



Vermont Division of Health Care Administration
2005 Vermont Household Health Insurance Survey

July 2006

Technical Documentation

Brian Robertson, Ph.D.
Jason Maurice, Ph.D.
Patrick Madden
Market Decisions

Table of Contents

	Page
I. Sampling Methodology	1
II. Questionnaire Design.....	9
III. Survey Pretesting.....	12
IV. Data Collection.....	13
V. Survey Response Rates and Final Dispositions	16
VI. Total Interviews	20
VII. Data Cleaning	21
VIII. Data Imputation	22
IX. Data Weighting	24
X. Precision	34
Appendices	37
Appendix 1. Mail Screening Survey.....	38
Appendix 2. Survey Pretest Report	41
Appendix 3. Weighting Plan	50
Appendix 4. Determination of Medicaid Eligibility	67

I. Sampling Methodology

This section outlines the sampling process used during the 2005 Vermont Health Insurance Survey. The sampling process consisted of two steps designed to meet overall statewide targets as well as specific targets for uninsured residents.

Target Population

The target population for the 2005 Vermont Health Insurance Survey consisted of all persons in families living in the state of Vermont, excluding those persons residing in households where no adult age 18 or over is present. Persons residing in group homes with nine or more persons, group quarters such as dormitories, military barracks, and institutions, and those with no fixed household address (i.e., the homeless or residents of institutional group quarters such as jails or hospitals) were also excluded from this survey¹. In addition, the sample excluded non-permanent residences and vacation residences (qualified households were considered those in which someone resided at least six months of the year). Since the sampling approach relied on the use of an RDD telephone sample, the sample population only included those households (and residents therein) with working telephones.

Sample Definition

The stated goal of the sampling approach was to obtain statewide population information on health insurance status, as well as gathering data on a number of demographic and health variables. The secondary goal was also to gather data from a sub-sample of uninsured residents with a precision of no less than plus or minus 4 percentage points. The methodology relied on a two stage RDD sampling methodology.

The sample was thus divided into two components with a set target for number of completed household interviews in a general population survey (GPS) as well as an over sample among the uninsured with a goal of gathering data from additional households with at least one uninsured household member. The target was to gather data on a minimum of 1,500 uninsured residents.

Table 1 Statewide and sub-population household interview requirements

Sampling Component	Target Household Interviews
General Population Survey	4,000
Uninsured Residents	430*

**This was the estimated number of uninsured household over sample interviews required to meet the target of 1,500 uninsured residents.*

¹ The initial screening coded as ineligible such group quarters. In this survey, group quarters' telephone numbers were considered those where a number of unrelated people living in more than one "unit" relied on the same telephone. An example of a unit in this case might be a fraternity house where all those residing in the house use the same phone.

It was anticipated that in conducting the uninsured over sample interviews that all household members would not necessarily be uninsured. Thus, the actual number of uninsured over sample interviews was approximated in the sampling design based upon the anticipated percentage of uninsured residents in Vermont and their distribution within households. The target for uninsured residents was set based upon precision requirements for this sub-population (a precision of plus or minus 4%) which required gathering data on a minimum of 1,500 uninsured residents. The sampling design anticipated that many uninsured residents would be identified during the general population survey and that the uninsured over sample would supplement the GPS in meeting the target goal of 1,500 uninsured residents.

Sampling Approach and Targeting of Sampling Components

The basic design of the sample process is presented in Table 2. The basis of all sampling used during the course of this research was Random Digit Dial (RDD) sampling protocols. During the first phase, a statewide RDD sample was generated. This resulted in statewide results with a minimum of bias introduced due to selective eligibility. During the course of the general population survey, all households were interviewed and data gathered on all residents within the household. During the GPS, the number of uninsured residents was evaluated to determine the total number of uninsured over sample interviews that would be required to meet the goals of this study.

Table 2. Sampling Process

Sampling Phase
Conduct Statewide General Population Survey
<i>(After 2000 GPS Interviews) Assess need for additional interviews sub-population targets</i>
Conduct remaining 2,000 GPS Interviews
Conduct additional uninsured over sample interviews to meet sub-population targets

This multi-stage approach was designed to most efficiently meet specified goals while minimizing any additional non-response bias due to inaccurate calculations of the number of over sample interviews (and drawing too few or too many sample records).

Market Decisions, LLC, generated the RDD sample in-house to derive the equal probability sample of telephone numbers. Within the data collection period, sample for both the general population component and the over sample component were entered in replicates to meet callback and refusal conversion goals.

Development of RDD Telephone Samples for this Research and Sample Generation

The proposed model relied on RDD samples as the sampling strategy. Any RDD sample used for this research must be designed to insure equal and known probability of selection (within each of the sampling stages). Market Decisions, LLC currently uses in-house software for generation of residential samples. The software is provided by Marketing Systems Group. The GENESYS sampling software is the first and only commercially available in-house sampling system with fully configured RDD design and generation capabilities. GENESYS supports RDD telephone sampling for any geographic area down to the census tract level. This includes state, county, metropolitan statistical area (MSA), ZIP Code, time zone, etc. The GENESYS system also contains telephone exchange-level estimates for over 48 demographic variables (e.g., age and income distributions) that can be used in conjunction with geographic definitions to produce truly unique geo-demographic sampling capabilities. The standard GENESYS RDD methodology produces a strict single stage, epsem sample of residential telephone numbers. In other words, a GENESYS RDD sample ensures an equal and known probability of selection for every residential telephone number in the sample frame.

In research projects that require a multi-stage sampling approach, our GENESYS software allows us to generate replicate samples, create new sampling cells in a matter of minutes, and has the capability to verify across all replicates and cells that numbers are not duplicated. For the 2005 Vermont Health Insurance Survey, two sampling stages were used:

RDD Samples
Statewide Sample (General Population)
Uninsured Over Sample (Statewide Sampling Frame)

Sample Generation

Both the general population sample and the uninsured over sample were generated in-house using GENESYS sampling software. The GPS sample was an RDD sample statewide, that is, the sampling frame covered the entire state. It was generated in proportion to the distribution of “population” of exchanges and telephone numbers throughout the state. Thus, a higher percentage of sample telephone numbers were generated in those areas with higher residential populations. The sample thus reflected the distribution of the population throughout Vermont. The uninsured sample also relied on a statewide sampling frame. That is, the sample of telephone numbers contained telephone numbers throughout the state.

In all, a total of 30,000 random telephone numbers were generated for the GPS component of the research and 20,000 random telephone numbers were generated for the uninsured component of the research.

Sample Cleaning

Any methodology that generates sample for RDD surveying produces non-household numbers. This is a simple fact that researcher must anticipate when the goal is to generate equal probability samples. Parameter estimates for a statewide sample generated through our GENESYS software provided several measures of size estimators to assist in the determination of the number of sample records needed.

Based on the GENESYS calculations, statewide samples generated for Vermont should have resulted in 44% of these numbers being residential telephone numbers. Given the inefficiency of such a high percentage of non-household numbers and potential impacts on response rates, Market Decisions used the GENESYS ID System to help remove non-productive numbers. GENESYS-ID is...

“ a process that takes any generated RDD sample, and then identifies non-productive numbers prior to the sample reaching the data collection phase of the project. The result is a sample that maintains its original statistical frame (providing full coverage of telephone households), but approaches the efficiencies of listed household samples. Consequently, interviewer productivity increases (as interviewers spend more of their time working with productive phone numbers), and data collection costs are reduced.”

GENESYS SAMPLING SYSTEMS

The system is designed to:

- Purge businesses (GENESYS-Plus is a part of the GENESYS-ID process).
- Identify non-working/disconnected phone numbers.

In the process, the system does not deter from a study's statistical validity.

No such system can remove all non-productive numbers. The GENESYS ID system will identify from 35-45% of non-productive numbers and eliminate them from the sample.

Following cleaning through GENESYS ID-Plus, a total of 21,902 telephone numbers were retained in the GPS sample and a total of 12,562 were retained in the uninsured over sample.

Sample Entry/Replicates

It is counter-productive to enter the entire potential sample at once. It is not possible to contact every potential respondent within the first few days of the study, given the large sample size. In addition, if efforts prove more efficient than anticipated it may result in the need for less sample than originally thought. Entering all sample at the beginning would then adversely affect response rates as numbers would not be resolved. Market Decisions entered sample as a set of replicates throughout the data collection process. The entry of each replicate was timed so that numbers in prior replicates had been sufficiently resolved and that later replicates were entered in order to provide adequate time to meet callback requirements.

In all, the general population sample was entered in 12 replicates throughout the data collection period. That is, the original file that was generated for the general population was divided into replicates. A sample replicate was entered and calls made to these cases. As numbers were resolved, an additional sample replicate was then entered. Spacing between entries of the replicates was generally one week. This allowed adequate callback attempts before new cases were entered. Call attempts were still made on all replicates of the sample throughout the data collection process based upon data collection protocols (that is, attempts weren't halted when a new replicate was entered). The over sample was entered in 6 replicates.

Uninsured Over Sample Screening

Given that the anticipated percentage of uninsured residents in Vermont was quite low, efforts were made to increase the data collection efficiency in conducting the uninsured over sample. Based on results of the 2000 Health insurance survey, only 8% of residents in Vermont were uninsured and approximately 12% of Vermont households had at least one resident that was uninsured. It was anticipated that only one out of every eight households would be eligible for inclusion in the uninsured over sample. Several steps were taken to improve the data collection efficiency in reaching households with uninsured residents while still maintaining the statistical properties of the sample. These steps included the use of a mail pre-screening survey sent to all sampled household for which an address could be obtained and using a screening question at the beginning of the telephone survey to identify and eliminate all households in which all residents were insured.

As noted, the actual sample of telephone numbers generated for the uninsured over sample included all telephone exchanges in Vermont. That is, the uninsured over sample did not target a specific area of the state².

During the course of the research, a pre-screening survey was mailed to all households in the uninsured over sample for which an address could be identified (5,425 total households). The purpose of the pre-screening survey was to identify household in which all residents were covered by some type of health insurance. That is, to identify households that would be considered ineligible for the over sample component of the research. Addresses were obtained using an address look-up based on the telephone number. Of the 12,652 random telephone numbers included in the uninsured over sample (after cleaning through GENESYS ID-Plus), addresses were obtained for 5,425 sample records. A copy of the mail survey is included in appendix 1.

The mail screening survey was mailed to these 5,425 households on November 29, 2005. Of these, 985 were returned as undeliverable. A total of 1,585 completed surveys were returned to Market Decisions. Among these, 1,343 respondents indicated that all members of the household

² Often times when generating telephone numbers for sub-population over samples one can restrict the geography in order to increase the incidence of the specific sub-population above the statewide average. This is done, for example, in conducting ethnic minority over samples since ethnic groups tend to concentrate in certain geographic areas. In the case of this over sample of uninsured residents, it was not possible to restrict the geography in order to increase the incidence of uninsured among the sample. This is due to the fact that there are no significant concentrations of uninsured residents in any specific area of Vermont.

were covered by some type of health insurance. Another 37 respondents indicated that residents in their home were not permanent residents of Vermont (either the home was a vacation home or no one resided in the household at least six months of the year).

The results from the mail survey were then noted in the telephone sample used during the uninsured over sample. As sample replicates were drawn during the course of data collection, any records flagged as a result of the response to the mail survey were assigned the appropriate disposition and set aside. That is, the records were classified as ineligible in cases where the mail survey indicated all household members were insured or in cases where the household was not a permanent residence. Call attempts were not made to these households. It is important to note that the final determination of eligibility was made as a sample replicate was entered and not prior to sample being entered for data collection. While sample records that were ineligible based on the response to the mail survey were flagged in the sample file they were not actually removed. This procedure was followed for two important reasons. First, we did not want to change the random order in which sample records were entered. Removing these numbers prior to sample entry would have changed the random selection order of the remaining numbers in the sample file and changed the probabilities of selection, impacting our ability to calculate sample weights. Rather these ineligible numbers were “allowed” to be selected in the entry of sample replicates and were then assigned a call disposition. In practical terms they were treated in much the same way as if they were called and through the telephone call were determined to be ineligible. The second reason was that the number of sample records required for the uninsured sample was estimated. There was the possibility that we had underestimated the incidence of the uninsured in Vermont and that sample would have exceeded our need to meet precision requirements. In such cases, all sample records that were unused during data collection should be treated the same in developing final disposition reports and sample weights regardless of whether one knows in advance that the household was technically ineligible. In the case of this specific research study, however, this was not a significant factor since our sample requirement estimates were accurate.

