondent hung up without explanation and firm if some type of reasoned

explanation was provided. The hostile designation was reserved for respondents who used vulgar language or were threatening.

Each refusal case was withheld from additional call attempts for a cooling off period of at least 13 days. After this cooling off period, specially trained interviewers attempted to recontact the phone number (for screener refusals) or the selected adult (for the extended interview), to persuade them to participate in the survey. If the respondent refused a second time, the case was finalized as a refusal and no further attempts were made to contact them. The interviewers selected for these conversion attempts demonstrated an above-average comprehension of the study objectives and ability to share this knowledge with the respondent. They received additional training to aid them in dealing with challenging situations.

3.7 Data Quality Assurance

Data quality assurance for MATS 2010 took a variety of forms prior to, during, and following data collection. Prior to data collection, data quality assurance was addressed through the questionnaire design, specification, and testing process described elsewhere. The valid generalizability of the collected data to the overall population was further assured by a well-designed and scientifically drawn sample. The techniques for designing and drawing the sample are described throughout chapter 2.

The present section addresses measures implemented to assure the quality of the data as collected during and following data collection. Such measures include monitoring interviewers and providing feedback to them, review of the actual data captured in the CATI system during data collection, ongoing monitoring of sample performance during data collection, and in the processing of the data into the final database once data collection is ended.

3.7.1 Interviewer Monitoring

Westat uses a silent monitoring system that allows supervisors to listen on the phone and to watch interviews on the CATI screen in real time without the interviewer or respondent knowing that they are being observed. Monitoring reports are completed for each monitoring session and reviewed with the



interviewer during each shift. This provided the opportunity to reinforce good skills and coach interviewers in areas needing improvement in a timely manner. Approximately 10 percent of all interview time was monitored.

3.7.2 Data Cleaning and Editing During Data Collection

The primary method for assuring the quality of the collected data is to address this objective, before the fact, in the design and programming of the CATI questionnaire, in the data collection protocol developed, and in the training of the interviewers in general best practices and the specifics of the MATS 2010 questionnaire and interviewing protocol. All of chapter 3 up to this point has addressed these issues in detail.

Even though the CATI system controls all skip patterns and allows only valid ranges of values to be entered by the interviewers, Westat data managers conducted additional reviews of the collected data after the fact.

The first review consisted of initial review of the frequency distributions of every survey variable during the data collection process. The CATI data manager reviewed the frequencies for each variable to check for any inconsistencies in the skip patterns or range violations. While rare in a well-tested CATI system, such errors may occur because of unusual situations not anticipated in the design or not revealed during testing. Such quality assurance allows discovering any such problems early in the process, making necessary corrections, and recalling affected respondents to obtain corrected data (data retrieval). This process did not reveal any errors in the CATI questionnaire programming for MATS 2010. Throughout the data collection process, the CATI data manager also reviewed comments noted by the interviewers in the CATI system. These comments might have been notes made by the interviewers themselves, or might have been extraneous comments made by the respondents and recorded by the interviewers. Often the comments required no further action. In some instances, the comment could be an indication that the respondent corrected their answer to a previous question, or the interviewer was unsure how to code a particular response. In these cases, the CATI data manager made any necessary edits to the data or referred the case to a MATS data collection manager for a decision. Any such edits were documented in the data edit log,

which contains both the original value that was recorded in the interview as well as the new, updated value for each respective variable where an edit occurred, along with a brief description detailing the reason for the edit.

3.7.3 Data Cleaning and Editing Following Data Collection

After data collection was completed, the MATS 2010 data delivery manager developed a SAS program that independently tested the data integrity rules and ascertained the follow through of all skip patterns. This SAS program served as a second layer of quality control to ensure the accuracy of the data integrity rules specified for the instrument. The program produced a detailed, case-by-case, variable-by-variable report if any errors were encountered. Errors in this context refer to instances where data were either missing, or data were present where they should not have been. The data manager reviewed the error report and made any necessary corrections to the data to accurately satisfy the data integrity rules of the instrument. The data delivery manager also rechecked each variable for values outside of the allowed ranges. All edits made to the dataset were documented in the same data edit log that was used for edits made by the CATI data manager during data collection.

3.7.3.1 Post-coding of Verbatim Text in 'Other-Specify' Questions

Once the data were cleaned, additional processing of the data occurred as a result of reviewing and recoding the text responses to the various open-ended 'Other-Specify' questions that appear in the MATS 2010 questionnaire. ClearWay, MDH, and Westat survey managers collaborated closely on all such post-coding and recoding decisions. In addition to creating post-codes for the 'Other-Specify' responses, this process also identified some instances where a categorical response to an earlier question in the same sequence as a given 'Other-Specify' question should be recoded, based on the additional information that the respondent furnished in the 'Other-Specify' response.

All updates made to the data as a result of this process were stored in the final dataset in newly created variables, and the data as originally collected were preserved in historical variables in the dataset. Westat documented the process in a review and recoding protocol document; a database at the record and variable level

that recorded all ‘Other-Specify’ text responses, post-codes, and recodes; and a set of summary tables of the outcomes of this process on each affected survey variable.

3.7.4 Sample Performance Monitoring During Data Collection

Throughout the data collection process, it was vital to monitor several outcomes of the data collection process, mainly to project estimated final totals from interim results and determine what adjustments were needed or possible, in order to support achieving the targeted number of completed interviews. Aside from standard weekly reports of case dispositions for the sample, there were two areas of particular interest to the survey sponsors and Westat technical managers.

1. From a sample management perspective, it was necessary to monitor the overall yield of completed interviews resulting from the sample release groups activated at the outset and then in subsequent releases, to assure achieving the target number of completed interviews.
2. Since overall smoking prevalence was the primary measure to be estimated by MATS 2010, it was desirable to monitor this during data collection, to assess whether there was an unreasonably large divergence from the trend or estimates produced by other surveys, such as BRFSS. If such divergence was observed, investigating whether it was being artificially affected by some aspect of the survey design would be in order, so as to make corrective adjustments early in the data collection process to compensate.

3.7.4.1 Monitoring the Overall Number of Interviews

It is standard practice in survey research to initially not release all of the originally projected number of sample release groups, and then monitor sample performance to determine if more sample is needed. In response to trends and patterns in sample performance identified through the monitoring process, the statisticians and operations managers updated the projections at several points during data collection to determine if the release of additional sample was warranted. Chapter 4 contains additional details about sample sizes.

The landline sample screener and extended response rates were close to expectations, only slightly lower than those observed in MATS 2007. However, for the cell phone sample, both the screener and extended response rates were higher than expected, resulting in a greater yield of completed interviews from the cell

phone sample than originally planned. As seen in Table 2-2 in Section 2.2.3, the actual yield of completed interviews was 1,502, substantially more than the design count of 1,050.

As mentioned above, fewer than the projected number of needed landline numbers were initially released for interviewing. With response rates coming in near expectations, this typically would have called for the release of additional sample; however this projected shortfall on the landline side was projected to be nearly “covered” by the higher yield emanating from the cell sample.

As data collection continued, it became clear that the lower number of landline completed interviews would not quite be covered by the higher number of additional cell completed interviews, resulting in projections of fewer than the target of 7,000 completed interviews. Given the latter stage of the data collection period, it was determined that the most efficient method to achieving the required goal was to allow refusal conversion attempts to be made on the 40 percent of landline screener refusal originally not selected for refusal conversion. Eliminating the subsampling of screener refusals and increasing the proportion of cell phone cases in the final sample were both likely to further reduce the sampling variance. (Subsampling adds another adjustment factor to the weighting process; cell phones were undersampled relative to the overall distribution in the population, so increasing their proportion moved the final sample closer to the natural distribution.) Both of these aspects of the original design were mainly for cost-efficiency trade-offs; at the point in data collection when this change was made, operational and cost analyses determined that making this change at this point was more efficient than adhering to the original design and releasing more landline sample to achieve the target sample size.

3.7.4.2 Monitoring Smoking Prevalence Rates

Smoking prevalence rates were monitored throughout data collection, and particularly at the point when the completed sample size was large enough to make the interim calculated rate predictive of the final estimates. The reported prevalence rate from the MATS 2007 survey was 17.0 percent. The expectation was that this percentage would drop slightly in MATS 2010, but a large difference between the MATS 2007 and the interim MATS 2010 prevalence rates in either direction could



indicate a potential problem with the questionnaire, the interviewing, or the data collection procedures that manifested itself in regards to smoking status classification.

There were no indications of any problems or issues based on the review of unweighted counts and proportions of respondents indicating their smoking status. The MATS 2010 unweighted proportions from the landline sample were very much in line with similar estimates from MATS 2007. This was true overall and by the two monitored age groups: 18-24 and 25+. In addition, comparisons were made between the landline and cell phone unweighted smoking status proportions. The observed relationships were all within expectations. Specifically, the unweighted smoking prevalence proportion from the cell phone sample was about twice the proportion for the landline sample. This should not be over-interpreted, but the general relationship was completely consistent with the fact that the unweighted cell sample is heavily skewed towards younger adults, who are more like to be smokers according to previous research on this population and their coverage in cell-only households.

4. Data Collection Results

Chapter 4 presents various statistics summarizing the outcomes of the MATS 2010 data collection, separately for the landline and cell phone samples and for the combined sample. The key statistics presented are the call dispositions and the response rates for these two sample groups.

4.1 Completed interviews

Table 4-1 presents the overall number of completed interviews for the landline and cell phone samples

Table 4-1. MATS 2010 completed interviews, by sample type

| | Landline | Cell | Total |
|----------------------|-----------------|-------------|--------------|
| Completed interviews | 5,555 | 1,502 | 7,057 |

4.2 Telephone Interviewing Results

4.2.1 Landline Sample Call Dispositions

Tables 4-2 and 4-3 show the detailed dispositions for all of the sampled landline telephone numbers that were released to data collection. Table 4-2 shows the dispositions for the screening of the 61,702 sampled landline numbers; Table 4-3 shows the dispositions of the 7,816 numbers for which a household screener was completed and from which an adult was selected for the MATS 2010 interview. These tables tabulate the actual disposition categories that Westat employed to manage the sample for the MATS 2010 interviewing operations. They also show the standard AAPOR disposition codes to which each lower-level MATS 2010 disposition category is mapped. (AAPOR is the American Association for Public Opinion Research, to which many survey researchers belong and whose members have established various standards for scientific survey research that are widely accepted.)

