



---

# COLORADO HOUSEHOLD SURVEY 2008–09

---

## *Methods Report*

---

Colorado Department of Health Care Policy  
and Financing  
1570 Grant Street  
Denver, Colorado 80203-1818  
Contact: Rene Horton, Business Analysis Section  
Manager, [Rene.Horton@state.co.us](mailto:Rene.Horton@state.co.us)

Contact for technical questions on data and  
methods: Anna Furniss, Colorado Health  
Institute, [furnissa@coloradohealthinstitute.org](mailto:furnissa@coloradohealthinstitute.org)

---

December 2009

## Table of Contents

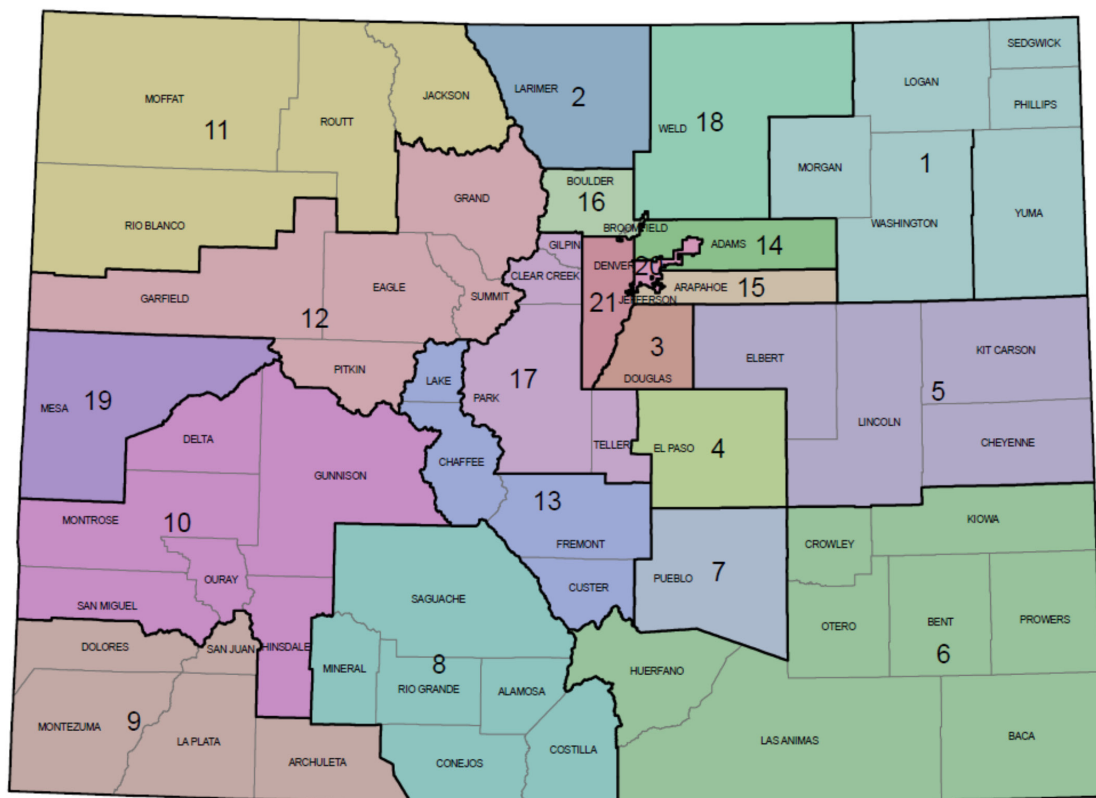
Introduction .....	3
I. Study Design.....	4
1. Sample Stratification.....	6
2. Household-level selection .....	8
3. Individual-level target person selection.....	8
2. Field Preparation, Fielding and Data Processing.....	8
3. Weighting Procedures .....	11
4. Survey response rate.....	16
5. Imputation.....	19

## Introduction

The Colorado Department of Health Care Policy and Financing (HCPF) contracted with the Colorado Health Institute (CHI) and its subcontractor, Social Science Research Solutions (SSRS), to conduct the 2008-09 Colorado Household Survey (COHS). The goal of the COHS is to document health insurance coverage and access to and use of health care for the non-institutionalized population in Colorado. This report provides information about the methods used to collect, clean and document the data in the COHS data files.