The mail screener represented the first phase of screening to identify households with at least one uninsured resident. After the mail survey results had been flagged in the telephone sample, sample for the uninsured over sample was entered in replicates. As noted, cases that were identified as ineligible through the mail survey were assigned a case disposition as entered into our CATI software and no call attempts were made. Call attempts were made to all other sample records. This included calls to:

- Sample records for which no address could be obtained.
- Sample records for which an address was obtained but the mail survey was returned as undeliverable.
- Sample records for which an address was obtained and the respondent indicated that at least one household member was uninsured.
- Sample records for which an address was obtained and the respondent was unsure of the insurance status of all household members.
- Sample records for which an address was obtained but the mail survey was not returned.

All records were treated the same regardless of their status after the mail screening survey. The order in which they were entered into the CATI program was determined prior to the mail screening survey and this order precedence was not changed as a result of the mail survey. For example, if a survey was returned that indicated people in the household were uninsured, this record was NOT automatically called. Rather, it was entered into the CATI program in a replicate based on the pre-assigned order.

When a telephone number was contacted and identified as a Vermont household, the person in the household most knowledgeable about the health insurance status of the household was then asked if all members of the household were covered by some type of health insurance. The question wording was based on the question from the mail screening survey. If the respondent indicated that “yes,” everyone was covered by some type of health insurance, the case was deemed ineligible and assigned a final disposition. In all other cases (if there were one or more uninsured individuals or the respondent was not certain of the health insurance status of ALL members of the household) the survey continued and respondents were asked the same questions included in the general population survey.

In all other aspects, the uninsured sample followed the same data collection protocols as the general population survey (number of attempts, callback scheduling, refusal conversions, and other data collection criteria).

Sample Representation

One important source of bias in telephone surveys is that households without telephones are artificially eliminated from selection as are those who experience an interruption in telephone service. Thus, a component of the population is not able to participate. In RDD telephone surveys, Market Decisions typically relies on households that have experienced an interruption in telephone service to represent this component of the population and also to adjust for the probability of selecting a home that may experience telephone service interruption:

Market Decisions relied on two questions to measure service interruption:

1. Was there anytime in the last 12 months that you did not have a working telephone for two weeks or more?
2. IF YES: For how many months of the past 12 months did you not have a working telephone for one week or more?

Households with an interruption in telephone service were then weighted to represent households without telephone service and to make appropriate weighting adjustments for households that experience service interruptions.

One other biasing factor is the fact that households may have more than one telephone. A household with more than one phone has a greater probability of selection (in proportion to the number of telephones in the household) than a household with only one telephone. To correct for this bias, we ask respondents a set of questions about the number of telephones in the household:

- The number of telephones in the household
- The number of telephones that are used exclusively for business
- Whether the contacted telephone is a business telephone exclusively

During the non-response weighting phase, data was weighted in proportion to the number of residential telephones in the household to balance out the greater probability of selection among those with more than one telephone. For the purposes of data collection, cellular phones are not included in the weighting adjustment, which are based on landline phones. There is one exception. Cellular phones are considered in cases where the cellular phone represents the only telephone for the household.

II. Questionnaire Design

The survey questionnaire that was used during the course of the 2005 Vermont Health Insurance Survey was based largely on prior surveys conducted in Vermont on this topic as well as health insurance surveys administered by Market Decisions in several other states. The initial steps in survey design focused on a review of the 2000 Vermont Health Insurance Survey as well as reviewing surveys administered by Market Decisions in Rhode Island and Pennsylvania. The initial design work involved Dr. Robertson of Market Decisions, Dian Kahn of the Vermont Division of Health Care Administration, and technical assistance provided by the University of Minnesota State Health Access Data Assistance Center (SHADAC). The core areas and most of the survey questions used in this study were derived from these consultations and from the 2000 Vermont, Rhode Island, and Pennsylvania survey instruments. Additional questions or subjects areas were provided or recommended by other agencies within Vermont to address specific topics or to obtain more specific input on existing survey topics.

A draft of the survey instrument was submitted to the Vermont Division of Health Care Administration on October 4, 2005. After incorporating changes, a final pretest version of the survey was completed on November 1, 2005.

The basic components of the 2005 survey gathered information from Vermont residents in the following areas:

1. Household Characteristics
2. Enumeration of the Household
3. Demographic Characteristics of each Household Member
4. Relationships Between Household Members
5. Type of Health Insurance Coverage
6. Private Health Insurance Coverage Characteristics
7. Medicare Supplement Coverage Characteristics
8. Medicaid, Dr. Dynasaur, and VHAP Awareness and Communication
9. Uninsured Characteristics
10. Interest in Medicaid Enrollment
11. Interruptions in Insurance Coverage
12. Concerns About Loss of Health Insurance
13. Dental and Vision Insurance Coverage
14. Prescription Drug Coverage through VHAP Pharmacy, VSCIPT, and Healthy Vermonters
15. Medicare Prescription Drug Benefits
16. Prescription Drug Expenses
17. Visits to Health Care Professionals – Point of Service
18. General Health Status and Chronic Conditions
19. Health Care Barriers
20. Employment Characteristics
21. Income, Debt and Assets Characteristics (family level)

Family Formation

One important concept that was incorporated into the 2005 Vermont Health Insurance Survey was that of family units. This concept is important because of the relationship between variables such as private or governmental insurance coverage and family level characteristics such as income. The survey logic was designed so that all members of a household were grouped into family units based upon their relationships. The survey was structured to ask the questions about each family unit separately.

Family units were identified by establishing the relationship of each member of the household to the identified head of the household. This was done by first collecting the number of people in the household and a name or other identifier for each person. The household was then rostered and basic demographic information gathered on each household member (age, gender, marital status, ethnicity, race, level of education, and whether those age 18 - 23 were still in school). The respondents were then asked to describe the relationship of each member of the household to the head of the household. Two follow-up questions then clarified marital relationships between household members besides the head of household and their spouse and any guardian/ward relationships. Based upon this sequence of questions, household members were classified into family units. In general, the rules to assign members to family units were:

1. The head of household and their spouse, domestic partner, or civil union partner were classified in the same family unit (always family unit 1).
2. Adults 23 and older who were not married, a domestic spouse, or civil union partner of the head of household were classified as a separate family unit (each considered separate unless there was a marital/parental/guardianship relationship to someone else in the household).
3. Married couples, domestic partners, and those in civil unions were classified in the same family unit with the exception noted below.
4. Married couples, domestic partners, and those in civil unions involving someone under 17 were grouped based upon their relationship to others in the household. If such a person was the child/ward of another household member they were classified in the same unit as their parent(s)/guardian and their spouse/partner in a separate unit. In those cases where they were not the child/ward of another household member, they and their spouse/partner were grouped as a separate family unit.
5. Children 17 and younger were classified in the same unit as their parent(s)/guardians. If their parent(s) or legal guardian did not live in the household, they were considered a separate family unit.
6. Children age 18 to 23 were classified based upon whether they were currently full time students in high school or in post secondary education institutions. Those who were full time students were classified in the same unit as their parent(s)/guardian (with exceptions noted below). Those who were not full time students were classified as a separate family unit.
7. Children who were 18 to 23 who were a spouse/partner of another household member or someone not residing in the household were considered a separate family unit.
8. Children who were 18 to 23 and who had a child of their own either within the household or outside the household were considered a separate family unit.

9. Finally, those who were identified as the ward of another household member were classified in the same unit as that household member unless prior rules determined the ward should be classified separately.

Eliciting Cooperation

Given the response rate requirements of the 2005 Vermont Health Insurance Survey, special attention was paid to survey elements designed to elicit cooperation. A number of design elements incorporated into the surveys helped maximize response rates. These elements included:

- Clear lead in and introductory statements that explained the nature of the research.
- Informing contacts who we are.
- Providing the name of the client.
- Persuader statements that explained why the research is important and why it is important for them personally to participate.
- A toll free telephone number and the name of the primary investigator (Dr. Robertson) so a potential respondent could verify that the research was legitimate or ask any questions about the research.
- A toll free telephone number and the name of the primary contact at the Vermont Division of Health Care Administration (Dian Kahn) so a potential respondent could verify that the research was legitimate or ask any questions about the research.
- A statement of implied consent that indicated the research is confidential and their name will in no way be associated with results; the results are reported in aggregate form only. The statement also indicated that the call may be monitored. Finally, it also indicated that if they do not wish to answer a question that is fine.
- Coded help screens that contained information about the research and selection process that interviewers provided to potential respondents.

III. Survey Pretesting

The design process for the 2005 Vermont Health Insurance Survey included an extensive survey pretest phase. This pretest phase was designed to finalize the survey instrument developed by Market Decisions and the Vermont Division of Health Care Administration staff by evaluating the survey logic, family unit formation logic, clarity of questions, anticipated survey length, and need for term definition. The pretest phase of the research project was begun on October 16, 2005 and was completed by November 6, 2005. The pretest phase relied on input from a number of sources including the research staff at Market Decisions, the staff of the Vermont Division of Health Care Administration, the field staff manager, field staff supervisors, interviewers, and finally residents of Vermont that were called and asked to complete the survey.

The survey was first programmed into our Computer Assisted Telephone Interviewing (CATI) software. The initial reviews of the survey questionnaire were conducted by Dr. Robertson and Patrick Madden in order to confirm questionnaire logic was correct and that the survey functioned as anticipated. After these initial logic tests, the research staff provided test copies of the program to the data collection staff. The field staff manager and supervisors were briefed on the project and then taken through the survey with explanations provided for the meaning, context, and intent of each survey item. The field staff was also provided with paper copies of the survey to allow them to assess logic and flow.

The field staff, including the field staff manager, supervisors, and interviewers, was then asked to go through the survey and note any problems that were observed. These problems were then passed back to the research staff and corrections were made to the survey questionnaire and CATI program logic to correct these problems.

The final step in the pretest phase involved live interviews with Vermont residents. A total of 60 pretest interviews were conducted with randomly selected Vermont residents to test the survey questionnaire. The pretest of the 2005 Vermont Health Insurance Survey was begun on November 2, 2005 and was concluded on November 6, 2005. A total of 60 pretest interviews were conducted with Vermont residents. These interviews included

- 17 interviews with households with children
- 15 interviews with households in which at least one resident was 65 or older and a Medicare beneficiary
- 10 interviews with households in which at least one person was uninsured

Pretest households ranged in size from 1 to 6 persons. On average, the pretest survey required 24 minutes to administer.

Vermont residents who took part in the pretest were asked to complete the survey as if they were a respondent, but they were also asked to provide feedback on questions. Specifically, they were asked to provide feedback if they were unclear about the intent of the question, if there were terms they did not understand or if the flow of the survey did not make sense or seemed confusing. Feedback from these pretest interviews was then used in the development of the final survey questionnaire. During this pretest process, changes were made to improve survey flow,

and to add clarity to questions that respondents indicated difficulty in understanding and the deletion of items to meet survey length requirements. A copy of the survey pretest report is included in Appendix 2.

IV. Data Collection

The data collection phase of the 2005 Vermont Health Insurance Survey (GPS) was begun on November 9, 2005 and data collection was completed by February 10, 2006. A total of 4,010 households were interviewed during this period for the general population survey component of the research.

The data collection phase for the uninsured over sample was begun on December 12, 2005 and data collection was completed by February 15, 2006. A total of 453 households with at least one uninsured resident were interviewed during this period for the uninsured over sample component of the research (a total of 5,655 households were ineligible because all households members had health insurance coverage).

In order to meet response rate requirements for this study, a rigorous data collection strategy was used in conducting this survey. This included the following:

- Rotation of call attempts across all seven days at different times of the day according to industry standards for acceptability and legality in telemarketing.
- A minimum of 15 call back attempts per telephone number at the screener level (before number was identified as a qualified residential number).
- 4 attempts to convert refusals (the exception were those households that made it clear they were not to be contacted again).
- A minimum of 10 callback attempts for “no answer” or answering machine only telephone non-contacts and for inappropriate contacts (contact only, no most knowledgeable adult home), and scheduled callback appointments.
- A brief message with a toll free number was delivered to answering machine only attempts to encourage participation (messages were left on the first, third and seventh answering machine dispositions).

Per industry standards, interviews were only conducted during the hours from 9 AM to 9 PM and seven days a week. The only exceptions were specific, scheduled appointments outside this range.