Note that AAPOR dispositions account for each sampled phone number. The two tables account for the results of the sampled household members in the 7,816

completed household screeners as a second level of detailed disposition codes. For this reason, there is no corresponding AAPOR code for these cases in Table 4-2; rather, the AAPOR codes for these cases appear in Table 4-3. Separating the cases into the two operational stages provides a clearer understanding of the landline data collection outcomes, while still allowing all 61,702 sampled landline telephone numbers to be classified as to their outcomes according to the standard AAPOR disposition codes.

4.2.2 Cell Phone Sample Call Dispositions

Tables 4-4 and 4-5 show the detailed dispositions for all of the sampled cell phone telephone numbers that were released to data collection. Table 4-4 shows the dispositions for the screening of the 18,629 sampled cell phone numbers; Table 4-5 shows the dispositions of the 1,861 numbers for which a household screener was completed and from which an adult was selected for the MATS 2010 interview.

4.2.3 Combined Sample Call Dispositions

Tables 4-6 and 4-7 show the detailed dispositions for all of the sampled telephone numbers (landline and cell combined) that were released to data collection. Table 4-6 shows the dispositions for the screening of the 80,331 sampled telephone numbers; Table 4-7 shows the dispositions of the 9,677 numbers for which a household screener was completed and from which an adult was selected for the MATS 2010 interview.

Table 4-2. MATS 2010 landline telephone number sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|---|--------------|----------------|
| 1.1 | Completed Screener | 7,816 | 12.7% |
| 4.70 | There is no one in the household age 18 or older to do the screener. | 30 | 0.0% |
| 4.10 | The sampled telephone number rings into a household not located in Minnesota. | 70 | 0.1% |
| 2.35 | All household members are currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 2 | 0.0% |
| 2.36 | Sampled telephone number is part of MATS Cohort Study. | 15 | 0.0% |

**Table 4-2. MATS 2010 landline telephone number sample dispositions
(continued)**

| AAPOR code | Description | Count | Percent |
|-------------------|--|---------------|----------------|
| 2.36 | Only eligible household member already completed study on another phone number. | 3 | 0.0% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 5 | 0.0% |
| 2.331 | Language Problem: unable to communicate due to a hearing or speech problem or unable to reach an English speaking household member. | 332 | 0.5% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one “human” contact has been made at the number and there are no refusals or language problems in the call history for the household. | 870 | 1.4% |
| 3.13 | The calling algorithm has been fulfilled with no “human” or answering machine contact. | 3,468 | 5.6% |
| 4.50 | The telephone number was identified as non-residential during business purge preprocessing prior to CATI load. | 2,798 | 4.5% |
| 3.14 | The maximum calling algorithm has been fulfilled for a telephone number and only answering machine contact was made. | 1,302 | 2.1% |
| 4.50 | The telephone number called was not a residential number. Included are businesses, institutions, agencies, modems, public facilities, vacation homes, group quarters. | 3,397 | 5.5% |
| 4.30 | The telephone number was identified as non-working during Tri-tone match processing prior to start of calling. | 32,302 | 52.4% |
| 4.30 | The telephone number was found to be not working when called. | 5,561 | 9.0% |
| 4.30 | The telephone number resulted in no ring-back during purge processing prior to CATI load. | 871 | 1.4% |
| 2.111 | Refusal – Household screener respondent refused to be interviewed or broke off during the screener interview. | 2,823 | 4.6% |
| 2.111 | Refusal – Screener refusal results from a call to the Minnesota Department of Health or the study’s toll-free line. | 37 | 0.1% |
| Total | | 61,702 | 100.0% |

Table 4-3. MATS 2010 landline extended interview sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|--|--------------|----------------|
| 1.1 | Completed interview. | 5,555 | 71.1% |
| 4.10 | Subject does not live in the state of Minnesota. | 1 | 0.0% |
| 2.35 | Subject is currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 1 | 0.0% |
| 4.10 | Subject is currently attending school out of state and will not be back before the end of data collection. | 12 | 0.2% |
| 2.332 | Language Problem: unable to communicate due to a hearing or speech problem or the selected respondent was unable to speak English. | 71 | 0.9% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one "human" contact has been made at the number. | 771 | 9.9% |
| 2.31 | Non-Response: subject deceased. | 4 | 0.1% |
| 2.22 | The maximum calling algorithm has been fulfilled for a telephone number and only answering machine contact was made. | 40 | 0.5% |
| 2.21 | Selected respondent not available in field period. | 6 | 0.1% |
| 4.50 | The telephone number called was not a residential number. Included are businesses, institutions, agencies, modems, public facilities, vacation homes, group quarters. | 7 | 0.1% |
| 2.32 | Non-Response: subject physically or mentally incapable of completing interview. | 34 | 0.4% |
| 2.20 | The telephone number was found to be not working when called. | 103 | 1.3% |
| 4.54 | Enumeration error – The respondent enumerated in the screener and selected for the extended interview is not a member of the household (typically occurs when visitors or family members living away are erroneously reported as household members.) | 55 | 0.7% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 2 | 0.0% |
| 2.112 | Refusal – The selected respondent or a gatekeeper refused the interview or the selected respondent broke off during the interview and refused to continue. | 1,129 | 14.4% |
| 2.112 | Refusal – Results from a call to the Minnesota Department of Health or the study's toll-free line. | 25 | 0.3% |
| | Total | 7,816 | 100.0% |

Table 4-4. MATS 2010 cell phone telephone number sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|--|---------------|----------------|
| 1.1 | Completed Screener | 1,861 | 10.0% |
| 4.70 | Cell phone not used by an adult over the age of 18. | 495 | 2.7% |
| 4.10 | The sampled cell phone number is not used by anyone living in Minnesota. | 446 | 2.4% |
| 4.70 | The sampled cell phone number is not used by a respondent who receives all or most of their calls on the cell phone. | 1,618 | 8.7% |
| 4.46 | The sampled number does not reach a cell phone. | 21 | 0.1% |
| 2.35 | All users of the sampled cell phone number are currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 3 | 0.0% |
| 4.70 | All users of the sampled cell phone number are currently attending school out of state and will not be back before the end of data collection. | 2 | 0.0% |
| 2.36 | Only eligible cell phone user already completed study on another phone number. | 3 | 0.0% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 2 | 0.0% |
| 2.331 | Language Problem: unable to communicate due to a hearing or speech problem or unable to reach an English speaking household member. | 140 | 0.8% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one "human" contact has been made at the number and there are no refusals or language problems in the call history for the household. | 536 | 2.9% |
| 3.13 | The calling algorithm has been fulfilled with no "human" or answering machine contact. | 6 | 0.0% |
| 3.14 | The maximum calling algorithm has been fulfilled for a cell phone number and only answering machine contact was made. | 2,775 | 14.9% |
| 4.50 | The cell phone number is not used for personal use. | 436 | 2.3% |
| 4.30 | The cell phone number was found to be not working when called. | 7,532 | 40.4% |
| 2.111 | Refusal –Screener respondent refused to be interviewed or broke off during the screener interview. | 2,563 | 13.8% |
| 2.111 | Refusal – Screener refusal results from a call to the Minnesota Department of Health or the study's toll-free line. | 190 | 1.0% |
| Total | | 18,629 | 100.0% |

Table 4-5. MATS 2010 cell phone extended interview sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|---|--------------|----------------|
| 1.1 | Completed interview. | 1,502 | 80.7% |
| 4.10 | Subject does not live in the state of Minnesota. | 2 | 0.1% |
| 2.35 | Subject is currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 1 | 0.1% |
| 4.10 | Subject is currently attending school out of state and will not be back before the end of data collection. | 3 | 0.2% |
| 2.332 | Language Problem: unable to communicate due to a hearing or speech problem or the selected respondent was unable to speak English. | 10 | 0.5% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one "human" contact has been made at the number. | 140 | 7.5% |
| 2.22 | The maximum calling algorithm has been fulfilled for a telephone number and only answering machine contact was made. | 27 | 1.5% |
| 4.50 | The cell phone number is not used for personal use. | 1 | 0.1% |
| 2.32 | Non-Response: subject physically or mentally incapable of completing interview. | 1 | 0.1% |
| 2.20 | The cell phone number was found to be not working when called. | 20 | 1.1% |
| 4.54 | Enumeration error – The respondent enumerated in the screener and selected for the extended interview is not a user of the cell phone number (typically occurs when friends or family members are erroneously reported as users of the cell phone.) | 6 | 0.3% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 2 | 0.1% |
| 2.112 | Refusal – The selected respondent or a gatekeeper refused the interview or the selected respondent broke off during the interview and refused to continue. | 138 | 7.4% |
| 2.112 | Refusal – Results from a call to the Minnesota Department of Health or the study's toll-free line. | 8 | 0.4% |
| Total | | 1,861 | 100.0% |

Table 4-6. MATS 2010 combined telephone number sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|--|--------------|----------------|
| 1.1 | Completed Screener | 9,677 | 12.0% |
| 4.70 | There is no one in the household age 18 or older to do the screener/Cell phone not used by an adult over the age of 18. | 525 | 0.7% |
| 4.10 | The sampled telephone number rings into a household not located in Minnesota/The sampled cell phone number is not used by anyone living in Minnesota. | 516 | 0.6% |
| 4.70 | The sampled cell phone number is not used by a respondent who receives all or most of their calls on the cell phone. | 1,618 | 2.0% |
| 4.46 | The sampled number does not reach a cell phone. | 21 | 0.0% |
| 2.35 | All household members/All users of the sampled cell phone number are currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 5 | 0.0% |
| 4.70 | All users of the sampled cell phone number are currently attending school out of state and will not be back before the end of data collection. | 2 | 0.0% |
| 2.36 | Sampled telephone number is part of MATS Cohort Study. | 15 | 0.0% |
| 2.36 | Only eligible household member/cell phone user already completed study on another phone number. | 6 | 0.0% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 7 | 0.0% |
| 2.331 | Language Problem: unable to communicate due to a hearing or speech problem or unable to reach an English speaking household member. | 472 | 0.6% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one "human" contact has been made at the number and there are no refusals or language problems in the call history for the household. | 1,406 | 1.8% |
| 3.13 | The calling algorithm has been fulfilled with no "human" or answering machine contact. | 3,474 | 4.3% |
| 4.50 | The telephone number was identified as non-residential during business purge preprocessing prior to CATI load. | 2,798 | 3.5% |
| 3.14 | The maximum calling algorithm has been fulfilled for a telephone/cell phone number and only answering machine contact was made. | 4,077 | 5.1% |