The study was conducted for CHI via a random digit dialing (RDD), computer-assisted telephone interview (CATI) by SSRS, an independent research company. Interviews were conducted from November 12<sup>th</sup> 2008 through March 13<sup>th</sup>, 2009 among a representative sample of 10,090 households containing at least one person age 18 and older. Interviews were stratified by 21 health statistics regions (HSRs) to ensure adequate representation within each of these important population aggregations within the State of Colorado.

### Colorado Health Statistics Regions



The Colorado health statistics regions were developed by the Colorado Department of Public Health and Environment (CDPHE) for public health planning purposes. The boundaries for the regions were determined according to the size of the population in each county—counties with smaller populations

were aggregated—and key demographic factors for each county, including the number of communities served by each county health department.

Table I. Completed interviews per HSR region

Region	Interviews
1	398
2	412
3	431
4	819
5	398
6	398
7	403
8	404
9	402
10	406
11	404
12	400
13	398
14	640
15	615
16	394
17	341
18	438
19	396
20	788
21	805
TOTAL	10,090

## I. Study Design

The study employed a dual-frame sampling design that includes a landline and cell phone-only sample. The dual frame design seeks to ensure that complete coverage of all households that own at least one type of phone (approximately 98 percent of all Colorado households are listed in telephone banks or own a cell phone). When the study was commissioned, our best-approximation was that fifteen percent of Colorado households owned only a cell phone. Since then, the CDC has released state-level data of wireless substitution and reported that 16.7 percent of all Colorado households owned only a cell phone in 2007. Of the 10,090 interviews, 400 were conducted with respondents who owned only a cell phone.

To reach cell phone-only respondents, SSRS developed a screening question to screen out persons who used a landline telephone. The cell phone-only sample yielded the terminations and completed interviews noted in Table 2.

Table 2. Final disposition of cell-phone only sample

Disposition	Sample Records	Percent
Completed interview	400	24%
Under 18 Years of age	154	9%
Owens landline	840	51%
Does not live in CO	227	14%
Cell-only but no one under 65*	17	1%
Cell-only but main residence not in Colorado	39	2%
Total completions and terminations	1,647	100%

\*Under 65 screen discussed in the next section

The overall sampling design contained a number of features across a number of dimensions which can be described in terms of sample stratification, household selection criteria and within household selection criteria.

1. Sample stratification
  - Set interview targets per region
  - Set interview targets within selected regions by telephone exchange based on incidence of African American households
  
2. Household-level selection
  - Screening to exclude out-of-state home owners and vacation homes
  - Within the cell phone-only frame, screening excluded landline owners and respondents under 18 years of age
  
3. Individual-level (target) selection
  - Screening to include adults who could answer health insurance questions for each member of the household and a computer-generated, randomly selected “target” person
  - Oversampling to ensure a 10 percent increase of target selection of persons under the age of 18
  - Additional screening out of households without anyone under the age of 65, which began mid-field on January 23, 2009

## **I. SAMPLE STRATIFICATION**

The number of regional interviews was set by the Colorado Health Institute to ensure adequate statistical power within each region. As we will describe later, each region was weighted individually to ensure within-region representation (see Table 1 for interviews completed by region).

Additionally, regions 4, 15 and 20 were further stratified by telephone exchange to maximize the number of African American interviews obtained. These three regions were selected because they are the only regions in Colorado with sufficient numbers of African American households to warrant an attempt at disproportionate stratification of telephone exchanges. Each of these three regions was disproportionately sampled with exchanges of higher incidences of African American households oversampled at the expense of exchanges with low incidence rates (see Table 3 below).

The stratification scheme illustrated in Table 3 was implemented to compensate for the expected bias created by telephone interviewing; that is, the distribution of most sampled populations tends to skew more heavily towards Whites than the general population. As such, the goal was to ensure an adequate sample of African Americans comparable to their proportion in the Colorado population. The proportion of African American households by region is compared to actual response in Table 4.