Responding to Vermont Residents Inquiries about the Survey

One strategy that was used in order to increase response rates was providing reluctant residents with the name and telephone number of the primary investigator (Dr. Robertson) as well as that of Dian Kahn of the Vermont Division of Health Care Administration. Over the course of data collection, both parties received a number of calls from potential survey respondents. Dr. Robertson responded to calls from 87 respondents. In most cases, the resident called either to simply verify the legitimacy of the survey, get more information about what the survey asked, or to respond to a message left on their answering machine. A few calls were made to have the

respondent's telephone number removed from our sample. Depending on the timing of the call, the resident was called back according to the callback protocol or the survey was completed at that time. Nearly all of those who contacted Dr. Robertson ended up completing the survey.

Scheduling Callback Appointments

The CATI system used by Market Decisions during the course of this survey is designed to allow interviewers to set callback appointments for a specific date and time. It is also designed to allow a respondent who has begun the survey and cannot complete it to complete it at a later time. This is done so that the respondent can complete the survey at a time that is most convenient for him or her. The interviewer enters the date and time the respondent provides and the respondent is then contacted at that time. Over the course of the data collection phase, 5,434 scheduled appointments were made. Approximately 31% of interviews that were completed were done so with respondents that had scheduled these specific appointments.

Survey Length

The 2005 Vermont Health Insurance Survey required respondents to provide a great deal of information about themselves and other family members. The goal was to obtain accurate information about all household members while limiting the time commitment required of the respondent. Our goal was a survey instrument that would require an average respondent about 20 minutes to complete.

On average, the 4,010 GPS interviews required 19.4 minutes. Sixty-five percent of the interviews were completed in 20 minutes or less. The shortest amount of time required was 7 minutes while the longest survey required 59 minutes.

On average, the 453 uninsured over sample interviews required 22 minutes. Forty-six percent of the interviews were completed in 20 minutes or less. The shortest amount of time required was 15 minutes while the longest survey required 47 minutes.

Exclusion of Household Members

In multiple family households, it was expected there would be cases where the respondent would not be able to provide accurate data on every person living in the household. During the course of the survey, the respondent was asked to identify any household member for which they felt they could not provide accurate information. During the course of this interview the respondent was not asked questions relating to these individuals. At the end of the survey, the interviewer asked the respondent to identify who in the household would be able to answer the survey questions about any excluded individuals. If the respondent could identify an individual this was noted and a callback was scheduled to complete the survey with this individual. At a later time, the household was re-contacted and the interviewer asked for this identified individual. The interviewer then asked this respondent all survey questions except the questions on household and person characteristics (demographics). These were skipped since these data were gathered during the initial interview with the household. In some cases, it was not possible for a respondent to identify an appropriate contact to answer questions for those excluded during the interview. In other cases, it was not possible to contact this identified individual³. In all there were 113 instances in the GPS component of the research and 31 instances in the uninsured over sample where it was not possible to contact this individual. In such instances, a weighting adjustment was made to take this into account.

³ In terms of callbacks, such instances were treated as new cases.

V. Survey Response Rates and Final Dispositions

The goal set for this research study was to obtain an overall response rate of 65% for the GPS component of the research. In calculating survey response rates, a system developed by Market Decisions in conjunction with Mathematica Policy Research Institute for use on similar research projects was adapted for use in calculating one of the response rates reported for this survey. This rate is referred to as the CASRO response rate. Additionally, AAPOR response, cooperation, and refusal rates are provided.

The CASRO response rate calculations were derived by examining the patterns of response at several stages of the interviewing process (from initial identification of a residential number through interview completion). The response rate calculation was designed to match the weighting scheme used to adjust the data by non-response. Table 3 outlines each of the steps in the response rate calculation.

Table 3. Definition of Response Rate

Step	Process
Working Residential Status	Identification of number as a residence.
Determination of Eligible Residence	Identification of a residence as meeting all eligibility requirements.
Family Unit Formation	Eligible Households completing section of survey on family unit formation.
Questionnaire Completion	Households that completed the survey.

At each of these stages, a stage response rate was computed. For example, the working residential status response rate is the number of identified residences divided by the number of identified residences plus those numbers for which residential status had not been determined. The overall survey response rate was the product of these four individual response rates.

The final disposition code assigned to each number was based upon the call outcome as well as whether the number had been identified as a household, identified as an eligible household, identified as ineligible, or undetermined. Based upon disposition and determination of residential status and eligibility, all disposition codes were classified into eight eligibility classes. These classes are presented in Table 4. Upon completion of the survey, a final disposition report was developed. This final disposition report is presented in Table 5. It reports dispositions separately for the GPS component of the research and for the uninsured over sample.

Table 4. Eligibility Classes Used in Reporting Final Case Dispositions

Eligibility Class Code (ELIGRESP)	Eligibility Class Description
1	Completed Interview – All
2	Completed Interview – Exclusions not completed
3	Eligible Household, Non-interview, Family Formation Completed
4	Eligible Household, Non-interview, Family Formation Not Completed
5	Working Residential Number – Ineligible
6	Working Residential Number – Undetermined Eligibility
7	Ineligible – Business, Institution/Non-working Number
8	Undetermined

Table 5. Final Sample Disposition Codes

Disposition code	GPS Sample	Uninsured Over Sample	Total
Complete	4010	453	4463
Partially completed survey	3	0	3
No answer	1040	704	1744
Busy	84	121	205
Answering machine	681	371	1052
Scheduled callback	8	0	8
Call blocking (caller ID blocks)	0	1	1
Contact only	8	2	10
Hang up	89	31	120
Hard respondent refusal	203	49	252
Hard household refusal	1247	269	1516
Partial terminate	122	17	139
Soft household refusal	102	16	118
Soft respondent refusal	5	1	6
Disconnected phone	3615	1679	5294
Wrong number	4	4	8
Not a working number	7	0	7
Fax or modem	1083	751	1834
No ring	345	93	438
Cell phone or pager	58	47	105
Number changed	101	65	166
Number not in service	5066	737	5803
Not available in time frame	8	1	9
Language barrier	7	0	7
Infirm	7	0	7
Business	2636	897	3533
Other	874	251	1125
Group quarters or institution	81	38	119
Not a permanent residence	36	23	59
Vacation home	372	286	658
No one in household is uninsured	0	5655	5655
Total	21902	12562	34464

Survey Response Rates

The response rates to the 2005 Vermont Health Insurance Survey are presented in table 6, along with cooperation and refusal rates. Both AAPOR (RR1) and CASRO response rates are provided along with AAPOR Respondent Cooperation (COOP1) and Respondent Refusal Rates (REF1).

Table 6. Survey Response Rates

Component	GPS Sample	Uninsured Over Sample⁴
CASRO		
Response Rate	58%	33%
AAPOR		
Response Rate (RR1)	53%	22%
Cooperation Rate (COOP1)	92%	87%
Refusal Rate (REF1)	4%	3%

⁴ Note that the response rates reported for the over sample are both conservative response rates. For these rates, all cases of undetermined eligibility are factored into the calculations. In all likelihood, the majority of these undetermined cases would be ineligible since the percentage of Vermont households with an uninsured household member is quite small. Less conservative response rates use formula that estimate the number of eligible cases among those that are undetermined and then calculate response rates based on the estimated counts of eligible respondents. Based on these less conservative formulas the AAPOR (RR3) response rate for the uninsured over sample would be 80% and the CASRO response rate would be 85%.

VI. Total Interviews

A total of 4,463 households were contacted and interviewed. For the GPS component a total of 4,010 interviews were conducted while a total of 453 uninsured over sample interviews were conducted.

The final data includes data on 10,976 Vermont residents. The general population survey component data set (which is used to make statements about the state as a whole and sub-populations of the population) includes data on 9,754 Vermont residents.

The final data set also contains data from 1,550 uninsured Vermont residents (combining those from the GPS and uninsured over sample)

Table 7. Number of Households Interviewed and Residents Providing Data

Strata	Households Interviewed	Residents Included
General Population	4,010	9,754
Uninsured Over Sample	453	1,222
Uninsured Residents*	---	1,550

**Total number identified in both the GPS and uninsured over sample. The number reported under "Households Interviewed" represents the number of households with at least one uninsured member.*

VII. Data Cleaning

Any survey process can result in erroneous reporting or recording of data. To insure the accuracy of the data, Market Decisions conducted data consistency checks on the data files. As a part of the data file preparation and analysis, the first stage of this process involved checking all data to insure that responses were consistent. This process involves insuring that respondents were asked appropriate questions based upon earlier responses to variables, skip patterns were followed based upon appropriate responses to earlier items, and that respondents provided consistent answers to questions on related concepts.

The initial steps of data consistency checks were programmed into the survey instrument themselves. These included verification items on key issues. An example includes the verification of Medicare coverage as opposed to Medicaid coverage among those under 65. The programmed data checks insured that respondents were directed to appropriate questions and that answers to some key issues were verified.

There are three possible sources of data errors that the survey programming could not fully account for in its design. These were

- Respondents who, after completing questions or entire sections of the survey changed their minds about the answer they had provided.
- Respondents, whether due to lack of information or unfamiliarity provided inaccurate information.
- Respondents who answered a question or question in one fashion and then provided a different answer to a related question later on in the interview.

In the first case, interviews could back up in the survey instrument and enter the corrected information. The CATI software used by Market Decisions would then correct answers based upon new branching or skip patterns.

The second case is primarily related to knowledge of specific insurance plans, primarily government sponsored plans, which provide coverage to family members. The two most notable examples were respondents who confused Medicare and Medicaid coverage and respondents that confused Medicaid coverage with coverage through private health insurance.

In the last case, the data was left coded as provided by the respondent. The decision was made not to challenge respondents by indicating they had provided conflicting answers to similar survey questions.

VIII. Data Imputation

Data Imputation

Given the nature of the survey data collected, it was decided that missing values would be imputed on certain key values, particularly weighting variables. Data imputation is a procedure that determines the likely value of a given variable based upon other known characteristics of the respondent. Imputation relies on answers to other questions to derive the most likely value for the missing value. Market Decisions used data imputation on several of the variables in this research. In those cases where a variable was imputed, the final data set contains a copy of the variable with imputed values, a copy of the original variable with missing values retained, and a flag variable which identifies which values were imputed and the method used. The research staff used three primary methods of data imputation:

Logical Imputation

This step involved an assessment of answers to other questions (within the case) to determine if it were possible to deduce the answer to a question with a missing value. In some cases, this was done by evaluating a question that was very similar in nature and content. In other cases, it involved assessing a number of related questions to derive the most likely value. The initial survey design anticipated this approach, somewhat. There were a number of consistency checks programmed throughout the survey on certain key variables. These consistency checks were used during the course of imputation to impute missing values to certain key variables.

Donor Substitution Imputation – Hot Deck Imputation

Hot deck imputation relies on the fact that individuals with similarities on a number of variables are likely to be similar on those variables with missing values. The process involves identifying an individual with similar values on other variables and substituting this person's response for the missing value. In each of these cases, a number of variables were used to identify those respondents that were similar to a respondent with a missing value for a specific variable. The types of variables that were used to define characteristics that are "similar" varied depending on the nature of the variable to be imputed. These included key demographic characteristics and variables with a high correlation to the variable imputed. Once defined, the process of imputing the missing value relied on replacement. Base upon defined characteristics, the file was sorted in "serpentine" fashion (alternating ascending and descending sorts on variables). The value from the "nearest neighbor" was then used to replace that of the missing value.

Regression Based Imputation

For certain variables, such as income, the use of regression-based imputation was the most suitable method. This process relied on regression analysis to predict the value of the variable. The process relied on the use of analytical software that is designed to conduct missing values analysis. As with hot deck imputation, the number and type of variables used during regression analysis varied by the variable that was imputed but this also relied on key demographic variables and those correlated with the variable containing missing data.