Table 4-6. MATS 2010 combined telephone number sample dispositions (continued)

| AAPOR code | Description | Count | Percent |
|-------------------|--|---------------|----------------|
| 4.50 | The telephone number called was not a residential number. Included are businesses, institutions, agencies, modems, public facilities, vacation homes, group quarters/The cell phone number is not used for personal use. | 3,833 | 4.8% |
| 4.30 | The telephone number was identified as non-working during Tri-tone match processing prior to start of calling. | 32,302 | 40.2% |
| 4.30 | The telephone/cell phone number was found to be not working when called. | 13,093 | 16.3% |
| 4.30 | The telephone number resulted in no ring back during purge processing prior to CATI load. | 871 | 1.1% |
| 2.111 | Refusal – Screener respondent refused to be interviewed or broke off during the screener interview. | 5,386 | 6.7% |
| 2.111 | Refusal – Screener refusal results from a call to the Minnesota Department of Health or the study’s toll-free line. | 227 | 0.3% |
| | Total | 80,331 | 100.0% |

Table 4-7. MATS 2010 combined extended interview sample dispositions

| AAPOR code | Description | Count | Percent |
|-------------------|--|--------------|----------------|
| 1.1 | Completed interview. | 7,057 | 72.9% |
| 4.10 | Subject does not live in the state of Minnesota. | 3 | 0.0% |
| 2.35 | Subject is currently living out of state and will not be back before the end of data collection ("Snowbirds"). | 2 | 0.0% |
| 4.10 | Subject is currently attending school out of state and will not be back before the end of data collection. | 15 | 0.2% |
| 2.332 | Language Problem: unable to communicate due to a hearing or speech problem or the selected respondent was unable to speak English. | 81 | 0.8% |
| 2.20 | The maximum calling algorithm has been fulfilled. At least one "human" contact has been made at the number. | 911 | 9.4% |
| 2.31 | Non-Response: subject deceased. | 4 | 0.0% |
| 2.22 | The maximum calling algorithm has been fulfilled for a telephone/cell phone number and only answering machine contact was made. | 67 | 0.7% |
| 2.21 | Selected respondent not available in field period. | 6 | 0.1% |
| 4.50 | The telephone number called was not a residential number. Included are businesses, institutions, agencies, modems, public facilities, vacation homes, group quarters/The cell phone is not used for personal use. | 8 | 0.1% |
| 2.32 | Non-Response: subject physically or mentally incapable of completing interview. | 35 | 0.4% |
| 2.20 | The telephone/cell phone number was found to be not working when called. | 123 | 1.3% |
| 4.54 | Enumeration error – The respondent enumerated in the screener and selected for the extended interview is not a member of the household (typically occurs when visitors or family members living away are erroneously reported as household members/users of the cell phone). | 61 | 0.6% |
| 4.10 | Other out of scope – The case is out of scope and no other final code applies. | 4 | 0.0% |
| 2.112 | Refusal – The selected respondent or a gatekeeper refused the interview or the selected respondent broke off during the interview and refused to continue. | 1,267 | 13.1% |
| 2.112 | Refusal – Results from a call to the Minnesota Department of Health or the study's toll-free line. | 33 | 0.3% |
| | Total | 9,677 | 100.0% |



4.3 MATS 2010 Response Rates

This section presents the survey response rates for MATS 2010. Since two independent samples were drawn for the cell phone and landline interviews, the response rates are reported separately for each sample. Section 4.3.1 describes the methodology, including the two phases at which non-response could be encountered, formulas for response rate calculation, and the rationale for focusing on the weighted response rates. Sections 4.3.2 and 4.3.3 report the response rates for the screener and the extended phase respectively, followed by the overall response rates across both phases.

4.3.1 Methodology

Two independent RDD samples were fielded on landline phones and cell phones. Landline phones are usually considered household devices. For MATS 2010, one adult was randomly sampled from each household for the extended interview once the screener had been completed. In contrast, researchers hold different opinions on whether a cell phone is a person-level or household-level device. Studies have shown that a non-negligible proportion of cell phone users share their phones with other household members, a common phenomenon even among cell phone-only and cell-mostly households. In MATS 2010, cell phones were treated as household devices. During the screener interview, the respondent was asked whether other household members shared the cell phone, and if the answer was yes, within-household sampling was conducted to select a respondent for the extended interview. In summary, household members were identified for interviews in a two-phase process for both RDD samples. Screener interviews were conducted to enumerate and sample household members, and then an extended questionnaire, the MATS 2010 instrument, was administered to the sampled person. Although the screener respondent was automatically selected for extended interview in single-person households in both samples, and in the cell sample when no cell phone sharing occurred, the logic of the two-phase interview still applies to these cases conceptually. The discussions will generally refer to the screening target as

“household” and the extended unit for the MATS 2010 interview as “person.” The response rates are provided at the following levels:

- i. Household-level response to the screening interview;
- ii. Person-level response to the extended interview, conditional on screener household response;
- iii. Overall response across the screener and extended phases, which is the product of (i) and (ii). The overall response rate indicates the percentage of possible interviews that were completed taking both survey phases into account.

For each phase, the response rate is generally defined as the ratio of the eligible responding units (i.e., households or persons) to the (estimated) number of units sampled and eligible for the interview in that phase. The MATS 2010 response rates are based on the AAPOR standards.⁷ All of the AAPOR response rate formulas collapse the numeric AAPOR disposition codes (Tables 4-2 through 4-5) and then assign them to the broad response categories in Table 4-8. The AAPOR formulas further collapse the latter into a few categories represented by the symbols that appear in the mathematical formula statements.

Table 4-8. AAPOR response rate formula categories

| AAPOR response category | AAPOR response category meaning | Screener response rate formula category |
|--------------------------------|--|--|
| I | Completed Interview | E |
| P | Partial Interview | N_e |
| R | Refusal and break-off | N_e |
| NC | Non-contact | N_e |
| O | Other | N_e |
| UH | Unknown if household occupied | N_u |
| UO | Unknown, other | N_u |

⁷ For reasons explained in section 4.3.2, MATS 2010 developed an allocation factor to account for the proportion of non-response screener cases that were outside the population covered by each the two sample frames. This factor, called f , is not an AAPOR standard, although it is comparable in purpose and application to the AAPOR e factor applied to screener non-response to account for non-residential phone numbers among the non-responding screener sample.

AAPOR has two similar response rate formulas that are relevant to MATS 2010, RR3 and RR4. The only difference between AAPOR RR3 and RR4 is that RR3 excludes partial completes from the numerator and RR4 includes them, resulting in a slightly higher response rate. RR4 can be used only if partial completes are weighted and included in the final data file. MATS 2010 did not include partial completes in the analysis file and did not assign a final sample weight to them. To be included in the weighted file used for analysis, an interview had to have reached the last question, J11, J11a, or J11b, as applicable based on the skip pattern. Those that broke off before this point are accounted for in Tables 4-3 and 4-5 among the breakoffs.

The screener response rates for both samples are calculated using the following formula:

$$R_{\text{screener}} = \frac{E}{E + f(N_e + eN_u)}$$

where

- E = number of responding households
- N_e = number of nonresponding households (known to be residential ; unknown about whether the households belonged to the MATS 2010 target population)
- N_u = number of cases with unknown residential status (due to non-contact)
- e = estimated residential rate among nonresponding cases with unknown residential status
- f = estimated proportion of the nonresponding residential cases that belonged to the MATS 2010 target population

For MATS 2010, since there is no auxiliary information about the residential status of the non-contact cases or the proportion of the nonresponding residential cases that belonged to the target population, it is necessary to estimate the factors e and f in the screener response rate formula using the residential and eligibility rates among those whose status could be determined through the screener. Details will be discussed in Section 4.3.2.

The extended interview was administered only among the eligible adults sampled from the screener responding households. At the extended stage, there is no

sampling unit with unknown residential and/or eligibility status. The extended response rate is calculated at the person level using the simple formula:

$$R_{\text{extend}} = \frac{E}{E + N_e}$$

where

- E = number of respondents to the extended interview
 N_e = number of nonrespondents to the extended interview

Response rates can be either unweighted or weighted. The unweighted response rate, computed using the raw number of cases, provides a useful description of the success of the operational aspects of the survey. The weighted rate, computed by summing the weights for all the cases in both the numerator and denominator, gives a better description of the success of the survey with respect to the sampled population. For MATS 2010, the unweighted and weighted response rates are essentially the same at the screener stage because an equal probability selection method was employed for selecting both samples. At the extended stage, the unweighted and weighted rates may differ moderately due to different weight adjustment factors associated with within-household selection. Sections 4.3.2 and 4.3.3 report the weighted response rates, since weighted response rates allow direct comparisons between different surveys with the same target population regardless of the sample design employed.