Table 3. Sample stratification scheme for African American sample

Strata	Overall population	African American population	Percent of all African Americans in region	Allocation of African American interviews / deff	African American interviews	Non-African American population	Percent of all Non-African Americans in region	Allocation of non-African American interviews / deff	Non-African American interviews
<b>Region 20</b>									
Low	403,973	15,289	32.2%	2.2%	2	388,684	72.3%	19.0%	89
Medium	114,854	16,804	35.4%	17.6%	16	98,050	18.2%	30.5%	143
High	66,500	15,417	32.5%	80.2%	73	51,083	9.5%	50.5%	237
<b>Total</b>	<b>585,327</b>	<b>47,510</b>	<b>100.0%</b>	<b>5.55</b>	<b>91</b>	<b>537,817</b>	<b>100.0%</b>	<b>2.88</b>	<b>469</b>
<b>Region 15</b>									
Low	293,382	15,929	33.2%	7.0%	3	277,453	56.8%	15.9%	82
Medium	147,925	16,671	34.8%	23.3%	10	131,254	26.8%	36.2%	187
High	95,486	15,334	32.0%	69.8%	30	80,152	16.4%	47.9%	247
<b>Total</b>	<b>536,793</b>	<b>47,934</b>	<b>100.0%</b>	<b>2.25</b>	<b>43</b>	<b>488,859</b>	<b>100.0%</b>	<b>2.28</b>	<b>516</b>
<b>Region 4</b>									
Low	293,466	11,285	34.0%	12.5%	4	282,181	52.7%	12.4%	81
Medium	170,850	11,683	35.2%	50.0%	16	159,167	29.7%	32.7%	213
High	104,707	10,216	30.8%	37.5%	12	94,491	17.6%	54.9%	358
<b>Total</b>	<b>569,023</b>	<b>33,184</b>	<b>100.0%</b>	<b>1.43</b>	<b>32</b>	<b>535,839</b>	<b>100.0%</b>	<b>2.56</b>	<b>652</b>

Table 4. Incidence of African Americans in three regions relative to completed interviews

	Incidence of African American households	Incidence in survey	Initial target interview completes	Completed interviews
Region 20	8.1%	11.9%	769	819
Region 15	8.9%	7.0%	724	724
Region 4	5.8%	4.3%	782	815

The initial targets were exceeded to ensure that sufficient numbers of African American interviews were completed across the state. SRSS completed 2,358 interviews from these three regions rather than the 2,275 projected in the original scope of work. The Colorado Health Institute requested a minimum of 200 African American interviews be completed overall. In the end, SRSS completed 233 interviews statewide.

## 2. HOUSEHOLD-LEVEL SELECTION

Screening questions excluded anyone living out-of-state or those called at a vacation home. Overall, 6 percent of all households reached were terminated because of an out-of-state status and 1.7 percent were terminated due to the resident being a vacation home. Cell phone-only screening information was provided earlier in the report.

## 3. INDIVIDUAL-LEVEL TARGET PERSON SELECTION

The survey was designed to collect data at the household level as well as the individual level, therefore it was important not only for the respondent to be able to answer questions about each household person's health insurance status, but the respondent also needed to be able to answer the entire battery of questions on the survey for one randomly selected "target" person in the household.

CHI also had a goal of oversampling children in households. Therefore, the probability that the computer would select a child under the age of 18 was increased by 10 percent once the household roster had been established.

In addition, CHI expressed concern that the COHS would have a greater proportion of completes from persons 65 and older than what would be seen in the general population because, in general, RDD telephone surveys have a higher completion rate for individuals 65 and older. Therefore, the number of completed interviews with targets age 65 and older was monitored. As of January 23, 2009, 23 percent of targets were 65 years of age or older. At this point, SSRS began terminating all interviews for which there was no one under the age of 65 in the household. By the end of time in the field, 14 percent of targets were ages 65 and older. This compares to 9.3 percent of Colorado's population in this age cohort.

All of the sampling steps were taken into account during the weighting procedure to correct for the disproportional subsamples created by each step, as will be described in later sections.

## 2. Field Preparation, Fielding and Data Processing

The questionnaire was developed by the Colorado Health Institute, based on questions contained in the 2008 Massachusetts, Oklahoma and Minnesota Household Surveys, which closely followed the State Health



Access Data Assistance Center (SHADAC) model of health interview survey questionnaires. Specific sections were modified for the State of Colorado. Table 5 presents a summary of the questionnaire domains in the survey. As illustrated in the table, the majority of questions were administered to the target household member, with demographics, socio-economic questions and health insurance questions asked for all household members. Additionally, employment questions and employer-based health insurance questions were asked of spouses of targets and for household members under the age of 26 since this younger age group has a higher prevalence of dependency on parents for their health insurance coverage.