The primary variables that were imputed were those used in weighting the survey data (gender, race and ethnicity). In addition, income (in terms of federal poverty level) was also imputed. This was important since missing values would cause problems with the post stratification weighting of the data. Those cases with missing values would not have appropriate adjustments made and this would lead to an increase in variance since their weights would differ from those cases with complete demographic data. The data imputation process “estimated” any missing values in those variables used in post stratification weighting to minimize their impact on data quality. The method of imputation used for these variables is as follows:

Gender	Logical Imputation
Age	Logical Imputation
Ethnicity	Logical and Hot Deck Imputation
Race	Logical and Hot Deck Imputation
Income	Regression Based Imputation

IX. Data Weighting

NOTE: The final data set contains two final weights (though both use the same steps as outlined in the weighting plan). The first is calculated only for the GPS component of the research. This weight is used when making statements about Vermont residents in general and when used the survey results can be said to reflect the views of Vermont residents. This weight is used in all analysis of survey data with the exception of sub-population analysis of the uninsured population of Vermont and their households. This weight is used when deriving estimates of the rate of health insurance coverage by various types demographic characteristics. The weight is also used to derive population estimates of the number and percentage of uninsured residents in Vermont.

A second weight was calculated for uninsured residents to factor in respondents from the uninsured over sample. Analysis using this weight does not provide results that can be said to reflect the view of all Vermont residents. Rather, this is the analytical weight used when analyzing the population of uninsured Vermont residents and the households in which they reside (though it is not used to estimate the uninsured population). This allows analysis of smaller sub-populations within the uninsured population and provides greater precision with the results that are reported.

The data has been weighted to adjust for non-response and also to match the state profile based upon sex, age, race, ethnicity, and area of residence for the GPS component of the research. The data has also been weighted to reflect the profile of the uninsured based on age, gender, race, ethnicity and area of residence. Weighting also adjusted for households with multiple phone lines and for interruptions in phone service. The weighting procedures involved two primary phases: Non-response weighting adjustments and post stratification weighting adjustments. Weighting was handled sequentially by weighting household level data, family level data, and finally person level data. The formulas and procedures used in weighting are provided in appendix 3.

An initial sample weight was assigned to each record in the sample file. This base weight was equal to the inverse of the probability of selecting a number within each of the samples (general population and uninsured over sample). Non-response weighting adjustments were then made at each determined level of eligibility (the eligibility status). The eligibility status codes are given in table 4 on page 18. For example, the first non-response weighting adjustment distributed the component of the probability of selection for the “undetermined” cases (cases where no information about the eligibility of the telephone number) among those cases with an eligibility status of one through seven. When completed, the non-response weighting adjustments allocated the probabilities of selection from all sample records into the final set of completed cases.

During the non-response weighting phase, an adjustment was also made for households with more than one telephone. The weighting adjustment simply adjusts for the probability that a household with more than one telephone has a greater probability of selection.

Interruption in Telephone Service

One concern with telephone surveys is the issue of under coverage because a household without a telephone cannot participate in the research. In order to adjust for this population as well as the probability that a household may have a temporary service interruption, a service interruption weighting adjustment was made to the data set. The weighting process assumes that those households that have experienced an interruption in telephone service have a lower probability of selection. Further, it assumes that those with an interruption in telephone service can be used to “represent” those without phones service in any analysis.

Market Decisions relied on two questions to measure service interruption:

- Was there anytime in the last 12 months that you did not have a working telephone for two weeks or more?
- IF YES: For how many months of the past 12 months did you not have a working telephone for two weeks or more?

The weighting adjustment applied to a case was $12/(12-m)$ where m represented the months of interrupted phone service. In order to avoid too significant a variation in the final sample weights, m was allowed to obtain a maximum value of 6 (with instances of an interruption of more than 6 months recoded to 6 for weighting purposes).

Post Stratification Weighting

GPS Component

The purpose of post stratification weighting is to standardize the weights so they sum to the actual population within Vermont as well as summing to the population by area, age, gender, race, and ethnicity. Post stratification weighting adjustments were made by age and gender within county. Post stratification weighting adjustments for ethnic origin and race were made at the state level.

Demographic data on population counts was developed from the 2004 population estimates, from the US Census Bureau. The final weighting numbers were based on this estimate of the 2004 population in Vermont.

An initial review of survey and census data was conducted to determine the appropriate steps in the weighting process. The general guideline in post stratification weighting is that no cell should have fewer than 20 cases. The original intent was to use two weighting steps: age by gender by area by ethnic origin and then a weighting adjustment by race. In examining survey and 2004 population estimates data, it was found that this method would result in cells in which there were fewer than 20 cases. Thus, the initial post stratification weighting was done in three steps:

- Age by gender by area
- Ethnicity (State)
- Race (State)

The categories used in the weighting adjustments were:

Area	Addison
	Bennington
	Caledonia
	Chittenden
	Franklin and Grand Isle
	Lamoille
	Orange
	Orleans and Essex
	Rutland
	Washington
	Windham
	Windsor

Age	0-9
	10-17
	18-24
	25-34
	35-49
	50-64
	65+

Gender	Female
	Male
Ethnic Origin	Hispanic
	Non-Hispanic
Race (based on primary race)	White
	African American
	Asian
	Other Race

The initial post stratification weighting applied to the data set was age within gender within area. This initial post stratification weight adjusted the survey data to match the population counts by age cohort and gender within each of these areas. An adjustment factor was calculated within each area by age by gender cell:

$$\text{Adj(AS)} = \text{AS}(\text{area} - \text{census} - \text{actual}) / \text{AS}(\text{area} - \text{survey})$$

Where:

- Adj(AS) was the age cohort by gender weighting adjustment within each area
- AS (area – actual) was the actual population within a specific area by age cohort by gender cell
- AS (area – survey) was the weighted survey counts within a specific area by age cohort by gender cell (weighted by final family weight)

Adjustments were made to this initial person level weight to adjust for the actual number of Vermonters by race and ethnic origin (two separate weighting adjustments). Since the application of any weighting adjustment to the initial person level weight may cause the age/gender/area survey counts to vary, a process called raking was utilized. That is, once the ethnic origin and race weighting adjustments were applied, the survey counts of age by gender by area did not match the actual population counts. The raking process alternates making weighting adjustments by variables for which there are only marginal counts (for example weighting by age/gender/area and then by ethnicity/area) by making alternating adjustments. Thus, the initial person level weight was adjusted by ethnic origin and in a separate adjustment by race. Then, this new weight was adjusted by age/gender/area so it again matched the demographic profile of Vermont by these characteristics. This weight was then adjusted to match the ethnic origin counts for the state and then the counts by race category for the state. The post stratification weighting process was repeated until the weighting adjustments converged and the weighted counts matched the state demographic profile by age/gender area, ethnic origin, and race.

Post Stratification Weighting Adjustments for Enrollment in Medicaid and Other State Sponsored Programs

An issue that is common in all studies that try to measure health insurance coverage is that the population enrolled in Medicaid and other state sponsored health insurance programs is generally undercounted. There are a number of reasons that might account for this, such as a greater difficulty in reaching these populations given their lower incomes and reluctance among some respondents to report enrollment in such programs. This is often referred to as a response driven by social desirability. Among many people there may be a sense of embarrassment associated with enrollment in a state sponsored health program. Another aspect is confusion of state sponsored insurance programs with Medicare or private insurance. Survey design elements were incorporated to identify cases where there was potential confusion.

In order to determine the potential for an undercount of Medicaid in the survey data, an analysis was undertaken using available administrative data on program enrollees. Based on administrative data a total of 129,940 Vermont residents were enrolled in Medicaid, VHAP, or Dr. Dynasaur. After post stratification weighting, the survey estimate of the population enrolled in Medicaid or another state sponsored health insurance program was approximately 102,000 Vermonters. This represents an undercount of 21%.

Given this undercount, post stratification weighting adjustments were recalculated to adjust for the undercount of enrollees in Medicaid and other state sponsored health insurance programs. These adjustments were based on the number of enrollees calculated from the administrative records. A post-stratification weighting adjustment was made by program by age by gender to correct for this undercount. The adjustments were made at the state level. The weighting cells are defined in Table 8.

Table 8. Weighting Cells for Medicaid Weighting Adjustments

Program	Gender	Age Categories
Medicaid	Male	0-18, 19-34, 45-64, 65+
Medicaid	Female	0-18, 19-34, 45-64, 65+
Dr. Dynasaur	Male	0-9, 10-17
Dr. Dynasaur	Female	0-9, 10-17
VHAP	Male	18-34, 45-64, 65+
VHAP	Female	18-34, 45-64, 65+

The adjustments were also made within county areas. Table 9 summarizes the geographic regions used in the Medicaid adjusted weights

Table 9. Medicaid Weighting Regions

County	Region				
	Northeast VT	Northwest VT	Mideast VT	Midwest VT	Southern VT
Addison				X	
Bennington					X
Caledonia	X				
Chittenden		X			
Essex	X				
Franklin		X			
Grand Isle		X			
Lamoille	X				
Orange			X		
Orleans	X				
Rutland				X	
Washington			X		
Windham					X
Windsor					X

The final post stratification weight was then recalculated to include an adjustment for enrollment in a state sponsored health insurance program as a fourth step.

- Age by gender by area
- Ethnic origin by gender by age (entire state)
- Race by gender by age (entire state)
- State Medicaid program enrollment by gender by age by area

Adjustments were made to the initial age by gender by area person level weight to adjust for the actual number of Vermonters by race and ethnic origin (two separate weighting adjustments). Then adjustments were made by enrollment in Medicaid or other state sponsored health insurance programs. Raking again was utilized until the weighting adjustment converged.

NOTE: Medicaid adjustments were made to the data for the GPS component only.

Post Stratification Weighting Adjustments for the Households with One or More Uninsured Residents

As indicated, the same basic strategies were used to weight the uninsured over sample as were used in weighting the GPS component of the data. The same system of base weights and non-response weighting adjustments were applied to the data prior to post stratification weighting.

In the case of post stratification weighting, the goal was to weight the data to allow results to be generalized not only to the uninsured population of Vermont but also to match the characteristics of all households with at least one uninsured resident. Since a household may contain both uninsured and insured residents, weighting adjustment were made to all residents in households with one or more uninsured residents. This was done to allow analysis of household and family level characteristics of those Vermont residences with one or more uninsured residents.

The first step in the post stratification weighting process was to develop population based estimates of uninsured residents and others residing in their households. These population estimates were derived by using the GPS component of the data to develop a profile of households with at least one uninsured resident and the demographic characteristics of all residents in these households (regardless of whether or not they had health insurance). An analysis was conducted to determine estimates of population counts for households with one or more uninsured residents based upon health insurance coverage, gender, age, race, ethnicity and the county of residence. These counts were then used as estimates of the population in developing final post stratification weighting adjustments.

The second step was to identify all households in the data set with at least one uninsured residents. This includes households whose data was gathered during the GPS component of the research as well as the uninsured over sample. All such households were flagged in preparation for the development of the post stratification weights.

In order to develop weighting cells in which there were at least 20 cases, it was necessary to collapse some of the areas used when weighting the GPS data. Otherwise, the weighting cells and process remained the same. The uninsured household post stratification weighting was done in three steps:

- Insurance coverage by age by gender by area
- Ethnicity (State)
- Race (State)

The categories used in uninsured household weighting adjustments were:

Health Insurance Coverage	Covered by Health Insurance
	Uninsured

Area	Bennington and Windham Counties
	Rutland and Windsor Counties
	Caledonia, Orleans, and Essex Counties
	Chittenden County
	Franklin, Grand Isle, and Lamoille Counties
	Addison, Orange, and Washington Counties

Age	0-9
	10-17
	18-24
	25-34
	35-49
	50-64
	65+

Gender	Female
	Male

Ethnic Origin	Hispanic
	Non-Hispanic

Race (based on primary race)	White
	African American
	Asian
	Other Race

The initial uninsured post stratification weighting adjustment applied to the data set was insurance coverage within age within gender within area. This initial post stratification weight adjusted the survey data to match the population counts by age cohort and gender within each of these areas and by whether the person was covered by some type of health insurance. An adjustment factor was calculated within each cell:

$$\text{Adj(AS)} = \text{AS(GPSest - actual)} / \text{AS(GPSest - survey)}$$

Where:

- Adj(AS) was the age cohort by gender by health insurance coverage weighting adjustment within each area.
- AS (GPSest – actual) was the actual population within a specific area by age cohort by gender by health insurance coverage cell based on estimates derived from an analysis of the weighted GPS data.
- AS (GPSest – survey) was the weighted uninsured household survey counts within a specific area by age cohort by gender by health insurance coverage cell (weighted by final family weight applied only to households with at least one uninsured resident).