4.3.2 Screener Response Rates

Table 4-9 shows the counts of the phone numbers fielded in MATS 2010 by response rate formula category. The four major types of residential status are 1) those identified as residential households that belonged to the target population of MATS 2010 (E and N_e), 2) those identified as residential households but not belonging to the MATS 2010 target population (I_{NT}), 3) those identified as nonresidential (primarily nonworking and business) phone numbers (I_{NR}), and 4) those phone numbers that, despite numerous attempts, could not be classified as either residential or nonresidential (N_u). The inclusion of the third type is driven mainly by the variety and large number of cases in the cell phone sample that were screened but not included in the interviewed sample (out of state, used exclusively



by minors, not cell phone-only or cell phone-mostly,), but some of these situations also occur in the landline sample, although far less often. Calculation of response rate is complex because of the possible ways to estimate the residential rate among the phone numbers whose residential status was unknown as well as the proportion of non-responding households that belonged to the target population. In the landline RDD survey for MATS 2007, the residential rate was estimated using subfactor e 's of 0.27 and 0.63 for "No Answers" and "Answering Machines," respectively, which produced a blended e of 0.443. For MATS 2010, due to the changing telephone industry and our incomplete knowledge of cell phone usage pattern, e is estimated using the Council of American Survey Research Organizations (CASRO) approach. The CASRO rate is computed by allocating the numbers with unknown residential status in the same proportion observed in the numbers with known residential status, which is considered a conservative approach.

$$e = \frac{E + N_e + I_{NT}}{E + N_e + I_{NT} + I_{NR}}$$

where

- E = number of responding households
- N_e = number of nonresponding households (known to be residential; unknown about whether the households belonged to the MATS 2010 target population)
- I_{NT} = number of residential households determined (through the screening interview) to be out of scope of the MATS 2010 target population
- I_{NR} = number of sampled phone numbers determined to be nonresidential

Table 4-9. Unweighted counts of phone numbers by screener response rate formula category

| Screener response rate formula category | Landline sample | Cell phone sample |
|--|------------------------|--------------------------|
| Residential respondents (E) | 7,816 | 1,861 |
| Nonrespondents known to be residential; unknown whether the households belonged to the MATS 2010 target population (N_e) | 4,082 | 3,435 |
| Residential households determined (through the screening interview) to be out of scope of the MATS 2010 target population (I_{NT}) | 105 | 2,563 |
| Sampled phone numbers determined to be nonresidential (nonworking or business) (I_{NR}) | 44,929 | 7,989 |
| Nonrespondents with unknown residential status due to non-contact (N_u) | 4,770 | 2,781 |
| Ring no answer | 3,468 | 6 |
| Answer machine | 1,302 | 2,775 |

Since “reverse-match” was conducted for all the sampled landline phone numbers in order to obtain the addresses for mailing advanced letters, it is possible take advantage of this information by estimating e separately for the landline phone numbers with and without matched addresses. Table 4-10 shows that, among the cases whose residential status was determined, the residential rates among the cases with matched addresses are much higher than those without matched addresses, which is consistent with our expectation. The “reverse-match” operation is not possible for the cell phone sample, so a single residential rate was computed for the “ring no answer” group and the “answer machine” group respectively. The weighted average of e is 0.95 for the cell phone sample, and 0.42 for the landline sample (similar to the MATS 2007 e estimate).

Table 4-10. Estimated residential rates (*e*) among different types of non-contact cases

| Screener disposition | Landline sample | | Cell phone sample |
|----------------------|------------------------------|---------------------------------|-------------------|
| | With matched mailing address | Without matched mailing address | |
| Ring no answer | 0.80 | 0.05 | 0.50 |
| Answer machine | 0.90 | 0.36 | 0.95 |

In the absence of any additional information for calculating the factor *f*, the proportion of nonresponding residential households that belonged to the MATS 2010 target population, this rate was estimated using the information collected during the MATS screening interview. The factor *f* is calculated as below. The estimated *f* is 98.7 percent for the landline sample and 41.9 percent for the cell phone sample. The low rate for the cell phone sample is due to the large proportion of residential cell phone numbers that were not the MATS target population (e.g. non-adult cell phone users, not cell phone-only or cell phone-mostly).

$$f = \frac{E}{E + I_{NT}}$$

where

- E* = number of responding households
- I_{NT}* = number of residential households determined (through the screening interview) to be out of scope of the MATS 2010 target population

The screener weighted response rates are calculated using screener result codes, the estimated *e* and *f*, and household base weights. The results are shown in Table 4-12, with 66.0 percent for the landline sample and 56.4 percent for the cell phone sample.

4.3.3 Extended Response Rates and Overall Response Rates

One adult was selected from each eligible, screened household for the extended interview. Table 4-11 gives the final status of all the adults sampled for the extended interview. A few cases were subsequently determined to be ineligible when contacted for the extended interview, because the person was found, for example, to live outside Minnesota or to be under age 18. This type of occasional screener response error occurs in every RDD survey. These persons (a total of 71 landline

cases and 14 cell phone cases) are excluded from both the numerator and denominator of the response rate formula. A person-level base weight is applicable at the extended phase, which is the product of the household-level base weight and the number of eligible adults sharing the phone number in the household. The weighted extended response rates are 68.1 percent for the landline sample and 78.9 percent for the cell phone sample, as shown in Table 4-12.

Table 4-11. Unweighted counts of sampled persons by extended response rate formula category

| Extended response rate formula category | Landline sample | Cell phone sample |
|--|------------------------|--------------------------|
| Respondents (E) | 5,555 | 1,502 |
| Nonrespondents (N_e) | 2,190 | 345 |
| Ineligible | 71 | 14 |

Table 4-12 indicates that it is easier to obtain response on the landline than on the cell phone at the screener phase. However, once the screener has been completed, a cell phone case is more likely to respond to the extended interview than a landline case. This is probably because the majority of cell phones are personal devices (i.e. not shared by other household members), so the screener respondent himself/herself is very likely to be selected for the extended interview. In contrast, within-household sampling is applicable to the majority of landline cases, and it is more difficult to gain cooperation when a different person other than the one who has responded to the screener is sampled for the extended interview.

The last row in Table 4-12 gives the overall weighted response rate for each sample, which is the product of the screener and extended rates. The two samples yield very close overall weighed response rates : 45.0 percent for the landline and 44.5 percent for the cell phone.

Table 4-12. Weighted response rates for landline and cell phone samples

| Weighted response rate | Landline sample | Cell phone sample |
|-------------------------------|------------------------|--------------------------|
| Screener | 66.0% | 56.4% |
| Extended | 68.1% | 78.9% |
| Overall | 45.0% | 44.5% |

4.4 Selected Operational Statistics

This section presents some statistics that characterize various operational aspects of MATS 2010.

4.4.1 Principal Sources of Non-response

4.4.1.1 Principal Sources of Non-response in Landline Sample

Table 4-13 summarizes the results for all landline cases, after eliminating the known non-working and non-residential numbers. Consistent with the AAPOR RR3 formula, this table collapses the screening and extended interview processes into a single set of results. For example, a screener refusal in one case and a completed screener that resulted in a refusal of the extended interview in another case are treated identically and count as two refusals in this table. The largest source of landline sample non-response was non-contact to the screener or extended interview: 59.4 percent out of the total 11,036 possible phone numbers that could have yielded a completed interview were not able to be contacted at one of the two stages. Refusal was the second-largest source of landline sample non-response. 4,014 cases (36.4 percent) resulted in a refusal at either the screener or extended stage. Combined, these two outcomes accounted for 95.8 percent of the total non-response. These results are summarized in Table 4-13.

Table 4-13. Primary sources of non-response in landline sample, collapsed across screening and extended interview stages

| | Count | Percent of non-respondents |
|-----------------------|--------------|-----------------------------------|
| Total non-respondents | 11,036 | 100.0% |
| Refused | 4,014 | 36.4% |
| No contact | 6,560 | 59.4% |
| Both | 10,574 | 95.8% |

There is one point to keep in mind in regard to the numbers cited in the previous paragraph. First, as explained in Section 4.3, some undetermined proportion of the non-contact cases at the screener level were not really households, and the response rate formula discounted a proportion of them as non-residential. The statistics presented in this section are purely operational and count all non-contact cases as non-response. It is also informative, and more straightforward, to look at sources of non-response for the landline extended interview, that is, among the 7,816 adults selected for the interview from the completed screeners. Table 4-14 shows there were 7,739 eligible cases among the 7,816 selected. Most of those who were not eligible were individuals who, upon being contacted for the interview, were found to have been erroneously included as members of the household during the screener (e.g., guests, family members not currently residing in the household). There were 2,184 non-respondents, of whom the largest number were 1,154 refusals (52.8 percent of non-respondents and 14.9 percent of all eligible sample). Most of the remainder were individuals who could not be reached despite repeated attempts to do so, including the extra call attempts made beyond the BRFSS protocol parameters. These were 811 cases, or 37.1 percent of non-response and 10.5 percent of all eligible sample. Combined, refusals and maximum contacts accounted for 1,965 non-respondents, or 90.0 percent of all non response and 25.4 percent of all eligible adult sample.

Table 4-14. Primary sources of non-response in landline extended interview sample

| | Count | Percent of non-respondents (n = 2,184) | Percent of total eligible sampled (n = 7,739) |
|--------------------------|--------------|---|--|
| Total sampled | 7,816 | - | - |
| Ineligible/out of scope | 77 | - | - |
| Total eligible sampled | 7,739 | - | 100.0% |
| Total complete | 5,555 | - | 71.8% |
| Total non-respondents | 2,184 | 100.0% | 28.2% |
| Refused | 1,154 | 52.8% | 14.9% |
| Maximum contact attempts | 811 | 37.1% | 10.5% |
| Both | 1,965 | 90.0% | 25.4% |

4.4.1.2 Principal Sources of Non-response in Cell Phone Sample

Table 4-15 summarizes the results for all cell phone cases, after eliminating the known non-working numbers and numbers not used for personal use. As in the landline non-response table above (Table 4-13), this table collapses the screening and extended interview processes into a single set of results. The largest source of cell phone non-response was non-contact to the screener or extended interview: 53.4 percent out of the total 6,561 possible cell phone numbers that could have yielded a completed interview were not able to be contacted at one of the two stages. Refusal was the second-largest source of cell phone non-response: 2,899 cases (44.2 percent) resulted in a refusal at either the screener or extended stage. Combined, these two outcomes accounted for 97.6 percent of the total non-response. These results are summarized in Table 4-15.

Table 4-15. Primary sources of non-response in cell phone sample, collapsed across screening and extended interview stages

| | Count | Percent of non-respondents |
|-----------------------|--------------|-----------------------------------|
| Total non-respondents | 6,561 | 100.0% |
| Refused | 2,899 | 44.2% |
| No contact | 3,504 | 53.4% |
| Both | 6,403 | 97.6% |

As in the landline non-response discussion, the statistics presented in this section for the cell phone sample are purely operational and count all non-contact cases as non-response. Again, it is interesting to look at sources of non-response for the cell phone extended interview, that is, among the 1,861 adults selected for the interview from the completed screeners. Table 4-16 shows there were 1,847 eligible cases among the 1,861 selected. Most of those not eligible were individuals who, upon being contacted for the interview, were found to have been erroneously included as users of the cell phone number (e.g., friends, family members who do not receive calls on the cell phone number). There were 345 non-respondents, of whom the largest number were 146 refusals (42.3 percent of non-respondents and 7.9 percent of all eligible sample). Closely following the refusals were individuals who could not be reached despite repeated attempts to do so, including the extra call attempts made beyond the BRFSS protocol parameters. These were 167 cases, or 48.4 percent of non-response and 9.0 percent of all eligible sample. Combined, refusals and maximum contacts accounted for 313 non-respondents, or 90.7 percent of all non-response and 16.9 percent of all eligible adult sample.