**Table 5. Summary of questionnaire domains by respondent type**

Domain	Survey Respondent	All Household Members	Target Household Member	Target's Spouse and/or Parents (Target age<26)
Demographic characteristics (age, race/ethnicity, gender, marital status)	X	X	X	X
Socioeconomics (education and employment status)	X	X	X	X
Length of residency in Colorado			X	
Health insurance coverage	X	X	X	X
Detailed employment questions			X	X
Avail. of employer sponsored insurance			X	X
Health status			X	
Access to and use of health care			X	
Family income			X	
Home ownership	X			
Household telephone status	X			

Prior to going into the field, SSRS programmed the study into a Computer Assisted Telephone Interviewing (CATI) program. Extensive checking of the program was conducted, given the large number of logic patterns that the skip patterns could generate. Household roster surveys with a specific target person require 3-4 times more manual labor to check when compared to a survey design with simply “last birthday” as the target selection criterion because of the complexity of the skip patterns.

The field period for this study was November 12<sup>th</sup> 2008 through March 13<sup>th</sup>, 2009. The interviewing was conducted by International Communications Research (ICR) in conjunction with SSRS in Media, PA. All landline telephone interviews were conducted using the CATI system which ensures that questions follow the logical skip patterns and that listed attributes are automatically rotated to eliminate “question position” bias.

CATI interviewers received both written materials on the survey and formal training. The written materials were provided prior to the beginning of the field period and included:

1. An annotated questionnaire that contained information about the goals of the study as well as detailed explanations of why questions were being asked, the meaning and pronunciation of key terms, potential obstacles to be overcome in getting good answers to questions and respondent problems that could be anticipated ahead of time as well as strategies for addressing them;
2. A list of frequently asked questions and the appropriate responses to those questions;
3. A script to use when leaving messages on answering machines; and,
4. Contact information for project personnel.

Interviewer training was conducted both prior to the study pretest (described below) and immediately before the survey was officially launched. Call center supervisors and interviewers were walked through each question in the questionnaire. Interviewers were given instructions to help them maximize response rates and ensure accurate data collection. Interviewers were instructed to encourage participation by emphasizing the social importance of the project and to reassure respondents that the information they provided was confidential.

A pretest of the survey instrument and procedures was conducted in late October 2008. The sample for the pretest was a basic RDD sample from Colorado landline telephone exchanges. Colorado Health Institute staff listened live with SSRS project managers for approximately two hours and an additional 4 hours of interviewing time was digitally recorded and placed on a secure FTP site for review. Overall, the flow of the survey was good and the respondents remained interested throughout. Interviewers were monitored throughout the field period and were provided feedback, where appropriate, to improve interviewing techniques and clarify survey questions.

SSRS maintained a staff of Spanish-speaking interviewers whom, when contacting a household, were able to offer respondents the option of completing the survey in Spanish or in English. A total of 168 interviews were conducted in Spanish.

SSRS treated this study as a “best practices” study given certain budgetary and methodological directives from the Colorado Health Institute. Fielding of the survey used the following best practice procedures:

- Manually dialed numbers—two weeks into the project, SSRS randomly split 1,000 pieces each of random, unreleased sample, and fielded the first 1,000 using manual dialing, 1,000 using “predictive” computer dialing and 1,000 using “proactive” computer dialing. The results showed a 10 percent decrease in refusal rate in the manual dialing procedure, as suspected. From that point forward, SSRS enacted manual dialing as part of its “SOP” for all “best practice” projects.
- Instituted a call rule of original plus up to 20 callbacks before considering a sampling unit “dead.” All landline samples with no answer, answering machine or busy disposition were dialed to a minimum of 12 calls.
- Varied the time of day and the day of the week when call-backs were placed using a programmed differential call rule.
- Explained the purpose of the study and stating as accurately as possible the expected length of the interview.
- Permitted respondents to set the schedule for a call-back and encouraged them to phone-back on an SSRS 800 number.

- Privacy managers were immediately called back on an open line (CRT systems do not transmit “caller ID” information, so any record that had a disposition for a privacy manager to call back were called back manually on phones that did relay caller ID information).
- Initial refused interviews were “put to bed” for a period of two weeks, when a refusal conversion attempt took place. Second refusals were put to bed for an additional 4 weeks, when a second conversion was attempted.
- The study included a sweepstakes for a pair of Denver Broncos tickets or a \$100 gift certificate to one randomly selected respondent.

Two analytical data files were created from the raw survey data: 1) a person-level file that includes all data elements collected for all persons in the household as well as characteristics of the household; and 2) a target-level file that includes all data elements collected for the target person in the household along with data on the characteristics of the target’s family and household. The Table 5 provides a summary of the variables included in each file. CATI range and logic checks were used to check the data during the data collection process. Additional data checks were implemented as part of the data file development work, checking for consistency across variables and family members, and developing composite measures of family and household characteristics.