Adjustments were made to this initial person level weight to account for the actual number of residents (in households with at least one uninsured resident) by race and ethnic origin (in two separate weighting adjustments). Since the application of any weighting adjustment to the initial person level weight may cause the age/gender/insurance coverage/area survey counts to vary, a process called raking was utilized. That is, once the ethnic origin and race weighting adjustments were applied, the survey counts of age by gender by health insurance coverage by area did not match the actual population counts. The raking process alternates making weighting adjustments by variables for which there are only marginal counts (for example weighting by age/gender/health insurance coverage/area and then by ethnicity and then by race) by making alternating adjustments. Thus, the initial person level weight was adjusted by ethnic origin and in a separate adjustment by race. Then, this new weight was adjusted by age/gender/insurance coverage/area so it again matched the demographic profile of households with at least one uninsured resident. This weight was then adjusted to again match the ethnic origin and race characteristics of these households. The post stratification process was repeated until the weighting adjustments converged and the weighted counts matched the state demographic profile by age/gender/insurance coverage/area, ethnic origin, and race.

As noted, Medicaid enrollment weighting adjustments were not made to the uninsured household data.

Population Size Reflected in the Final Data Set

The weighted data set is designed to provide data that can be generalized to the population of Vermont and to allow statements to be made about the state as a whole as well as for various sub-populations with a known standard error and confidence. The population size reflected in the final data set is the total population of Vermont, or 621,394 residents.

Population Size Reflected in the Uninsured Population Data Set

The weighted data set is designed to provide data that can be generalized to the uninsured population of Vermont and to allow statements to be made about the uninsured residents as well as the household in which they reside. The population size reflected in the final data set is the estimated uninsured population of Vermont, or 61,057 uninsured residents. The final data set also contains data on residents with insurance that reside in households with at least one uninsured member.

X. Precision

The determination of precision in surveys of this nature is more involved because of the combination of a general population component as well as an over sample component. Another factor is that data was gathered from all household members rather than a randomly selected household member (a cluster sample design). The sampling approach introduces design effects into the survey process that must be taken into account when calculating the final sampling errors for the study. The design effect can be thought of as the impact of the sample design in terms of the departure from what would be expected from simple random sample of the same size. The two-phase design of the sample introduces a design effect because the probabilities of selection are not the same in the general population sample and the over sample. That is, there was a greater probability that a household in the over sample would be selected than a household in the general population sample. The second component of the design effect arises from the rostered nature of the data collection process. That is, the data collection process relied on contacting households and obtaining information about all household members rather than contacting households and gathering data about only one randomly selected household member. This is referred to as the design effect due to intracluster correlation. The reason for the effect is that members of the same household share a number of similar characteristics. For example, all members of a family are likely either all insured or not insured.

In order to accurately report sampling error, it is important to incorporate the overall design effect into sampling error calculations. The standard formula for calculating sampling error is derived by assigning a confidence level to the standard error (for a proportion), typically 95%. At 95%, the sampling error is considered to be the standard error multiplied by 1.96:

$$\text{Sampling Error (95\% confidence)} = \pm 1.96 * \sqrt{(p * (1-p)) / n}$$

Where p is the observed proportion in the sample and n is the number of completed surveys. In calculating sampling error, p is always set to 50%, which results in the most conservative measure of sampling error. In the case of the 2005 Vermont Health Insurance Survey, the sampling error calculations were adjusted by the design effect:

$$\text{Sampling Error (95\% confidence)} = \pm 1.96 * \sqrt{[(p * (1-p)) / n] * \text{deff}}$$

Where deff is the product of the design effect due to stratification, the design effect due to intracluster correlation, and the design effect of the oversample.

Table 10 provides a summary of the sampling errors for the state as a whole, which is derived from the GPS component of the research and is based on 4,010 interviews (9,754 residents). A separate sampling error is reported for the uninsured residents. In this case this is based on 1,550 uninsured respondents interviewed during both the GPS component of the research and the uninsured over sample. All reported sampling errors include design effect adjustments.

Table 10. Precision for the 2005 Vermont Health Insurance Survey

Area/Group	Precision (+/-)
State (GPS Survey)	1.2%
Uninsured Residents	1.3%

Analysis of the 2005 Vermont Health Insurance Survey Data

Point estimates from the data set can be obtained using any standard analytical program. Point estimates represent elements such as the percentage of respondents answering “yes” to a question or the mean value. The nature of the survey and sampling process, however, does introduce design effects into the process that typical analytical programs do not take into account. The design effect of the sampling process has an effect on estimates that involve variance, such as the calculation of confidence intervals. In order to calculate these statistics accurately, the use of software specifically designed to analyze data gathered using complex sampling methodologies is required. Market Decisions uses SUDAAN software from RTI International for all analysis of correlated and weighted data. The use of analytical packages not tailored to analysis of correlated data will underestimate variance of the sample and provide misleading survey results.

In this study, sample design effects arise from three sources:

- Sample stratification (in this case general population and uninsured over sample strata)
- Intracluster correlation (and cluster based sampling in general)
- Oversampling of sub-populations

Sample stratification impacts the probability of selection of a given household. That is, the probability that a household would be selected differed between the general population component and the uninsured over sample component of this research.

The intracluster correlation is the relationship between those within a Primary Sampling Unit (PSU). In this case, the PSU is the household (within each household data was collected on each member). The intracluster correlation impacts precision because of the similarity of members within a PSU. That is, members of the same household will tend to have attributes in common, such as where they live, whether they are covered by insurance and so on. In analytical terms it means that the “value” gained by adding another person within the same household is less than that of adding a randomly selected individual from among the entire population.

Finally, over sampling impacts precision because certain groups are preferentially selected over others. The overall design effect is the product of these factors.

Estimates of precision reported in table 8 were calculated using SUDAAN. These estimates are based on the standard errors at 95% confidence and take into account the design effects of the sampling process.

To aid in future analysis of this data, a sample SUDAAN program is included below. The key components for analysis (in SUDAAN) are:

1. Setting the design effect parameter as DEFT1 (stratification, clustering, and over sample).
2. For the nest command,
 - The strata variable is FILE (General Population or Over Sample) – ***you will need to include this when running analysis on the uninsured population. It is not needed when running analysis using only the GPS data set.***
 - The PSU variable is CSID (household ID).
3. For the weight, the person level weight is finalmwt1 when analyzing data from the GPS data set. When analyzing data from the uninsured household data set you will need to use the weighting variable finaluwt1.
4. Use design effects estimator to account for sample stratification, clustering, and over sample.

```
PROC CROSSTAB DATA="TYPE IN FILENAME HERE"
FILETYPE=SPECIFY TYPE OF FILE
DESIGN=WR DEFT1;
NEST (FILE) CSID;
WEIGHT Finalmwt1;
SUBGROUP gend01x;
LEVELS 2;
TABLES gend01x;
SETENV PAGESIZE=55 LINESIZE=78 COLWIDTH=8 DECWIDTH=4 LABWIDTH=20
LEFTMGN=2 COLSPCE=1;
PRINT NSUM="SAMPLE SIZE"
      WSUM="WEIGHTED SIZE"
      COLPER="Column Percent"
      TOTPER="Total Percent"
      SECOL="SE Col %"
      SETOT="SE Tot %"
      DEFFCOL="DE COL"
/STYLE=NCHS
NSUMFMT=F6.0 WSUMFMT=F8.0;
```

Appendices

Appendix 1. Mail Screening Survey



Vermont . . .

**Department of Banking, Insurance, Securities
and Health Care Administration
Division of Health Care Administration**

Dear Vermont Resident:

I am writing to ask for your participation in an important study about health insurance in Vermont.

The Division of Health Care Administration is doing this survey to gain information about how many Vermonters have health insurance. We also want to gather more information about Vermonters who do not have health insurance. Health Care Administration regulates and monitors many parts of Vermont's health care system including insurance companies.

Market Decisions is conducting this important health insurance survey under a contract with the state of Vermont. **Your information is strictly confidential.** Your answers will be combined with those of others so that you cannot be identified. Your participation in this study is important since you represent other Vermonters in your community.

There are a few questions about your household on the back of this letter. This information will help us understand more about the health insurance coverage in Vermont. We would appreciate it if the person who is most familiar with health insurance coverage in your household would answer these questions. Please place this letter with your answers in the business reply mail envelope and return it to Market Decisions. The postage is already paid.

We may contact your household by telephone to ask additional questions. This will be within the next few weeks. Again, we would greatly appreciate your participation in this important study.

If you want to verify or learn more about the study you can contact Dr. Brian Robertson of Market Decisions at 1-800-293-1538 ext. 102 or Dian Kahn of the Vermont Division of Health Care Administration at 802-828-2906

Thank you again for your assistance with this important study.

Please answer these few questions about your household by placing a check or “x” in the appropriate box.

1. Is this a private residence where **SOMEONE** lives at least 6 months of the year?

YES NO

Please think about the definition of health insurance and types of health insurance programs described below before answering the next question.

Health insurance includes any program or insurance plan that pays any part of hospital or doctor bills. This includes Medicare, Medicaid, VHAP, Dr. Dynasaur, and insurance you might get through an employer or pay for yourself. This includes insurance offered by Blue Cross Blue Shield, MVP, CIGNA or other companies. It also includes health insurance you may receive through the military or as a veteran.

Medicare is a NATIONAL health insurance program for people 65 years and older and for certain people with disabilities.

Medicaid pays for medical care for adults with lower incomes who are caretakers for children, are aged, or have a disability. **Dr. Dynasaur** is a state program that pays for medical care for children under 18 or pregnant women. **VHAP or The Vermont Health Access Plan** pays for medical care for adults with incomes below a certain level. Enrollees in any of these programs may have green AIM or gold insurance cards.

People in your household may be covered by different types of health insurance.

2. Is **EVERYONE** who lives in this household covered by some type of health insurance? Please include yourself, other adults and any children living in the household.

	Yes, Everyone who lives in this household has some type of health insurance.
	Some people in this household are covered by health insurance but there are others who are not.
	No one in this household has health insurance.

3. If there is **ANYONE** in the household that you are unsure whether or not they have health insurance please check the box below:

	I am unsure about the health insurance coverage of someone in this household.
--	---

Please place this letter in the envelope provided and mail to:
Market Decisions, P.O. Box 2890, South Portland, ME 04116

Appendix 2. Survey Pretest Report



Survey Pretest

Vermont Division of Health

Vermont Health Insurance Survey

Prepared for:

Vermont Division of Health

November 2005

Prepared by:

Brian Robertson, Ph.D.

Director of Research

brianr@marketdecisions.com

Vermont Health Insurance Survey Pretest Survey Results

The pretest survey involved conducting live interviews with randomly selected Vermont residents to evaluate the functioning of the survey instrument. The respondents were asked to go through the survey and answer the questions. In addition, respondents were asked to provide feedback during the course of the survey. This was done to determine if there were questions that did not make sense to the respondent and if there were terms they did not understand or found confusing. Respondents were asked to bring up any issues they had during the course of the survey by telling the interviewer. The interviewer then entered the comments from the respondent verbatim. These comments were added at the specific question where the respondent raised a concern or had a comment. This allowed our staff to view the comment in context; that is, our staff could evaluate the specific question about which the comment was made. Following the survey, the respondents were then debriefed about the survey to get their overall feeling about the research. Interviewers again entered any comments provided by the respondent verbatim. After completing the interview, the interviewer would then write a brief summary of the survey from their perspective, especially noting instances where they felt the respondent was confused, instances where there was some hesitancy in answering survey questions, the general tone of the respondent, and the respondent's reaction to the survey.

The respondent notes from the pretest survey as well as the comments provided by interviewers were compiled and reviewed by Dr. Robertson.

The pretest of the 2005 Vermont Health Insurance was begun on November 2, 2005 and was concluded on November 6, 2005. A total of 60 pretest interviews were conducted with Vermont residents. These interviews included

- 17 interviews with households with children
- 15 interviews with households in which at least one resident that was 65 and older and a Medicare beneficiary
- 10 interviews with households in which at least one person was uninsured

Pretest households ranged in size from 1 to 6 persons. On average, the pretest survey required 24 minutes to administer.

Bases on the pretest, the survey instrument functioned as expected. The feedback from respondents and interviewers indicates there was a need only for some minor wording adjustments to some questions and the need for expanded definitions in one instance.

Of the 60 households that were interviewed, 47 agreed to allow the state to re-contact them in the future – indicating a high level of resident interest in the topic.

Respondents and Interviewer Comments

HCB12

Over the last 12 months, did you or anyone in your family have to change their life significantly in order to pay medical bills?