Table 4-16. Primary sources of non-response in the cell phone extended interview sample

| | Count | Percent of non-respondents (n = 345) | Percent of total eligible sampled (n = 1,847) |
|--------------------------|--------------|---|--|
| Total sampled | 1,861 | - | - |
| Ineligible/out of scope | 14 | - | - |
| Total eligible sampled | 1,847 | - | 100.0% |
| Total complete | 1,502 | - | 81.32% |
| Total non-respondents | 345 | 100.0% | 18.7% |
| Refused | 146 | 42.3% | 7.9% |
| Maximum contact attempts | 167 | 48.4% | 9.0% |
| Both | 313 | 90.7% | 16.9% |

4.4.2 Refusal Conversion Results

Recontacting people who initially refuse to participate in an interview is designed to increase the sample size and response rates, and also to reduce bias associated with including in the sample only those who are most inclined to respond. This section describes the results of the refusal conversion efforts for the landline and cell phone screeners, and the landline and cell phone extended interviews, as summarized in Table 4-17 and Table 4-18.

4.4.2.1 Screener Refusal Conversion Results

Table 4-17. MATS 2010 screener refusal conversion

| | Landline screener | Cell phone screener | Combined screener |
|-----------------------------------|--------------------------|----------------------------|--------------------------|
| Assigned to data collection | 25,731 | 18,629 | 44,360 |
| Initially refused (#) | 6,605 | 4,788 | 11,393 |
| Converted (#) | 1,016 | 363 | 1,379 |
| Converted (%) | 15.4% | 7.6% | 12.1% |
| Total completed | 7,816 | 1,861 | 9,677 |
| Converted as percent of completed | 13.0% | 19.5% | 14.3% |

Landline Screener Conversion. At the screener stage, the initial telephone contact resulted in a refusal to respond to the landline screener questions at 6,605 landline phone numbers. After conversion attempts with the initial landline refusals, 1,016 of these cases became completed screener interviews, representing a conversion rate of 15.4 percent and constituting 13.0 percent of the total 7,816 completed screeners.

Cell Phone Screener Conversion. The initial cell phone contact resulted in a refusal to respond to the cell phone screener questions at 4,788 cell phone numbers. After conversion attempts with the initial cell phone refusals, 363 of these cases became completed screener interviews, representing a conversion rate of 7.6 percent and constituting 19.5 percent of the total 1,861 completed cell phone screeners. As compared to the landline screener, it was more difficult to convert a cell phone screener refusal. While there are many possible explanations for the lower conversion rate in the cell phone sample, one explanation could be due to the fact that cell phones are often used by one individual, while landline phones are often

used by multiple members of a household. When calling back to convert a landline refusal, there is a greater chance of reaching a different (and willing) respondent than when calling cell phone refusals.

Combined Screener Conversion. Overall at the screener stage, the initial contact resulted in a refusal to respond to the screener questions at 11,393 phone numbers. After conversion attempts with the initial refusals, 1,380 of these cases became completed screener interviews, representing a conversion rate of 12.1 percent and constituting 14.3 percent of the total 11,393 completed screeners.

4.4.2.2 Selected Household Member Conversion

Table 4-18. MATS 2010 extended interview refusal conversion

| | Landline extended | Cell phone extended | Combined extended |
|-----------------------------------|--------------------------|----------------------------|--------------------------|
| Assigned to data collection | 7,816 | 1,861 | 9,677 |
| Initially refused (#) | 1,940 | 266 | 2,206 |
| Converted (#) | 308 | 50 | 358 |
| Converted (%) | 15.9% | 18.8% | 16.2% |
| Total completed | 5,555 | 1,502 | 7,057 |
| Converted as percent of completed | 5.5% | 3.3% | 5.1% |

Conversion of Selected Household Members in Landline Sample. Among the household members selected for the MATS 2010 interview in the landline sample, 1,940 initially refused to respond to the interview. After conversion attempts, 308 of these completed the interview. This is a conversion rate of 15.9 percent, representing 5.5 percent of the total 5,555 completed landline interviews.

Conversion of Selected Cell Phone Users in Cell Phone Sample. Among the cell phone users selected for the MATS 2010 interview in the cell phone sample, 266 initially refused to respond to the interview. After conversion attempts, 50 of these completed the interview. This is a conversion rate of 18.8 percent, representing 3.3 percent of the total 5,555 completed landline interviews. While it was easier to convert a landline *screener* refusal than a cell phone screener refusal, there is not a large difference in the landline and cell phone extended conversion rates. Using the same logic discussed above, the landline respondent has already been selected at

the extended level, so the conversion attempt must be made with the same respondent. The landline extended interview conversion no longer has the possibility of reaching a different household member who may be more willing to complete the interview.

Conversion of Selected Respondents in Combined Sample. Among the respondents selected across both samples, 2,206 initially refused to respond to the extended interview. After conversion attempts, 358 of these completed the interview. This is a conversion rate of 16.2 percent representing 5.1 percent of the total 9,677 completed interviews.

4.5 Interview Timing

Tables 4-19 and 4-20 present statistics about the length of time it took to administer the MATS 2010 interview. Table 4-19 shows the mean, minimum, and maximum lengths for the cell phone and landline screener; Table 4-20 shows the mean, minimum, and maximum lengths of extended interview for the overall sample and for different smoking statuses.

Table 4-19. Screener lengths (minutes)

| Sample Group | Mean | Minimum | Maximum |
|--------------|------|---------|---------|
| Cell phone | 2.4 | 1.4 | 6.8 |
| Landline | 1.9 | 0.4 | 11.2 |

Table 4-20. Extended interview lengths (minutes)

| Sample Group | Mean | Minimum | Maximum |
|----------------------|------|---------|---------|
| Current smokers | 20.8 | 12.5 | 73.4 |
| Former smokers | 15.9 | 9.7 | 57.5 |
| Experimenter smokers | 13.6 | 8.3 | 33.1 |
| Never smokers | 12.8 | 7.7 | 42.8 |
| All respondents | 15.0 | 7.7 | 73.4 |



Overall, the average extended interview lasted 15 minutes. The design target for the MATS 2010 extended questionnaire was an average interview length of 15 minutes across all respondents. As seen in the table, the actual experience was exactly 15 minutes on average. Also as expected, interviews with current and former smokers took longer to complete than did those with experimental and never smokers, with the current smokers requiring the most time on average, owing to the largest number of questions applying to their circumstance. With the exception of current smokers, no interview exceeded an hour in length, and this occurrence was rare indicated by the much lower mean.

As expected, the cell phone screener took longer to administer on average than the landline screener due to the higher number of questions necessary to determine a cell phone number's eligibility. There was no explicit target time for the screener interview.





5. Weighting

The objective of the MATS 2010 survey is to make inferences about the entire civilian, non-institutionalized adult population in Minnesota. Sample weights are created so that population estimates can be calculated using the results of a survey from a sample of a finite population. Weighting is necessary to account for differential probabilities of selection and to reduce potential bias due to non-response and differential coverage of subpopulations. Although weighting adjustments are aimed at reducing bias, these adjustments typically introduce variation in the weights, which increases the variances of survey estimates. Care was taken in the development and implementation of the weighting methodology to balance the bias reductions against the potential increases in variance.

Two Random Digit Dialing (RDD) samples were selected and fielded for MATS 2010, one on landline phones, the other on cell phones. The weighting process consists of the following stages:

1. Create household-level base weights that reflect the different probabilities of selection for the sampled phone numbers from landline and cell phone frames. Then adjust these weights for screener non-response and multiple phone numbers in a household.
2. Calculate person-level weights to account for within-household sampling, and then adjust for non-response to extended interviews.
3. Conduct a composite weighting adjustment on the overlap component between the two samples – the cell phone-mostly group – as described in Chapter 2.
4. Calibrate the composite weights to the population totals estimated from reliable external sources to further reduce remaining non-response and non-coverage errors. A trimming and re-calibration step was also incorporated into this stage.

Parallel weighting adjustments were conducted for the landline and cell phone samples separately in stages 1 and 2 before the two samples were put together for



stages 3 and 4. Only one set of weights was created for the combined dataset including the respondents from both samples, which can be used to produce estimates for the entire adult population in Minnesota as well as its subgroups. Sections 5.1 through 5.4 cover each of the four weighting stages. Some of the stages involve multiple steps, which will be discussed in greater detail below.

5.1 Household-Level Weights

As discussed in Chapter 2, both cell phone and landline phones were treated as household devices in MATS 2010. The primary purpose of the screening interview was to provide information required to assess the eligibility of household members for the MATS 2010 interview. Household weights were calculated solely for use as a basis for computing person-level weights for the analysis of interview data.

The household-level weight for household j in sample s , $HHW_{j(s)}$, is the product of three factors:

- Base weight associated with the differential sampling of telephone numbers in the two samples ($A_{j(s)}$);
- Adjustment for screener non-response ($B_{j(s)}$); and
- Adjustment for the number of telephone numbers in a household ($C_{j(s)}$).

The procedure for computing the household-level weights for each sample is described in Sections 5.1.1 through 5.1.3.

5.1.1 Household Base Weights

Each sampled telephone number was assigned an initial base weight, and this was done separately for the two RDD samples. This initial base weight was computed as the inverse of the probability of selection of the telephone number from the sampling frame. Since an equal probability selection method was used in each frame, the base weights ($A_{j(s)}$) are 54.54 for all the 61,702 landline phone numbers sampled from the landline frame. As mentioned in Section 2.2, the cell phone sample included two groups of cases – 18,451 selected from the cell phone frame with the base weights of 99.04 and 178 selected from the landline frame with the

base weights of 54.54. The latter group (178 cases) was determined to be cell phone numbers through the process of purging non-working and non-residential telephone numbers from the landline sample after sample selection; these cases were assigned the same base weights as the landline phone numbers because base weights should reflect the actual probability of sampling the phone number.