### 3. Weighting Procedures

Survey data were weighted to: 1) adjust for the fact that not all survey respondents were selected with the same probability and 2) account for gaps in coverage in the survey frame. Base weights (survey design weights) address the differential sampling rates described earlier in this report. Subsequently, the resulting base weights were post-stratified along several dimensions to reflect the control totals obtained from the 2009 national estimates of the U.S. Census Bureau’s American Community Survey. These counts were indexed by region, gender, education, age, race/ethnicity and home ownership.

In the first weighting stage, SSRS developed design weights to compensate for a range of known biases that occur in telephone interviewing in general and the COHS sample design specifically. These are summarized below:

- *Non-response weight* = Exchange weight x eligibility rate, where the exchange weight equals the number of telephones called /number of telephones available to call and the eligibility weight equals the number of completes /number eligible to be completed.
- *Sub-sampling weight* = Corrections for regions 4, 15, and 20 by race and strata.
- *Post-stratification weight* = Rebalancing completes by region to population counts.
- *Number of persons weight* = Correction for the number of persons in the household (capped at 3 or more).
- *Phone use weight* = Correction for the number of landline telephones used in the household, capped at 3 (3 phones = weight of .33) or more.
- *Age weight* = 18 years and younger down-weighted to rebalance the increase of 10% of being selected as a target.
- *Cell phone-only weight* = 16.7% of the file is cell phone-only
- *Design Weight* = Non-response x sub-stratification x stratification x persons x phones x age x cell phone-only

Each step was normalized so that the sum of weights equaled the un-weighted number of completes. The final post-stratification procedures that followed included:

- *Final Weight* = Design weight with a two-step raking procedure. The first raking occurs at the region level, where targets were set by age, education, gender, race and home ownership by the 21 statistical regions. However, because children 0-18 years of age had disproportionately large weights, the cell phone-only population became inflated to 23.3 percent of the sample population; therefore, a final statewide rake was conducted to reapportion cell phone-only households to 16.7 percent.
- *Trimmed Final Weight* = when weighting complex surveys, such as the COHS, it is not uncommon to have cases with inordinately large or small weights. Often, researchers trim weights so that no one case is disproportionately influential (in the case of a large weight) or inconsequential (small weights) to the estimates generated by the data. For the COHS sample, 11 percent of its weights were in excess of 8 (with a high of 98) or under 0.1. Weights were trimmed to 0.1 and 8 and the residual proportionally reallocated to the weights of the non-trimmed cases.

The final weights were developed using a procedure known as *Iterative Proportional Fitting* (IPF) or “raking” using the statistical software, QBAL. Post-stratification targets were entered for age, race/ethnicity, gender, region, tenure of homeownership and education based on U.S. Census Bureau’s American Community Survey (ACS) estimates. The ACS reports data according to Public Use Microdata Area (PUMA), which is an area that defines the extent of territory for which the Census Bureau tabulates public use microdata sample data. The raking process was carried out at the regional level for which population estimates had to be developed—since the ACS only provides super-PUMA and PUMA designations for in-state geography.

A method for overlaying PUMA population estimates over the 21 statistical regions was developed by CHI. Each PUMA represents a proportion of the population for a certain county in Colorado. A map of PUMA-to-county was obtained from the University of Missouri for all counties in Colorado and a map of county to region was developed in order to calculate PUMA weights for each region. The regional PUMA weights were applied to the ACS data to generate regional population estimates of gender, education, race, etc. Final counts are provided below.