Comment: does the question include changing life to include health insurance: if so yes, but if not no.

An interesting comment but I think this is appropriate.

HCB13

Have any of the following happened to your family because of medical bills? You or others in the household were...

Comment: are you talking about the last 12 months, last 12 months the answer is no, if it is prior to that years or so the answer is yes

I think we should add a time qualifier: During the past 12 months...

HSTAT02

In general, FILL NAME 's health is

Comment: would you say Jennifer's health in general is would be a better way to say it

Interviewer indicated this was more of a statement than a question. I would recommend using the more traditional wording:

Would you say FILL NAME 's health, in general, is

INC04

When thinking about this family income, who would you consider to be the PRIMARY wage earner. That is, which person in the household contributes most to the family's total income?

Comment: our wages are 50/50 and there should be a spot for that.

Respondent indicated that couples income was equally 50/50. This is perhaps a rare case but the survey might need to allow for more than one "primary" wage earner in some instances.

INSP12

Does FILL NAME 's health insurance plan cover **all or** at least some of the cost of prescription drugs?

Need to remove “all or” to clean up grammar.

INSV01

Is anyone now covered by an insurance plan that pays for routine vision care including regular eye exams?

Comment: Does that mean separate from regular health insurance?

For this question we'll need to add a prompt: This includes any coverage you may have through your health insurance or through a separate plan.

RX03

The Healthy Vermonters Program provides Vermonters who do not have pharmacy insurance with access to prescription drugs for the same price as is paid by Medicaid.

Does FILL NAME receive any support for drug costs through the Healthy Vermonters program?

Comment: i don't understand that one

says that on card, not sure if it is healthy Vermonters program, restate. i do a get discount on one of my medications but it's not

Comment: one of the state ones that listed, in the new Medicare book.

For this question there seems to be some lack of understanding or awareness of the program – we should include a more expanded definition of the program – who it is offered to, who provides the program, etc.

RXU02

Which category best represents the amount that FILL NAME pays per month for prescription drugs that FILL NAME uses on a regular basis?

Comment: Respondent suggests that we clearly define how people pay - is this asking what people pay out of pocket or does this include what insurance covers

Comment: hard to answer that because he gets most of his drugs from the VA

For this question, since most people will not know the actual cost of the prescription medication they pay for, it should be qualified to amount that is paid out of pocket.

General Comments

Comment:	<i>I didn't feel too bad about it, too personal about debts and income and such, but other than that okay.</i>
Comment:	Might have a difficult time asking the health questions (i.e. mental health/depression) if people don't have insurance they might not reveal them, also income levels, lower incomes may not want to answer
Comment:	Harvard Group Health plan available through the state or from the state for elder folks, retirees, etc it's affordable for older people
Comment:	<i>I think you guys are pretty nose (debt and assets) - I answered more questions than I thought I was going to.</i>
Comment:	I just wish that my epilepsy wasn't the way it is, they are trying to get things to go better for us, they [the state] are trying to make it easier for insurance to cover things - it is hard because my husband makes just a little too much money, we fall into the wrong category for more assistance and don't even qualify for food stamps - I try not to worry because worrying isn't good for my condition, but I do worry
Comment:	the only thing i would say, some things are not black and white for example are you worried that about not having health insurance in the next 12 months, I'm not real worried but if my husband lost my job i might be worried-qualify the question how worried are you rather than are you worried yes or no. other than that it was fine
Comment:	I think it went well, I think we covered everything.
Comment:	about 19 yrs ago when Dr Dinosaur didn't exist we desperately needed insurance for one of our children and there was nothing available then, even called the governors hotline ... Would like to see programs that ensure everyone gets the health insurance they need.
Comment:	when I was taken off of vhap.... there seems like there should be something income based. I would be willing to pay something for that other insurance. it should be income based like, if you make this much you pay this much...these other insurances are so high. it is costing us 180/week for insurance. it is so high.
Comment:	aside from the couple questions that I noted as kind of confusing... what i like was as opposed to asking specific questions and getting specific answers there were choices, as far as confidentiality that felt good. Also I liked the option to use labels rather than first names.
Comment:	No specific comments, I don't think too clearly...
Comment:	nothing in detail but I'm guessing that you're hoping to help people that are close to low income levels and not accessing insurance programs that if you are trying reach them, if you weren't used to surveying, its leaning more to those that are more articulate to these issues then those that aren't. the people that know the terms, know there income, know there assets and debts. I don't know how you would replace that, it would probably be a design of the questions because it sounded like you were listing at least four or five options for some of the questions, that sounded like a lot. maybe like for the income breakdown, like the big chunks first, then go to a smaller increment rather then ask for a direct estimate.
Comment:	they might it is possible talk about barriers for getting insurance like in daughters boyfriend he didn't hand in papers on time there is no advocate for anyone to get insurance human resources should be little more of an advocate about health resources instead of being a machine i really thing the survey was good very important issue
Comment:	<i>I feel this should not be asked it is no one business (respondent comment in regards to debt and asset questions)</i>

Other Questionnaire Issues

Medicare Prescription Drug Follow-up Questions (MCRX01 – MCRX05)

The intent of these questions is to assess awareness, familiarity, and interest in the new Medicare prescription drug benefit. As currently structured, these questions are asked if there is anyone in the household that is 65 and older. The issue that has occurred is when the respondent is not 65 and older. In this case, the questions are not directly relevant to the respondent (though they may be relevant to others in the household).

In cases where the respondent is under 65, their awareness and familiarity is not necessarily topical since they are not technically eligible for the program. Further, MCRX04 – which asks about enrollment in the program, would not be applicable to anyone under age 65.

While it would mean losing some data in cases where the household has a person over 65 but where the person answering the questions is not – I would recommend only asking these questions if the respondent is over 65.

RX03

The Healthy Vermonters Program provides Vermonters who do not have pharmacy insurance with access to prescription drugs for the same price as is paid by Medicaid.

Does FILL NAME receive any support for drug costs through the Healthy Vermonters program?

Currently this is asked of those 65 and older. The questions will be re-worded and asked of all household members:

RX03

The Healthy Vermonters Program provides Vermonters who do not have pharmacy insurance with access to prescription drugs for the same price as is paid by Medicaid.

Does anyone in the household receive any support for drug costs through the Healthy Vermonters program?

[IF YES: ASK: WHO IN YOUR FAMILY IS ENROLLED IN HEALTHY VERMONTERS PROGRAM? SELECT ALL RESPONSES]

PROMPT: If you are interested in finding about this program, you can get more information by calling 1-800-250-8427

10
11
12
13
14
15
16
17

96 NO MORE

97 NO ONE IS ENROLLED IN HEALTHY VERMONTERS

98 DK

99 REF

Pretest Introductory Script

Hello, I'm _____ calling for the Vermont Division of Health Care Administration.

We are doing an important study to help the state evaluate the health insurance coverage and health insurance needs of Vermont residents.

We will be surveying over 4000 households in the state of Vermont over the next several months, but first we need to pre-test the study with 30 households to evaluate the proper functioning of the survey instrument, and to get your feedback on it from the respondent's point of view. Your participation will count for a lot in helping shape this important study that the state will use to evaluate the health insurance and healthcare needs of Vermont residents.

The study will take about 20 minutes, depending on the size of your household. All of the information you provide will be kept confidential. Your answers will be combined with those of others. This survey may also be recorded but that is just to evaluate my performance if that is okay with you.

Will you help us?-

IF YES:

Great, I really appreciate your help. I'm going to start at the beginning of the survey. I'd like you to answer the questions for your household but also let me know if you have any comments or suggestions about the questions as we go along. For instance, you might find some questions confusing or not understand some terms that I use. I'll be taking notes on these and we'll also take some time at the end of the survey to discuss them and get any other feedback you have. Do you have any questions before we begin? Okay, then let's get started...

IF ASKED:

Your telephone number was randomly generated by a computer program.

The study is being sponsored by Vermont Division of Health Care Administration.

If you would like to find out more about our study, you can call Dr. Brian Robertson of Market Decisions at 1-800-293-1538 ext 102. or Dian Kahn of the Vermont Division of Health Care Administration.

TAKING NOTES:

Use Control-N to make notes of questions or comments in the survey question they relate to. Do the same at the end of the survey after you've got their final feedback.

Appendix 3. Weighting Plan

Vermont Division of Health Care Administration Weighting Plan

The survey data is weighted to adjust for non-response and also to match the state profile based upon sex, age, Hispanic origin, race, and area of residence. Weighting is also adjusted for households with multiple phone lines and for interruptions in phone service. A final weighting adjustment was made (in the GPS data) to adjust for an undercount in the Medicaid population in the data set.

The weighting procedure involved three primary phases: Probabilistic base weights, non-response weighting adjustments and post stratification weighting adjustments. Probabilistic base weights and non-response weighting adjustments were made at the state level for the GPS component of the survey (since this sample was drawn as a separate stratum). Separate probabilistic and non-response weighting adjustments were calculated for the uninsured over sample. Weighting was handled sequentially by weighting household level data and person level data.

The basic weighting strategies outlined below were applied to both the GPS survey component and the uninsured over sample through the set of non-response weighting adjustments. Note that the probabilistic and non-response weights were calculated separately for the GPS and uninsured household components of the research.

There were some differences in the approach used in the post stratification weighting of the survey data components. For the GPS component, 2004 population estimates from the Census Bureau were used to calculate population counts for the demographic characteristics used in weighting.

These 2004 estimates were NOT used in weighting of the uninsured population sample since they do not accurately reflect the population of the uninsured. Note that the uninsured population in this report refers to both uninsured residents that were interviewed during the GPS component of the research and the uninsured over sample component of the research. In order to derive the most accurate population counts for key demographic characteristics for the uninsured (and for insured residents who also resided in these households), Market Decisions relied on the GPS survey results to estimate these population characteristics.

The same basic set of variables was used in post stratification weighting for both the GPS and uninsured population (and insured residents residing in households with one or more uninsured residents): age, gender, area, race and ethnicity. The one exception was a post stratification adjustment for enrollment in Medicaid or other state sponsored health insurance programs. A post stratification adjustment to correct for an undercount of this population was made in the GPS weights, but no such adjustment was made in weighting the uninsured population.

Base Sample Weight

An initial sample weight was assigned to each record in the sample file regardless of the final outcome of the case. This base weight was equal to the inverse of the probability of selecting a number within the specific sampling frame. The base weights will be calculated separately for the GPS component and the uninsured over sample. This process allocated the probability of selection from all sample records to the final set of completed cases. Our GENESYS sampling software provides several sample characteristics. Of interest here is the total number of phone numbers within the specified sampling frame. This total number represents the maximum number of RDD telephone numbers that could be generated within the specific sampling frame. For our sampling strata, this represents the maximum number of RDD telephone numbers in the GPS survey component and separately the uninsured over sample.

The sampling base weight $BW_{samp}(ci)$ for the i th sampled phone number from sample c is calculated as the inverse of the probability of selection or:

$$BW_{samp}(ci) = \frac{N(c)}{n(c)}$$

where $N(c)$ is the total number **of telephone numbers** in the hundred-number blocks that have at least one listed number in sample c and $n(c)$ is the total number of telephone numbers sampled for sample c before GENESYS ID pre-screening of nonworking, nonresidential telephone numbers. The base weight is based upon all RDD telephone numbers included in the final sampling frame, not just those cases identified as residential.

Initial Classification of Case Records

The non-response weighting adjustments relied on weighting sample records based upon the eligibility status assigned to the telephone number. The eligibility status tracked the number resolution in terms of identification of eligible households and completing interviews with these households. The eligibility status of a sample record ranged from undetermined (no information had been obtained that would help determine if this was even a residence) through interview completion (completed interview with data on all household members). The table below provides a summary of the eligibility status codes:

Eligibility Classes Used in Non-response Weighting

Eligibility Class Code (ELIGRESP)	Eligibility Class Description
1	Completed Interview – All
2	Completed Interview – Exclusions not completed
3	Eligible Household, Non-interview, Family Formation Completed
4	Eligible Household, Non-interview, Family Formation Not Completed
5	Working Residential Number – Ineligible
6	Working Residential Number – Undetermined Eligibility
7	Ineligible – Business, Institution/Non-working Number
8	Undetermined

In the initial sample file, all sample records were assigned an eligibility status code of 8. As additional information was gained about the case, the eligibility status was changed to reflect this information. It is important to note that eligibility status is determined by evaluating all call attempts to the telephone number and not simply the last attempt.