5.1.2 Adjustment for Screener Non-response

This step adjusts for households that did not respond to the screener. Each sampled telephone number was classified as a respondent (*R*), a nonrespondent (*NR*), or an ineligible case (*I*). The base weights of the nonrespondent cases were distributed to the weights of the respondent cases. Separate adjustments were made for the two samples because the response propensities for cell phone and landline screeners were different. No additional information was available at the screener phase to form more refined non-response adjustment cells within each sample type. The non-response adjustment factor $B_{j(s)}$, applied to each responding household j in sample type s is

$$B_{j(s)} = \frac{\sum_{j(s) \in R} A_{j(s)} + \sum_{j(s) \in NR} A_{j(s)}}{\sum_{j(s) \in R} A_{j(s)}}$$

5.1.3 Adjustment for Multiple Telephone Numbers in Household

At the end of the landline screener interview, information about the existence of additional telephone numbers and their use in the household was collected. If the household had an additional telephone number for residential use (not solely for business, fax or computer use), the selection probability associated with the household was higher and the weight had to be adjusted to account for this. For the cell phone sample, if there were multiple persons in the household, each with a cell phone that was not shared by other household members, then the multiple phone number adjustment factor should be cancelled out by the weighting factor for within-household selection (to be discussed in Section 5.2.1). To keep the implementation simple, a factor of 1 was applied for all the cell phone sample cases in this step.

The adjustment factor for multiple phone numbers is:

$$\begin{aligned}
 C_{j(s)} &= 1 \quad \text{if } s \text{ indicates cell phone sample} \\
 &= 1 \quad \text{if } s \text{ indicates landline sample and household } j \text{ has no more than one} \\
 &\quad \text{telephone number} \\
 &= 0.5 \quad \text{if } s \text{ indicates landline sample and household } j \text{ has more than one telephone} \\
 &\quad \text{number}
 \end{aligned}$$

In this adjustment, it is standard practice to assume that there is at most one additional residential use telephone number in the household. Research has shown that there are too few households with more than two such numbers to significantly affect the weight distribution even if the inverse of the exact number of phone lines is used in the formula.

Thus, the household-level weight for household j in sample s , $HHW_{j(s)}$, is given by:

$$HHW_{j(s)} = A_{j(s)} \cdot B_{j(s)} \cdot C_{j(s)}$$

Note that the original sample design included protocols for a subsampling for refusal conversion within the landline screener process. However, as discussed in Chapter 2, it was determined during data collection monitoring that the most efficient method for achieving the required numbers of completed extended interviews was to allow all of the landline screener refusals into the conversion process, not just the original 60 percent subsampled. By doing so, there was no longer any need for computing and applying a household adjustment factor for refusal conversion subsampling (as was done for MATS 2007), because it would equal 1 for all cases.

5.2 Person-Level Weights

Household-level weights were used as the starting point for creating person-level weights. The person-level weight for sampled person k in household j , sample s , $PW_{jk(s)}$, is the product of the household-level weight and two weighting adjustment factors:

- Adjustment factor associated with within-household selection ($D_{jk(s)}$);
- Adjustment for the MATS 2010 extended interview unit non-response ($E_{jk(sc)}$).

The procedure for computing the person-level weights for each sample is described in Sections 5.2.1 and 5.2.2.

5.2.1 Adjustment for Within-Household Selection

For both samples, the extended interview was administered to only one person per household. The within-household sampling adjustment factor for person k from household j in sample s is:

$$D_{jk(s)} = N_{j(s)}$$

where $N_{j(s)}$ is the number of eligible adults in household j , sample s , that shared the sampled telephone number.

For the landline sample as well as the cell phone cases where the phone was shared, the adjustment factor is equal to the number of eligible adults in the household (standard landline RDD protocol simply assumes that all adults in a household “share” the sampled phone number). For the cell phone sample, if the screener respondent reported that the phone was not shared by any other adult household members, then he/she was invited to continue with the extended interview directly, and thus received an adjustment factor of 1 for within-household selection. The MATS 2010 statisticians decided to set the maximum value for this adjustment factor to 3 for the shared cell phone cases because the proportion of households with more than three adults sharing a single cell phone is very small. For the landline sample, no upper limit was set for this adjustment factor.

For each sampled person k from household j in sample s , the person-level weight before extended non-response adjustment, $UPW_{jk(s)}$, can be calculated as the product of the household-level weight and the adjustment factor for within-household sampling:

$$UPW_{jk(s)} = HHW_{j(s)} \cdot D_{jk(s)}$$

5.2.2 Adjustment for Extended Interview Non-response

The adjustment for extended interview non-response is very similar to what was done for the screener phase, except that it is possible to use the variables collected during the screening interview to form non-response adjustment cells. The non-



response adjustment factor, $E_{jk(sc)}$, applied to each respondent k from household j in adjustment cell c of sample s is

$$E_{jk(sc)} = \frac{\sum_{jk(s) \in R_c} UPW_{jk(s)} + \sum_{jk \in NR_c} UPW_{jk(s)}}{\sum_{jk(s) \in R_c} UPW_{jk(s)}}$$

For the landline sample, household size (i.e. number of eligible adults in the household) was used for forming the non-response adjustment cells. For the cell phone sample, since phone usage questions were asked during the screening interview, the non-response adjustment cells were formed by crossing three pieces of information: 1) whether the sampled person was in a cell phone-only or cell phone-mostly household; 2) household size; and 3) whether the cell phone was for personal use only or for both personal and business use. Any missing values for the above variables were imputed in order to categorize each person into a non-response adjustment cell. Random allocation was used to impute the data because the percent of missingness was very small. The algorithm ensures that the distributions of the imputed values are the same as the distributions of the respondent-reported non-missing values.

The person-level weight for person k from household j in sample s , $PW_{jk(s)}$, is calculated as:

$$PW_{jk(s)} = UPW_{jk(s)} \cdot E_{jk(sc)}$$

5.3 Composite Weights

Although separate person-level weights were created for landline and cell phone cases, as described in Section 5.2, it is inappropriate to generate population estimates using either sample by itself. Each sample covers only a subset of the Minnesota adult population and part of the subsets overlap. Specifically, the landline sample is missing the cell-only households and the cell phone sample includes the cell-only and cell-mostly households; the cell sample and the landline sample both include the cell-mostly households. Given this, the next step was to combine the two samples into one dataset and develop a single set of weights

(referred to as composite weights). The composite weight, $PW_{jk(comp)}$, for person k in household j , is calculated as:

$$\begin{aligned}
 PW_{jk(comp)} &= \lambda \cdot PW_{jk(cellphone)} && \text{for cell phone-mostly cases in cell phone sample} \\
 &= (1 - \lambda) \cdot PW_{jk(landline)} && \text{for cell phone-mostly cases in landline sample} \\
 &= PW_{jk(cellphone)} && \text{for other cases in cell phone sample} \\
 &= PW_{jk(landline)} && \text{for other cases in landline sample}
 \end{aligned}$$

where λ is the compositing factor for the overlapping cell phone-mostly cases.

Careful considerations were given to the determination of λ associated with the cell phone-mostly cases covered by both samples to balance the trade-off between bias and variance. For MATS 2010, λ was calculated to be 0.574 to minimize the mean squared error for the estimated percentage of cell phone-mostly persons among the Minnesota adult population.

5.4 Calibration to External Population Totals

The last stage of weighting was to calibrate the composite person-level weights to Minnesota adult population estimates. Two sources were used to obtain the population estimates: 1) demographic information from the 2008 American Community Survey; and 2) distribution of phone status from the most recent National Health Interview Survey. Compared to the MATS 2007, phone status was included as an additional dimension for calibration because tobacco use behavior was believed to be associated with landline/cell phone ownership and usage (e.g., cell-only or cell-mostly) pattern. Calibration to population control totals is a commonly used estimation procedure to reduce potential bias and variance. The calibration method used on the MATS 2010 was raking, an iterative procedure where the weights are benchmarked to multiple sets (or dimensions) of marginal control totals in a sequential order until the sums of the raked weights equal the

marginal control totals for all the dimensions or the differences are within a specified tolerance level. The raked weight, RPW_{jk} , can be expressed as

$$RPW_{jk} = PW_{jk(comp)} \cdot \prod_{d=1}^D RF_{dl}$$

where RF_{dl} is the raking factor for dimension d , level l which respondent jk is in. For example, if the 4th dimension ($d=4$) is sex with two levels ($l=1$ for male and $l=2$ for female), then the raking factor for this dimension is RF_{41} for the adult male. The raking factors are derived so the following relationship holds for every raking dimension d , level l :

$$CNT_{dl} = \sum_{jk} \delta_{dljk} \cdot RPW_{jk}$$

where CNT_{dl} is the control total, and $\delta_{dljk} = 1$ if adult k in household j is in level l of dimension d and $\delta_{dljk} = 0$, otherwise. The MATS 2010 weights were raked to the five dimensions defined in Table 5-1. Raking to these five dimensions simultaneously controlled for phone status, gender, age, race/ethnicity, educational attainment, and location (inside or outside of the Twin Cities metropolitan area). A very small proportion of these variables had missing values and were imputed using the same method as described above in Section 5.2.2.

Due to the accumulative effect of all the weighting steps described above, very large weights resulted for a small number of cases, which would drive up the variance of the estimates. These weights were trimmed to three times the average weight of all the responding cases. The portions of the trimmed weights above this average were redistributed to other cases by sample type (i.e. cell phone and landline), and the calibration operation described above was then reapplied to the trimmed/redistributed weights. Weight trimming is commonly used in regular survey practice. The potential small bias introduced through trimming is outweighed by the variance reduction, such that a reduction in the overall mean squared error was expected.