Table 6. Demographic characteristics by 21 statistical regions in Colorado

Region	Gender		Home Ownership		Education			
	Male	Female	Rent	Own	>H.S.	H.S. Diploma	Some College	College degree +
1	50.8%	49.2%	26.4%	73.6%	10.5%	27.7%	24.4%	12.7%
2	50.5%	49.5%	28.1%	71.9%	3.6%	17.3%	27.9%	28.3%
3	50.1%	49.9%	13.8%	86.2%	2.0%	12.7%	22.1%	34.4%
4	49.2%	50.8%	28.4%	71.6%	5.2%	19.9%	24.7%	22.7%
5	50.8%	49.2%	26.4%	73.6%	10.5%	27.7%	24.4%	12.7%
6	50.0%	50.0%	25.3%	74.7%	12.6%	26.6%	23.5%	13.1%
7	48.2%	51.8%	27.6%	72.4%	11.6%	24.1%	24.4%	15.8%
8	49.6%	50.4%	25.0%	75.0%	13.6%	25.9%	22.9%	13.6%
9	50.2%	49.8%	31.0%	69.0%	7.4%	25.2%	23.0%	22.7%
10	51.2%	48.8%	32.2%	67.8%	7.4%	24.2%	22.2%	24.4%
11	52.6%	47.4%	29.7%	70.3%	5.8%	26.6%	23.7%	21.8%
12	54.5%	45.5%	34.7%	65.3%	6.9%	22.0%	20.5%	28.4%
13	50.1%	49.9%	26.7%	73.3%	6.7%	23.8%	24.8%	23.0%
14	50.7%	49.3%	27.4%	72.6%	9.6%	24.5%	22.8%	15.0%
15	49.4%	50.6%	30.5%	69.5%	7.1%	18.8%	22.3%	25.4%
16	50.6%	49.4%	28.8%	71.2%	4.4%	15.8%	22.0%	35.5%
17	50.0%	50.0%	23.1%	76.9%	5.3%	22.3%	24.2%	26.4%
18	50.3%	49.7%	28.3%	71.7%	8.1%	22.8%	26.0%	16.4%
19	48.7%	51.3%	26.0%	74.0%	6.0%	26.8%	25.7%	18.7%
20	50.7%	49.3%	42.1%	57.9%	12.2%	18.1%	17.1%	28.0%
21	49.8%	50.2%	26.5%	73.5%	5.3%	21.2%	25.0%	25.9%

Table 7. Age and race/ethnicity distribution by 21 health statistic areas in Colorado

Region	Age				Race / Ethnicity			
	0 – 17 yrs	18 –34 yrs	35 – 64 yrs	65+ yrs	White	African American	Hispanic	Other
1	24.8%	20.2%	40.5%	14.5%	79.9%	1.0%	17.1%	2.0%
2	22.8%	28.0%	39.2%	10.0%	86.1%	1.2%	10.1%	2.6%
3	28.8%	21.7%	44.1%	5.4%	87.2%	1.8%	7.3%	3.7%
4	27.5%	22.5%	40.6%	9.4%	75.3%	7.6%	13.3%	3.8%
5	24.8%	20.2%	40.5%	14.5%	79.9%	1.0%	17.1%	2.0%
6	24.3%	21.2%	39.7%	14.8%	64.9%	1.0%	32.5%	1.6%
7	24.0%	23.5%	37.0%	15.5%	56.2%	2.4%	39.3%	2.1%
8	24.0%	21.9%	39.3%	14.7%	57.2%	1.0%	40.5%	1.4%
9	21.7%	22.2%	41.5%	14.6%	81.9%	0.2%	12.8%	5.2%
10	21.8%	23.3%	42.1%	12.8%	81.7%	0.3%	13.6%	4.4%
11	22.2%	21.7%	46.4%	9.8%	84.9%	0.4%	13.8%	1.0%
12	22.2%	26.0%	45.2%	6.6%	82.4%	0.6%	16.0%	1.0%
13	21.7%	20.3%	43.7%	14.3%	90.1%	0.2%	7.7%	2.0%
14	28.1%	25.8%	38.4%	7.7%	60.4%	3.4%	32.7%	3.5%
15	26.4%	22.4%	41.6%	9.5%	66.1%	9.7%	18.7%	5.5%
16	22.3%	25.9%	43.9%	7.9%	81.4%	0.6%	13.6%	4.4%
17	21.8%	17.5%	47.0%	13.7%	91.7%	0.3%	5.6%	2.3%
18	26.7%	28.2%	37.3%	7.9%	70.4%	0.5%	27.1%	2.0%
19	22.7%	24.3%	37.7%	15.2%	84.5%	0.7%	13.1%	1.7%
20	24.6%	24.4%	41.1%	9.9%	50.6%	9.6%	35.0%	4.8%
21	22.6%	21.0%	45.5%	10.8%	80.4%	1.6%	14.5%	3.4%

Finally, the full weighting procedure was applied to both the total sample and the landline frame only in order to develop comparisons of estimates and design statistics. A summary of the resulting design statistics, specifically the design effect, bias, variance and mean squared error, is provided in Table 8. These estimates are based on the final untrimmed, full sample weighted estimate of the uninsured rates in Colorado.