The eligibility status is determined by the call dispositions assigned to the sample record. A table of potential disposition codes for this research is presented below along with their assigned eligibility status. The eligibility status of 11 of these dispositions can vary depending on the overall call history. For example, the last disposition of a case may have been a hang-up (which would have an assigned eligibility of 8), but if earlier calls had determined that this was a residential telephone number, its eligibility status would be 6. The sample file recorded both the disposition codes and eligibility status to allow assignment of these particular cases to their actual eligibility status.

Final Disposition Codes and Eligibility Status

Final Disposition Code	Eligibility Class Code (ELIGRESP)
Complete	1
Complete With Exclusions	2
Partially Complete - Terminated Interview	3
Partially Complete (Unresolved Callback)	3
No One 18 or Older	5
Not a Vermont Residence	5
Vacation Residence	5
All Household Members Covered by Health Insurance (<u>Uninsured Over Sample Only</u>)	5
Call Blocking/Screening	6
Business	7
Disconnected Phone	7
Fast Busy	7
Fax/Modem	7
Group Quarters/Institution	7
No Ring	7
Number not in Service	7
Pager/Cell	7
Telephone Number changed	7
Temp Out of Service	7
Call Intercept	8
Answering Machine	VARIES
Busy	VARIES
Contact only	VARIES
Hard Refusal	VARIES
Hang-up	VARIES
Infirm	VARIES
Language Barrier	VARIES
Not Available in Time Frame	VARIES
No Answer	VARIES
Scheduled Callback	VARIES
Soft Refusal	VARIES

Non-Response Weighting Adjustments

1. Working Residential Status Non-response Adjustment Factor

The working residential status non-response adjustment is defined as follows:

- For records where residential status was determined (designated by ELIGRESP=1,2,3,4,5,6 or 7), the working residential status non-response adjustment $ADJ_{res}(ci)$ for record i in sample c is defined as:

$$ADJ_{res}(ci) = \frac{\sum_{i=1}^{n_c} BW_{smp}(ci)}{\sum_{i=1}^{n_c} \delta_{resdet}(ci) BW_{smp}(ci)}$$

where $BW_{smp}(ci)$ is the sampling base weight for record i in sample c , n_c is the number of records in sample c , $\delta_{resdet}(ci)$ is equal to 1 for cases where residential status was determined (ELIGRESP=1,2,3,4,5,6, or 7) and 0 otherwise.

- For records of unknown residential status (designated by ELIGRESP=8), the residential status non-response adjustment $ADJ_{res}(ci)$ for record i in sample c is defined as:

$$ADJ_{res}(ci) = 0.$$

The first non-response adjusted weight $W_1(c)$ is then calculated as the product of the initial sampling base weight and the residential non-response adjustment factor as follows:

$$W_1(ci) = BW_{smp}(ci) \times ADJ_{res}(ci)$$

Following this step, all telephone numbers where residential status is unknown (ELIGRESP=8) will have adjusted weights of zero. Note that this adjustment assumes that the same proportion of residences that were identified in the known residential categories (ELIGRESP=1-7) occurs in the unknown residential category (ELIGRESP=8).

2. Eligible Residence Non-response Adjustment Factor

The second step in data collection was to identify whether or not the residence was eligible for an interview. For the adjustment, a response was considered to have been obtained for the i th residence from the c th sample when it was determined whether the residence was an eligible⁵ household. Thus, non-response at this stage implies that one has determined residential status but did not get through the screening questions to determine eligibility.

The eligible residence non-response adjustment adjusts the sampling weights of records for which eligibility was determined to account for those sampled cases for which eligibility could not be determined. The eligible residence non-response adjustment is then defined as follows:

- For records where household eligibility was determined (designated by ELIGRESP=1,2,3,4 or 5), the eligible residence non-response adjustment $ADJ_{eligres}(ci)$ for record i in sample c is defined as:

$$ADJ_{eligres}(ci) = \frac{\sum_{i=1}^{n_c} \delta_{workres} W_1(ci)}{\sum_{i=1}^{n_c} \delta_{eligHH}(ci) W_1(ci)}$$

where $W_1(ci)$ is the working residential status adjusted weight for record i in sample c , n_c is the number of records in sample c , $\delta_{workres}(ci)$ is equal to 1 for working residential numbers (ELIGRESP=1,2,3,4,5, or 6), $\delta_{eligHH}(ci)$ is equal to 1 for eligible households (ELIGRESP=1,2,3,4 or 5) and 0 otherwise.

- For records of unknown household eligibility (designated by ELIGRESP=6 or 8), the eligible residence non-response adjustment $ADJ_{eligres}(ci)$ for record i in sample c is defined as:

$$ADJ_{eligres}(ci) = 0.$$

- For ineligible records that are nonworking/nonresidential numbers (designated by ELIGRESP= 7), eligible residence non-response adjustment factor $ADJ_{eligres}(ci)$ for record i in sample c is defined as:

$$ADJ_{eligres}(ci) = 1.$$

⁵ An “eligible residence” for this study is a residential household located in Vermont which has an adult present and is not group quarters, a group home (with 9+ members), institution, or hospital. Also excluded were non-permanent residences and vacation homes.

The second non-response adjusted weight $W_2(c)$ is then calculated as the product of the first non-response adjusted weight and the eligible residence non-response adjustment factor as follows:

$$W_2(ci) = W_1(ci) \times ADJ_{eligres}(ci)$$

Following this step, all residences with unknown eligibility (ELIGRESP=6) and all numbers for which residential status is unknown (ELIGRESP=8) will have adjusted weights of zero. Note that this adjustment assumes that the same proportion of eligible residences that were identified in the eligible residence categories (ELIGRESP=1-5) occurs in the unknown residential category (ELIGRESP=6).

3. Weighting Adjustment for Households with Multiple Telephone Numbers

This adjustment should be performed after the first two non-response adjustments are computed. The reason for this is that this information is collected after determining the eligibility of the household, yet prior to collecting data on the family unit formation. The first two weighting adjustments are thus not affected by this adjustment, while the next adjustments are impacted.

This adjustment converts the sample of telephone numbers to a sample of households. The adjustment factor accounts for the fact that households with more than one residential telephone number had a greater chance of selection than those that did not. Households with multiple phone numbers are given lower weights, since these households had multiple chances of being selected.

The weighting adjustment factor for record i in sample c is defined as:

$$ADJ_{multel}(ci) = \frac{1}{n_{HHtel}(ci)}$$

where $n_{HHtel}(ci)$ is the number of telephone numbers on which the household could receive personal calls. In order to avoid large variations resulting from a few household with many telephones, weighting is done using three categories: 1, 2, and 3 or more residential telephones.

4. Weighting Adjustment for Telephone Interruption

This adjustment factor attempts to adjust for under-coverage due to an inability to capture households with no telephones in the sample. Households with substantial recent interruptions in telephone service receive higher weights because they are conjectured to represent a class of households with a lower chance of selection than households with no interruption. In addition, these households are assumed to resemble the chronic non-telephone households more closely than do households with no service interruptions. The adjustment factor for record i in sample c is defined as:

$$ADJ_{telint}(ci) = \frac{12}{12 - M_{int}}$$

where M_{int} is the number of months out of the last 12 for which the respondent reported an interruption in telephone service ($M_{int} < 12$). Again to avoid large variations in adjustments, service interruption will be classified by 1,2,3,4,5,and 6 or more months.

5. Telephone-Adjusted Sample Weight

The telephone-adjusted sample weight $W_{tel}(c)$ is then calculated as the product of the second non-response adjusted weight, the adjustment for multiple phone numbers and the adjustment for telephone interruption as follows:

$$W_{tel}(ci) = W_2(ci) \times ADJ_{multel}(ci) \times ADJ_{telint}(ci)$$

This weight represents the probability of selection of a particular household in the sample.

6. Family Unit Formation Non-response Adjustment Factor

The third non-response adjustment factor accounts for non-response to the family unit formation questions. The family unit formation is considered to be complete when a listing of all family members in the household is obtained and the family units are determined. Thus, non-response at this stage indicates that a complete family unit formation was not obtained from the household.

The family unit formation non-response adjustment adjusts the eligible residence non-response adjusted weights to account for data loss from eligible residences that did not complete the family unit formation. The family unit formation non-response adjustment is defined as follows:

- For records where the family unit formation was completed (designated by ELIGRESP=1,2, or 3), the family unit formation non-response adjustment $ADJ_{famunit}(ci)$ for record i in sample c is defined as:

$$ADJ_{famunit}(ci) = \frac{\sum_{i=1}^{n_c} \delta_{eligres}(ci) W_{tel}(ci)}{\sum_{i=1}^{n_c} \delta_{famunitcom}(ci) W_{tel}(ci)}$$

where $W_{tel}(ci)$ is the telephone adjusted sampling weight for record i in sample c , $\delta_{eligres}(ci)$ is equal to 1 for known eligible residential cases (ELIGRESP=1,2,3, or 4) and 0 otherwise, $\delta_{famunitcom}(ci)$ is equal to 1 for eligible households where the family unit formation is complete (ELIGRESP = 1,2,or 3) and 0 otherwise.

- For records where the family unit formation was not completed and records of unknown status (designated by ELIGRESP=4,6,or 8), the family unit formation completion non-response adjustment $ADJ_{famunit}(ci)$ is defined as:

$$ADJ_{famunit}(ci) = 0.$$

- For ineligible records (designated by ELIGRESP= 5 or 7), the family unit formation completion non-response adjustment factor $ADJ_{famunit}(ci)$ is defined as:

$$ADJ_{famunit}(ci) = 1.$$

NOTE: For the uninsured over sample, a large percentage of identified households were ineligible for completion of the survey since all household members were covered by health insurance. The weights for all such households will be set to zero at this point.

The third non-response adjusted weight $W_3(c)$ is then calculated as the product of the telephone adjusted sampling weight $W_{tel}(c)$ and the family unit formation non-response adjustment factor as follows:

$$W_3(ci) = W_{tel}(ci) \times ADJ_{famunit}(ci)$$

7. Questionnaire Completion Non-response Adjustment Factor

The last step in creating household level analysis weights is to adjust for household non-response to the questionnaire. Following the family unit formation, respondents were afforded the opportunity to exclude members of the household if they felt they could not answer survey questions about these individuals. Later attempts were made to follow-up in these cases and gather data. In some cases, it was not possible to complete the survey with such individuals.

The questionnaire completion non-response adjustment adjusts the family unit formation non-response adjusted weights to account for data loss from households that completed the family

unit formation but for which data on all household members was not gathered. The questionnaire completion non-response adjustment is then defined as follows:

- For records where the primary family unit completed a questionnaire (designated by ELIGRESP=1 or 2), the questionnaire completion non-response adjustment $ADJ_{quest}(ci)$ for record i in sample c is defined as:

$$ADJ_{quest}(ci) = \frac{\sum_{i=1}^{n_c} \delta_{famunitresp} W_3(ci)}{\sum_{i=1}^{n_c} \delta_{questresp}(ci) W_3(ci)}$$

where $W_3(ci)$ is the third non-response adjusted weight for record i in sample c , $\delta_{famunitresp}(ci)$ is equal to 1 for cases which responded to the family unit formation (ELIGRESP=1,2, or 3) and 0 otherwise, $\delta_{questresp}(ci)$ is equal to 1 for households where the primary unit responded to the questionnaire (ELIGRESP=1 or 2) and 0 otherwise.

- For records where the primary family unit did not complete the questionnaire (but did complete the family unit formation), eligible records where the family unit formation was not completed, and records of undetermined eligibility (designated by ELIGRESP=3,4,6, or 8), the questionnaire completion non-response adjustment $ADJ_{quest}(ci)$ is defined as:

$$ADJ_{quest}(ci) = 0.$$

- For ineligible records (designated by ELIGRESP= 5 or 7), the questionnaire completion non-response adjustment factor $ADJ_{quest}(ci)$ is defined as:

$$ADJ_{quest}(ci) = 1.$$

Following this step, all eligible households not completing the family unit formation (ELIGRESP=4) and records of unknown status (ELIGRESP = 6 or 8) will have non-response adjusted weights of zero.

8. Final Sample Household Weight

The non-response adjusted final sample household level weight $HHW(ci)$ is then calculated as the product of the third non-response adjusted weight $W_3(c)$ and the questionnaire completion non-response adjustment factor as follows:

$$HHW(ci) = W_3(ci) \times ADJ_{quest}(ci)$$

At this stage, all households where the primary unit completed the questionnaire and ineligible units will have positive household weights.