Table 5-1. Description of variables used to define MATS 2010 RDD raking cells

| Raking dimensions | Raking cell definitions |
|--------------------------|--|
| Dimension 1 | |
| Telephone status/usage | Having cell phone only Having cell phone mostly Having landline and cell phone, but not cell phone mostly Having landline only |
| Dimension 2 | |
| Gender x Age groups | Male, 18 to 24 years Male, 25 to 29 years Male, 30 to 34 years Male, 35 to 44 years Male, 45 to 54 years Male, 55 to 64 years Male, 65 years and older Female, 18 to 24 years Female, 25 to 29 years Female, 30 to 34 years Female, 35 to 44 years Female, 45 to 54 years Female, 55 to 64 years Female, 65 years and older |
| Dimension 3 | |
| Race/Ethnicity | Hispanic Non-Hispanic, White Non-Hispanic, African American Non-Hispanic, Asian Non-Hispanic, Native American, Pacific Islander Non-Hispanic, Other race, 2 or more races |

Table 5-1. Description of variables used to define MATS 2010 RDD raking cells (continued)

| Raking dimensions | Raking cell definitions |
|------------------------------|--|
| Dimension 4 | |
| Location x Collapsed Race | Inside Twin-Cities, Hispanic |
| | Inside Twin-Cities, Non-Hispanic, White |
| | Inside Twin-Cities, Non-Hispanic, African American |
| | Inside Twin-Cities, Non-Hispanic, Asian |
| | Inside Twin-Cities, Non-Hispanic, Native American, Pacific Islander, Other race, 2 or more races |
| | Outside Twin-Cities, Hispanic |
| | Outside Twin-Cities, Non-Hispanic, White |
| | Outside Twin-Cities, Non-Hispanic, African American, Asian, Native American, Pacific Islander, Other race, 2 or more races |
| Dimension 5 | |
| Educational attainment x Age | Less than HS degree by the age groups in dimension 2 |
| | HS degree or equivalent, by the age groups in dimension 2 |
| | At least some college, by the age groups in dimension 2 |
| | BS degree or above, by the age groups in dimension 2 |

6. Comparability with Previous MATS Surveys and Limitations of the Data

It is helpful for users of the MATS 2010 data to be aware of the survey methods described in this methodology report when analyzing the data, interpreting the findings, reading reports, and applying the results to historical assessment and planning for the future. It is likewise helpful for them to be informed of any known or potential limitations that apply to the use of the data. Finally, when comparing the results of MATS 2010 to those of previous MATS surveys, it is important to consider methodological factors that may affect the comparability of the data from one round to the next. This report focuses mainly on comparability of MATS 2010 with the immediately previous round, MATS 2007, but also examines comparability across all four rounds when relevant and feasible.

6.1 Comparability Issues

Several issues related to comparability, or factors affecting it, have been discussed in the relevant sections of this report. Briefly, these are aspects of the following issues:

Sampling issues:

- Addition of cell phone sample and elimination of Blue Cross sample
- Elimination of oversampling of young adults and African Americans

Weighting issues:

- Weighting design
- Combining and weighting the landline and cell phone samples, versus combining and weighting the RDD-Blue Cross sample



Data collection issues:

- Screener refusal conversion

Questionnaire design and specification issues:

- Questionnaire content
- Questionnaire skip patterns

6.1.1 Sampling

6.1.1.1 Inclusion of Cell Phone Sample and Elimination of Blue Cross Sample

The inclusion of the cell phone sample in MATS 2010 was to address particular concerns of not covering the cell phone-only population and potential undercoverage of the cell-mostly population in the landline-based RDD samples in MATS 2007. As discussed in Chapter 2, the distribution of the cell-only population (and the cell-mostly population to a lesser extent) is highly skewed. This is not only true for basic demographic characteristics, such the cell-only group containing young adults at a greater rate than landline-based households, but also for behavioral characteristics. Even when controlled by age groups, the population from cell-only households is known to report higher levels of risky behavior, such as smoking.

Thus, the general expectation is that the inclusion of the cell phone sample and in particular the population in cell-only households would contribute to a higher smoking prevalence rate than what would have otherwise been observed.

However, quantifying the degree of the effect is very difficult and assessing the impact on making comparisons to MATS 2007 is even more complicated: MATS 2007 included the Blue Cross sample; Blue Cross members in cell-only households were thus included in the frame and sampled members that provided their cell phone as a contact number could have been included in the completed interview results. So, some of the cell-only household population most likely contributed to the MATS 2007 estimates, even though this was not an explicit part of the MATS 2007 design.

The overall implication of this is that the observed decline in the smoking prevalence rate between MATS 2007 and MATS 2010 is likely suppressed due to the vastly improved coverage of the cell-only and cell-mostly population in MATS 2010, but to what degree is unknown.

Going back to the Blue Cross sample, other than the cell phone issue discussed above, the dropping of this frame from MATS 2010, in general, should not generate comparability issues. The inclusion of the Blue Cross sample in MATS 2007 was, in essence, an oversampling of this subpopulation that was already covered in the RDD sample. Oversampling impacts reliability of estimates, but not the expected values of the estimates themselves.

6.1.1.2 Elimination of Oversampling of Young Adults and African Americans

MATS 2010 did not oversample young adults and African Americans as was done in MATS 2007. In MATS 2007 this oversampling not only yielded a higher proportion of young adults (from the RDD and Blue Cross samples) and African Americans in the RDD sample, but it also yielded a higher proportion of non-African Americans who live in geographical areas with high concentrations of African Americans. However, this difference would not be expected to affect the point estimates themselves, in that the weighting and raking processes adjust for much of the difference resulting from oversampled populations in MATS 2007 relative to the more natural distribution of all Minnesotans on major demographic characteristics in the MATS 2010 sample. That said, the actual proportion of young adults and African Americans in the respective samples could affect overall point estimates, to the extent that differences between age groups and between race/ethnicity groups on the characteristic in question are not adjusted back to their natural distribution by the weighting process. However, the expected effect on the comparability of MATS 2010 and MATS 2007 would be minimal.

6.1.2 Weighting

6.1.2.1 Weighting Design

MATS 2010 used a similar weighting approach to MATS 2007, as the objective of both the MATS 2010 and MATS 2007 surveys was to make inferences about the entire civilian, non-institutionalized adult population in Minnesota. While the

particular steps employed and discussed in Chapter 5 for MATS 2010 may have been different than the particular steps employed in MATS 2007, the goal was the same. The differences in the weighting steps were dictated by the different sample designs as discussed above. So, in this regard the weighting design for MATS 2010 itself did not introduce additional comparability issues between MATS 2007 and MATS 2010 estimates.

6.1.2.2 Combining and Weighting the Landline and Cell Phone Samples, versus Combining and Weighting the RDD and Blue Cross Samples

MATS 2007 combined and weighted the (landline) RDD and Blue Cross samples. MATS 2010 combined and weighted the landline and cell phone. The specific steps employed, and even the ordering of the steps, to combine the overlapping sample frames and to compute and apply the compositing lambda factors were different in MATS 2010 and MATS 2007. However, as discussed in the previous section, these weighting step differences, driven by the sample design differences, were designed to achieve the same overall goal of assigning final weights to support making valid inferences about the entire civilian, non-institutionalized adult population in Minnesota. Therefore, the differences in the combining and compositing process did not, in and of themselves, directly contribute to additional comparability issues.

6.1.3 Data Collection

All of the MATS survey data were collected using computer-assisted telephone interviewing, and adhered to the CDC BRFSS data collection protocol. They all used standard survey research interviewer training and interviewing protocols.

6.1.3.1 Screener Refusal Conversion

MATS 2007 adopted refusal conversion of randomly subsampled screener refusals and accounted for them in the weighting process. MATS 2010, 1999 and 2003 did not subsample refusals. The additional subsampling in 2007 may have produced a very slight increase in the sampling variance and therefore in the statistical precision of the estimates, but not in the point estimates themselves. Like the previous MATS surveys, MATS 2010 attempted to convert 100 percent of those who initially refused to respond to the extended interview. None of the MATS surveys attempted to convert refusals that were classified as “adamant” or “hostile.”

MATS 1999 did not use refusal conversion letters; MATS 2003 sent refusal conversion letters at the end of data collection to a subset of the remaining screener refusals; and MATS 2007 and 2010 sent refusal conversion letters to all screener refusals that had a matched address associated with the phone number.

6.1.3.2 Home-based Interviewers

MATS 2007 and 2010 employed a mix of home-based interviewers and center-based interviewers. MATS 1999 and 2003 employed only center-based interviewers. There is no basis to ascribe any incomparability to the survey estimates resulting from the data collected under these two staffing models.

6.1.4 Questionnaire Design and Specification

There are two main areas where questionnaire design may affect comparability. The first area is the questionnaire content, which refers to the selection of questions, response categories, and the formulation of their specific wording and ordering. This area also includes the introductory text and transition phrases, as well as prompts, probes, and instructions to be used by the interviewers.

The second area is the determination of which respondents are administered each question and, for some questions, an alternative, more suitable phrasing of the question. This concept is commonly referred to as the “skip patterns” for the questionnaire. Some questions will not apply to certain groups because of who they are (questions about quitting smoking are not relevant to never smokers) or how they answered a specific question (if a person has not seen a health care provider in the past 12 months, it is logical to skip the questions about their experiences with health care providers in the past 12 months). The group who receives each question is usually referred to as the “base” for that question. In administering the questionnaire, the skip instructions determine and control who receives each question. All of these concepts are captured in detailed questionnaire specifications and in the programming instructions for CATI questionnaires.

6.1.4.1 Questionnaire Content

Section 3.1 described the general questionnaire design process and general issues and factors considered in formulating the question items, wording, and response categories. As noted there, and as elaborated in the Minnesota Adult Tobacco



Survey 2010 Comparability Report and its item-by-item crosswalk comparison between MATS 2007 and 2010, there are a number of questions that appear in only one of the two most recent MATS questionnaires. Such questions, by definition, have no comparable items for trend comparison across the survey rounds. When previous items were omitted from the MATS 2010 questionnaire, the resulting absence of trend data was consciously anticipated in the design, either because the items were no longer of interest or had not been useful in the past, or because some items needed to be eliminated as a trade-off to accommodate new items. In addition, when new items were added, it was because of new or changing research objectives. While historical trends cannot currently be analyzed for new MATS 2010 items, MATS may choose to retain such items in the future and monitor the trend from MATS 2010 forward.

In addition to noting questions that exist only in MATS 2007 or MATS 2010, the MATS 2010 comparability report and questionnaire crosswalk fully document a few wording changes made to clarify meaning, add or improve response categories, or simplify administration of questions appearing in both rounds. For the most part, the changes were minor and would not hinder meaningful comparisons across time. Appendix G discusses the nature and possible effect on comparability of MATS 2007 questionnaire items that were significantly modified for MATS 2010. The question numbers refer to the MATS 2010 questionnaire attached as Appendix A.

6.1.4.2 Skip Patterns in MATS 2010 Compared to MATS 2007

With the exception of two skip pattern changes discussed below, there were no significant skip pattern changes for MATS 2010. There were two instances where the process of defining MATS 2007 skip patterns and specifying them for the CATI questionnaire inadvertently excluded respondents from a series of questions that were needed for several desired analyses of both the MATS 2007 cross-sectional data and for trend comparisons with previous MATS. MATS 2010 rectified these problems. As a result, trend data for the affected items is available for 1999, 2003, and 2010, but not for 2007.