Complex survey designs and post-data collection statistical adjustments affect variance estimates and resulting tests of significance and confidence intervals. The impact of the survey design on variance estimates is measured by the design effect, which represents the extent of departure from a simple random sample where all sample units respond. The *design effect* measures the variance inflation of the sample estimate relative to the variance of an estimate based on a hypothetical random sample of the same size. *Bias* measures the systematic deviation of a population estimate compared to the estimate produce from the survey data. The *mean squared error* (MSE) is a statistic that measures overall bias and variance. In practice, the goal is to attain a weight that corrects for as much bias as possible without creating too much variance in the weights. About 11 percent of the final untrimmed weights were in excess of 8 and under 0.1. In order to reduce the variance inflation due to such extreme weights,

trimmed weights were developed. As shown in Table 8, this trimming depresses the uninsured estimate by 0.4 percent. Variance was cut from 3.1 to 1.8 for a significantly lower MSE than the final weight.

It is significant to note the comparison between the full sample and the landline-only frame. A key question that should be considered for all dual frame survey designs is whether the additional frame, in this case the cell phone-only frame, was worth the added expense. This appears to be true for the COHS, at least with regard to MSE, but not with regard to the design effect. The design effect for the final full sample, using the untrimmed weight, is 4.02 compared to 3.56 for the landline-only weight. This increase in the variance, and consequently the design effect, was introduced during the post-stratification weighting when the sample was balanced so that the cell-phone users represented only 16.7 percent of the sample. However, in the end, the added benefit of having a representative younger population in the full sample, via the cell-phone frame, was worth the slight increase in the design effect, and especially because it was off-set by the reduced MSE.

Dual frame designs are preferred when there are known under coverage issues associated with a landline-only sample. For the COHS, the full sample uninsured rate (landline and cell phone-only) was assumed to be the “gold standard” estimate of the uninsured. It was calculated to reduce the landline-only bias and MSE based on the full sample final untrimmed estimate of the uninsured. The landline-only sample produced an uninsurance rate 0.9 percent lower than the full sample estimate, with an MSE equal to that of the full sample. In the final analysis, the full sample weights performed better than the landline-only weights in terms of both bias and variance.

Table 8. Applied weights to the full sample and landline-only sample

Weight	Full Sample				Landline-only Frame			
	deff	Bias*	Variance	MSE	deff	Bias	Variance	MSE
Exchange	1.53	1.8	0.57	3.81	1.71	2.1	0.56	4.97
Eligibility	1.01	2.6	0.01	6.77	2.22	3.3	0.01	10.90
Non-response	1.69	4.2	0.69	18.33	2.67	4.8	0.71	23.75
Sub-stratification	1.68	4.1	0.68	17.49	2.67	5.6	1.22	32.58
Stratification	2.17	4.9	1.17	25.18	3.51	5.5	1.69	31.94
Number of persons	2.64	5.0	1.64	26.64	5.49	5.5	1.69	31.94
Phone user	2.64	5.0	1.64	26.64	1.56	5.5	1.65	31.90
Age	2.60	4.9	1.60	25.61	1.53	5.5	2.05	31.21
Cell phone	2.85	2.6	1.85	8.61	--	--	--	--
Final Region Weight	3.41	-0.2	2.41	2.45	--	--	--	--
Final Weight	4.02	0	3.02	3.02	3.56	0.9	2.56	3.37
Trimmed Final Weight	2.64	0.4	1.64	1.80	2.71	1.3	1.71	3.40

\* Based on an uninsurance rate attained using the final weight, 14.2%

The weighting procedures detailed above were conducted for both the *target file* using the target’s demographic data for post-stratification, and the *person file* using each individual’s demographics as their own target. There were, however, some differences in the procedure used to develop the person file. First, the adjustments for sub-stratification and stratification were made based on number of persons

rather than completed interviews. Secondly, the 'number of persons' adjustment was not made to the person file since each case in the person file represents a person and not a randomly selected household member. Further the age correction in the target file adjusts for the fact that targets 0 to 18 years of age were 10 percent more likely to be randomly selected by the computer than targets 19 and older. This selection procedure was not applied to the person file.

Because of the large differences in the design effect (DEFF) and variance between the final weight and the trimmed weight, the final *trimmed* weight will be used when analyzing the target file data.

#### **4. Survey response rate**

The response rate for this study was 37.5 percent for the landline sample and 27.3 percent for the cell phone sample using AAPOR<sup>1</sup>'s RR3 formula. This translates into an overall response rate of 35.5 percent. Following is a full disposition of the sample selected for this survey.