9. Family Unit Non-response Weight

For this survey, attempts were made to obtain data on all household members regardless of household composition and the number of family units. Thus, the family weight will be equivalent to the household level weight since there is no need for non-response adjustment:

$$FW(cif) = HHW(ci)$$

This represents the final family weight.

$$FFW(ci) = HHW(ci)$$

10. Person Weights

The initial person level weight is equal to the weight for the family. That is, the person weight for person j in family f in record i sample c is:

$$PW(cifj) = FFW(cif)$$

11. Household Member Non-response Adjustment Factor

The next step is to adjust for household member non-response within the household. Following the family unit formation, respondents were afforded the opportunity to exclude members of the household if they felt they could not answer survey questions about these individuals. Later attempts were made to follow-up in these cases and gather data. In some cases, it was not possible to complete the survey with such individuals. Non-response at this stage indicates that data were not obtained for all members of the household. We recommend that the weights of additional household members that did respond to the questionnaire be adjusted to account for the weights of additional household members for which data were not gathered. The adjustment will be done for the general population component of the survey and the over sample separately. It is assumed that people for whom data are not obtained are not likely members of the primary family unit that resides within the residence and thus their family unit will likely differ in some way from the primary family unit that resides in the household. Market Decisions would recommend weighting the data for this non-response to be done within family unit. That is, weighting adjustments for these individuals will be developed based upon all respondents in the same family unit. For example, for those cases in which data was not obtained from a person in family unit 2, weighting adjustments will be made based upon all respondents classified as residing in family unit 2. The household member non-response adjustment is then defined as follows:

- For all completed households (data obtained on all household members) (defined by ELIGIBILITY STATUS=1), the unit non-response adjustment $ADJ_{family}(cif)$ for household f , record i , in sample c is defined as:

$$ADJ_{family}(cif) = 1.$$

- For cases where data was not obtained from all household members (defined by ELIGIBILITY STATUS=2), the unit non-response adjustment $ADJ_{family}(cif)$ for household f , record i , in sample c is defined as:

$$ADJ_{family}(cif) = \frac{\sum_{i=1}^{n_c} \delta_{addfam} FW(cif)}{\sum_{i=1}^{n_c} \delta_{addfamresp}(cif) FW(cif)}$$

where $FW(cif)$ is the initial family weight for record i , family f , in y c , $\delta_{addfam}(cif)$ is equal to 1 for all respondents in a family unit (defined by ELIGIBILITY STATUS=1,2), $\delta_{addfamresp}(cif)$ is equal to 1 for all additional respondents in households (defined by ELIGIBILITY STATUS=2).

This represents the last non-response weight. All household members for which data were obtained will have positive weights while those for whom survey data were not gathered will now have an adjusted weight of 0.

Post Stratification Adjustment

NOTE: Post stratification adjustments were made to the GPS component of the survey first, relying on 2004 population estimates as the basis for the actual population counts. After the final set of post stratification adjustments to the GPS data, the data reflected the actual population of Vermont residents based on age, gender, area, race, ethnicity, and Medicaid enrollment. The final person level weight should be used in all analysis when the goal is to look at statewide results or at sub-populations BESIDES the uninsured.

The weighted GPS data set was then used to provide population estimates for the uninsured population of Vermont and the insured that also resided in households with one or more uninsured residents. The estimates derived from the weighted GPS data set were used in the post stratification weighting adjustments for the uninsured population. The final uninsured person level weight should be used ONLY in analysis of the uninsured population and their households.

GPS Component

The last step in the weighting process was to standardize the person level weights so they sum to state's total of the civilian, non-institutionalized population (for the GPS component of the survey). This adjustment served two purposes: (1) to adjust family level results to person level results, and (2) to restore proportionality among groups of the population that were over- or under-represented in the survey due to differential non-response or representation in the sample frame.

Post stratification adjustments force respondent weight totals to known population totals for specified groups – referred to here as post strata p . The most effective totals for reducing bias in survey estimates will be related to (1) propensity to be under-covered or failure to respond and (2) response to survey variates of interest.

A number of post stratification adjustments were made at this stage. Market Decisions made adjustments based on the following set of demographic characteristics:

- Area of the state
- Gender
- Age
- Race
- Hispanic Origin

In addition, a weighting adjustment was applied to adjust for undercounts in Medicaid or other state health insurance programs (Medicaid, VHAP, Dr. Dynasaur).

The post stratification adjustment was completed through iterative adjustments, sometimes called raking. Raking is used to produce post stratified weights when only the marginal population counts are known (i.e., counts by age and gender may be available by sample but not for age by gender). Before making each adjustment, post strata were defined within sample based on the above variables. Analysis was conducted to insure that each post strata contained at least 20 persons. In cases where 20 persons were not contained within a cell, post strata were collapsed with a neighbor post strata. The list of the post strata used in the weighting process for the GPS component of the survey is provided on pages 27 and 28.

Adjustments were applied sequentially during the weighting process. After adjusting for the first characteristic, the next adjustment is made based on the second characteristic, and so on. For this research, Market Decisions accomplished post stratification weighting adjustments in four iterative steps:

- Region by gender by age
- Race (statewide)
- Hispanic origin (statewide)
- Medicaid undercount adjustment (by age, gender, and area)

For example, assume $ADJ_{st}(j)$ is the j th post stratification adjustment and $PW(j-1)$ is the weight after post stratification adjustment ($j-1$). And, assume C_{jk} is the k th post stratification cell used in making the j th adjustment, then the adjustment for persons in that cell is:

$$ADJ_{st}(j) = \frac{N(C_{jk})}{\sum_{jkl=1}^{n_{jk}} PW(j-1)_{jkl}}$$

where $N(C_{jk})$ is the population count for cell C_{jk} , n_{jk} is the number of persons in cell C_{jk} , and $PW(j-1)$ is the weight after post stratification adjustment ($j-1$). The process of raking is continued until the poststrata totals following the adjustment converge to the known totals.

Final Person Weight

The final person weight $FPW(cifj)$ is then calculated as the product of the initial person weight $PW(cifj)$ and the final standardization adjustment factor as follows:

$$FPW(cifj) = PW(cifj) \times ADJ_{st}(cifj)$$

Uninsured Population Component

The uninsured population component refers to all survey respondents identified as uninsured during both the GPS component of the research as well as the uninsured over sample. It also includes insured residents that reside in a household with one or more uninsured members.

The same general process was used in the post stratification weighting adjustments for the uninsured population as outlined for the GPS component of the survey. In developing post stratification weights for the uninsured population, population estimates of the uninsured and their households were derived from the weighted GPS data set. This provided the most accurate estimates of the actual number of uninsured residents, their demographic characteristics, as well as the characteristics of insured people who also resided in households with the uninsured. The GPS analysis provided:

- The estimated total count of the uninsured population
- The counts of the uninsured population by gender, age, area, race, and ethnicity

In order to allow analysis of the household characteristics of the uninsured, analysis was conducted to determine the characteristics of the insured population that reside in households with one or more uninsured residents. This analysis provided:

- The estimated total count of residents that reside in households with at least one uninsured resident (this included both insured and uninsured residents).
- The proportions between uninsured and the insured residents in such households.
- The counts of the insured population residing in such households by gender, age, area, race, and ethnicity.

The goal of the post stratification adjustments for the uninsured population was to allow:

- Analysis of the uninsured population and to allow analysis of sub-populations within the uninsured population in Vermont
- Analysis of the household characteristics of the uninsured population. That is, to allow analysis not only of the characteristics of the uninsured resident themselves but also the characteristics of their household and all that reside in such households.

The final goal was an important concern to allow an examination of the uninsured population in relation to their overall household structure and the context in which uninsured Vermonters reside within the household and family unit. Examples include analysis of the employment characteristics of adults in households with uninsured children or the characteristics of children that reside in households in which one or both parents are uninsured, among other types of analyses.

The GPS component of this survey provided the best available population estimates for appropriate post stratification weights for the uninsured population component. The GPS data (after appropriate post stratification weighting) provides reliable population characteristics for the uninsured population and insured residents that reside in households with one or more uninsured residents. That is, it provided the most accurate estimates for development of weighting adjustments to allow analysis of the uninsured population of Vermont as well as analysis of their household and family characteristics.

The data from the uninsured over sample as well as households from the GPS component with uninsured residents was normalized to the population profile developed from an analysis of the complete GPS data set. The demographic characteristics used in post stratification adjustments to the uninsured population component of the research were the same as those used in post stratification adjustments to the GPS component of the research. That is:

- Area of the state
- Gender
- Age
- Race
- Hispanic Origin

The same formulas described on page 62 were used in post stratification weighting adjustments to the uninsured population component of this research.

NOTE: The final data set contains two separate weights for analysis.

The first weighting variable is limited to the GPS data. This weight is used when the goal is to make statements about Vermont residents in general and for specific sub-populations based on demographic, health, income, and/or employment characteristics. This weight is also used when deriving estimates for various types of health insurance coverage.

The second weighting variable is limited to analysis of the uninsured and their households. It is only used when analysis is conducted among uninsured residents or in analysis of households in which one or more uninsured residents reside. This weight CAN NOT be used to make general statements about the population of all Vermont residents.

Appendix 4. Determination of Medicaid Eligibility

Vermont Division of Health Care Administration Determination of Medicaid Eligibility

An analysis was conducted to evaluate the characteristics of the uninsured population that might meet current eligibility requirements for enrollment in Medicaid, Dr. Dynasaur, or VHAP. The analysis first determined the percentage of uninsured residents that might meet requirements based upon their age, their family income, family assets, receipt of SSI, whether they had a disability, and prior insurance coverage. The criteria for defining a resident as eligible were provided to Market Decisions by the Department for Children and Families. Eligibility was determined for ANFC related Medicaid, SSI related Medicaid (disables and 65 and older), Dr. Dynasaur, and VHAP. The eligibility criteria for each of these specific programs is outlined below.

Eligibility Criteria for Medicaid, Dr. Dynasaur, and VHAP

ANFC-related Medicaid

Requirements:

- Under age 21 or a parent (or caretaker relative) with a child under age 18 living with the parent (caretaker relative).
- At or below the following income and asset levels:

ANFC-related Medicaid

	Household Size							
	1	2	3	4	5	6	7	8
Income limit outside Chittenden County	1,271	1,271	1,494	1,671	1,871	1,994	2,217	2,394
Income limit inside Chittenden County	1,361	1,361	1,583	1,761	1,961	2,083	2,305	2,483
Asset limit in Vermont	2,000	3,000	3,150	3,300	3,450	3,600	3,750	3,900

SSI-related Medicaid (disabled)

Requirements:

- Disabled
- At or below the following income and asset levels

SSI-related Medicaid -- disabled

	Household size	
	1	2
Income limit	861	1,023
Asset limit in Vermont	2,000	3,000

SSI-related Medicaid (age 65 or older)

Requirements:

- Age 65 or over
- At or below the following income and asset levels:

SSI-related Medicaid -- 65 and over

	Household size	
	1	2
Income limit	1,746	2,070
Asset limit in Vermont	2,000	3,000

Dr. Dynasaur

Requirements:

- Under age 18
- Household income under 300% FPL
- No asset test

VHAP

Requirements:

- Age greater than 18 and less than 65
- Household income \leq 185% of FPL if a parent or caretaker; otherwise \leq 150%.
- No asset test
- Currently have no health insurance

If household income is at or below 75 percent of FPL and within the age ranges, the person is eligible without regard to health insurance status.

Otherwise, assuming that the person is within the age ranges, health insurance status needs to be clarified:

Did person lose insurance providing hospital and physician services at some time within the past 12 months?

If no, the person is eligible and we do not need to go further. If yes

Was the insurance providing hospital and physician services lost in the past 12 months ...

1. Purchased by an individual directly from an insurance company
2. Purchased through an association
3. Provided at least in part by an employer
4. Provided at least in part by a college or university

[If 1 or 2, the individual is not eligible]

[If 4 the person is eligible]

If 3: then:

Was the insurance provided in part by an employer, lost because the employee's hours were reduced?

[If coverage was lost because of a reduction in hours, the individual is not eligible.]

[If coverage was lost for some other reason, the individual is eligible.]

Within this general framework, eligible residents were then grouped into specific programs based on these eligibility requirements.