Stages of Change. The first of these oversights in 2007 impacted the ability to report on the standard concept of stages of change, a construct that classifies current

smokers as to their readiness to quit smoking and former smokers as to the durability of their currently quit status. Questions E16 and E17 (MATS 2007 question numbering) are needed to classify the stage of change for current smokers. The intention, consistent with MATS 2003, was for all current smokers, and all former smokers who smoked in the past 30 days, to be administered these questions, and for all others to skip to question G1. However, a skip instruction at an earlier point in the questionnaire (following E10) caused current smokers who had not made a quit attempt in the past 12 months to skip from that point to G1, thereby excluding them from the base for the stages of change questions. This affected approximately 46 percent (unweighted) of the smokers in the sample (48 percent weighted). As a consequence, MATS 2007 was unable to report on the stage of change for all current smokers in 2007 or to use this variable as an independent variable for other outcomes where all current smokers are the population of interest. Likewise, it was not possible to extend the trend from MATS 2003 to 2007 for any analysis that requires stage of change for all current smokers.

This problem did not affect determining the stage of change for former smokers or for the large subset of current smokers who had made a quit attempt in the past 12 months.

Self-efficacy for quitting and beliefs about stop-smoking medications. The same skip instruction that adversely affected collection of the stage of change data in 2007 also affected the two questions about self-efficacy for quitting (E18) and beliefs about stop-smoking medications (E19a –e). These data were unavailable in MATS 2007 for current smokers who did not make a quit attempt in the past 12 months. As a consequence, MATS 2007 was unable to report on these data for all current smokers in 2007 or to use any of these variables as an independent variable for other outcomes where all current smokers are the population of interest. Likewise, it was not possible to extend the trend from MATS 2003 to 2007 for the self-efficacy variable. The questions about beliefs about stop-smoking medications were new to MATS 2007.

Like stage of change, self-efficacy is most useful when looked at across all smokers, not only those with a quit attempt in the past 12 months. As a result, the MATS

2007 report, *Creating a Healthier Minnesota: Progress in Reducing Tobacco Use*, did not examine either of these constructs.

In comparison, analyzing their various beliefs about stop-smoking medications is useful even when limited to current smokers with a quit attempt in the past 12 months. Perceptions of quitting assistance may affect smokers' interest in them or willingness to use them. Further, successful quitters have usually made multiple quit attempts before being successful. Those current smokers who had recently tried to quit are the most likely to make another attempt soon, and supporting this group of smokers in future attempts to quit is critical to Minnesota's tobacco control efforts. Consequently, this topic was explored in the MATS 2007 report.

As noted above, the problems discovered in the 2007 instrument were rectified in 2010. The MATS 2010 data support analysis of the stages of change and self-efficacy for quitting, and allow for trend analysis of these topics from 1999 to 2010, with an interruption in 2007.

6.2 Potential Limitations of the Data

All of the Minnesota Adult Tobacco Surveys yield data that provides highly accurate and detailed representations of the smoking-related attitudes, beliefs and behaviors of Minnesota's adult residents at various points in time. However, statistics produced from sample surveys are subject to two general types of error, technically referred to as sampling error and nonsampling error. The term "error" does not refer to a mistake or a known error but to the fact there may be some difference between the survey statistic and the actual statistic for the entire population that the sample survey is meant to represent. It is for this reason that statistics produced from a sample are referred to as "estimates": they estimate what the actual statistics are for the entire population, or for any subgroup in the population.

6.2.1 Sampling Error

Sampling error is a purely statistical phenomenon, resulting from the fact that the data are collected from a sample that represents the entire population, rather than from everyone in the population, as in the case of a census. Sampling "error" is a

technical term; it does not refer to any known error, but rather to the fact that an estimate produced from a sample has some amount of uncertainty associated with it.

It is possible to quantify the uncertainty of an estimate produced from the survey sample data, to the extent that the uncertainty is caused by the use of sample with a known probability of being selected for the survey. There is no one number that can be assigned to every survey statistic to indicate the uncertainty; rather, it depends on the type of statistic (percentage, mean, ratio, difference, etc.), the size of the sample used to calculate the estimate, and the effects of complex sampling designs such as those used for MATS.

Common measures of uncertainty include standard errors and confidence intervals. The MATS technical reports utilize confidence intervals, which express the likely range of the actual value of a population statistic, around the “point” estimate produced from the survey data. For example, the statement that MATS 2010 found the 2010 smoking prevalence among adult Minnesotans to be 16.1 ± 1.2 percent means the expectation is that the true value falls somewhere within the confidence interval ranging from 14.9 percent to 17.3 percent. The confidence interval is commonly expressed as a “half-width,” plus or minus around the point estimate, as in this smoking prevalence example. Like nearly all sample surveys, MATS reports the 95 percent confidence interval, which means that there is a 95 percent certainty that the interval for any given estimate contains the true value.

All statistics presented in the MATS technical reports utilize weighted data. The survey weights reflect the complex MATS 2010 sample design, as described in Chapter 5. This means that the reported statistics are reflective of the entire population or subgroup for which they are calculated. The weighted estimates for the MATS 2010 technical report and their associated confidence intervals were all calculated using SAS, a widely used statistical software package that accounts for the complex sample design and sample weights.

6.2.2 Potential Sources of Nonsampling Error

As in the case of sampling error, it is nearly impossible for a survey to avoid other sources of error. Unlike sampling error, it is not typically possible to quantify potential nonsampling errors in a specific survey.

6.2.2.1 Frame Coverage Issues

In addition to the sampling error that is common to all sample surveys, MATS 2010 was also subject to a form of nonsampling error known as coverage error. All survey samples use a “frame” from which to draw the sample. Ideally, the frame “covers” the same population about which the survey seeks to provide information, but frames seldom perfectly cover the population. Those in the population who are not covered by the frame may be different from those who are covered by it, in terms of the characteristics, behaviors, attitudes, and beliefs that the survey addresses. The greater this difference is (if any), the greater the likelihood that there is some error in the reported statistics, in terms of their ability to accurately reflect the entire population of interest.

In the case of RDD surveys, which historically used landline telephone numbers as the frame, the coverage historically was above 95 percent, since less than 5 percent of the population was without landline telephone service at any point in time. In recent years, many households have begun using their cell phones as their household telephone line, abandoning totally their regular landline telephones. As of the first half of 2007, it was estimated that, nationally, 12.6 percent of adults lived in households that used cell phones exclusively.⁸ By the second half of 2009, that estimate had increased to 22.9 percent. The prevalence of cell phone-only households is considerably higher among younger adults (18-29), African Americans, Hispanics, and those with household incomes less than 200 percent of poverty level, and marginally higher in the Midwest.⁹ Such cell phone-only households are not covered by the RDD landline frame. Further, members of such

⁸ Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2007. National Center for Health Statistics.

⁹ Blumberg SJ, Luke JV. Wireless Substitution: Early Release of estimates From the National Health Interview Survey, July-December 2009. National Center for Health Statistics. Released 5/12/2010.

subgroups in households that rely solely on cell phones may be different in important ways from other members of the same subgroups.¹⁰

For these reasons, MATS 2010 added the cell phone sample. Adding the cell phone sample likely improved the overall coverage sample frame compared to 2007, and probably more so among certain groups more likely to be impacted by tobacco use, such as young adults and lower income individuals.

While the raking step for creating the sample weights can adjust for some of the difference in sample resulting from undercoverage, it cannot adjust for all of it, particularly for those characteristics that are not associated with the characteristics used for the raking – age, gender, race, and education. To the extent that adults who have only cell phone service, and younger or poorer adults in general, are different from other adults in relation to their smoking behavior and other characteristics of interest to MATS, the statistics reported for the overall population may have been affected by the under-representation in the sample in 2007. To the extent that younger adults in cell phone-only households are different from all young adults, the statistics reported for the young adult subgroup may have been affected by the absence of the cell-phone only households in the MATS 2007 sampling frame. Both of these situations may have been improved, if not fully rectified, by the addition of the cell phone sample in MATS 2010.

6.2.2.2 Measurement Error

Nonsampling errors in surveys may be attributed to a variety of sources, many of which fall under the type called measurement error. These sources of potential error may result from how the survey was designed, how respondents interpret questions, how able and willing respondents are to provide correct answers, and how accurately the answers are recorded and processed. MATS 2010 took several steps to minimize these types of errors. Important ones for MATS 2010, as described throughout this methodology report, include the careful and deliberate design of the questionnaire with review by multiple individuals and organizations;

¹⁰It is important to emphasize that the issue is the percentage of individuals living in cell phone-only households. An individual may personally choose to rely exclusively on a cell phone, but if he or she is a member of a household with a traditional landline phone, that individual was covered by the MATS 2007 RDD frame and still could be included in the sample.



continuing improvement to the clarity of several questions, balanced against possible effects on comparability of the responses across survey rounds; the use of a CATI system to administer the questionnaire and record responses; the internal testing of the CATI questionnaire; the pilot test of the instrument and survey procedures; the monitoring of the sample and of the collected data throughout data collection; and the thorough review of that data prior to finalizing the file for analysis.

The weighting process – especially the raking/post-stratification adjustments – partially corrects for bias due to minor discrepancies in the representativeness of the sample. During the weighting process, extensive diagnostic examination of the effects of the weighting design and of draft weights on the weighted estimates of demographics, smoking prevalence, and other characteristics further supported the “calibration” of the sample into closer conformity with the overall Minnesota population. Biases may be present when people who are missed in the survey differ from those interviewed in ways other than the categories used in weighting. People who are missed in the survey include those missed because of the frame coverage issue or because sampled individuals did not respond to the survey. As with most surveys that rely on telephone interviewing, it is likely that some subgroups, such as specific racial and ethnic minority communities, are under-represented; again, the use of race/ethnicity in the raking process helps reduce this effect.

All of these considerations affect comparisons across different surveys or data sources. Although most of these limitations are inherent in all surveys, MATS 2010 made every effort to minimize these limitations.



Appendixes

- A. MATS 2010 Questionnaire
- B. MATS 2010 Questionnaire Skip Pattern Table
- C. MATS 2010 Landline Screener Questionnaire
- D. MATS 2010 Cell Phone Screener Questionnaire
- E. MATS 2010 Letters
- F. MATS 2010 Web Page Content
- G. Significant Modifications of MATS 2007 Questionnaire Items for MATS 2010

NOTE: Appendix A is available at www.mnadulttobaccosurvey.org. The remaining appendixes and other reports referenced in this report can be requested by contacting Ann St. Claire at ClearWay Minnesota at astclaire@clearwaymn.org or by calling 952-767-1400.