---

<sup>1</sup> American Association of Public Opinion Research



Table 9. Response rates by 21 health statistics regions

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10	Region 11	Region 12
<b>Eligible, interviewed (Category 1)</b>												
Complete	396	373	391	767	399	397	383	394	395	390	390	387
<b>Eligible, not interviewed (Category 2)</b>												
Refusal	358	310	339	762	321	376	344	415	298	289	278	313
Break off (callback)	72	25	45	114	14	59	52	32	25	42	28	68
Answering machine household	107	113	217	282	73	104	110	96	147	145	49	160
Physically or mentally unable/incompetent	4	6	2	14	4	8	6	7	6	7	10	14
Language problem	22	4	6	37	8	13	3	16	4	7	13	33
<b>Unknown eligibility, not interviewed (Category 3)</b>												
Always busy	14	7	8	31	4	12	10	2	3	7	9	9
No answer	303	228	341	382	201	317	256	367	350	319	122	809
Answering machine	0	0	0	0	0	0	0	0	0	0	0	0
Call blocking	0	2	10	25	0	3	8	1	1	1	0	1
No screener completed	0	0	0	8	0	1	0	1	1	1	0	0
<b>Not eligible (Category 4)</b>												
Fax/data line	177	172	316	381	211	302	184	182	209	203	248	578
Non-working number	2819	1674	1635	4557	1769	2822	2752	3271	2066	1841	2482	3406
Business/government office	241	280	377	554	182	257	225	337	295	291	399	640
No eligible respondent	227	171	155	485	160	257	238	238	201	237	232	413
Quota filled	12	0	0	842	0	0	105	7	5	0	606	257
<b>TOTAL</b>	<b>4751</b>	<b>3365</b>	<b>3842</b>	<b>9241</b>	<b>3345</b>	<b>4927</b>	<b>4675</b>	<b>5366</b>	<b>4005</b>	<b>3780</b>	<b>4865</b>	<b>7087</b>
RR3	38.6%	41.7%	35.4%	36.9%	45.8%	38.7%	40.2%	38.2%	41.2%	40.5%	49.4%	35.1%

Table 9. Response rates by 21 health statistics regions (continued)

	Region 13	Region 14	Region 15	Region 16	Region 17	Region 18	Region 19	Region 20	Region 21
<b>Eligible, Interviewed (Category 1)</b>									
Complete	398	529	685	423	401	387	390	758	663
<b>Eligible, not interviewed (Category 2)</b>									
Refusal	407	372	605	265	286	358	284	605	623
Break off (callback)	32	85	140	52	31	56	55	185	53
Answering machine household	138	227	547	249	137	141	119	718	194
Physically or mentally unable/incompetent	6	6	10	5	4	6	3	10	11
Language problem	7	42	84	16	10	21	8	64	16
<b>Unknown eligibility, not interviewed (Category 3)</b>									
Always busy	2	8	22	13	3	9	12	72	11
No answer	266	469	805	414	329	283	198	1269	639
Answering machine	0	0	0	0	0	0	0	0	0
Call blocking	1	16	31	15	0	2	3	26	5
No screener completed	0	3	13	0	0	1	0	21	1
<b>Not eligible (Category 4)</b>									
Fax/data line	148	291	466	272	187	220	172	653	368
Non-working number	1987	2989	6452	3060	1719	1970	1295	6573	3864
Business/government office	268	423	606	477	256	283	251	1020	510
No eligible respondent	223	230	480	215	178	207	161	510	325
Quota filled	0	96	1	63	33	6	58	2	0
<b>TOTAL</b>	<b>3882</b>	<b>5786</b>	<b>10946</b>	<b>5538</b>	<b>3573</b>	<b>3950</b>	<b>3009</b>	<b>12486</b>	<b>7282</b>
RR3	37.5%	38.4%	30.5%	38.6%	41.9%	37.0%	42.2%	28.8%	38.7%

Table 10. Response rate for landline and cell phone-only samples

	Cell Sample	Total landline and cell phone-only	Total
<b>Eligible, Interviewed (Category 1)</b>			
Complete	400	11,906	12,306
<b>Eligible, not interview (Category 2)</b>			
Refusal	71	10,180	10,251
Break off (callback)	52	1,704	1,756
Answering machine household	0	5,620	5,620
Physically or mentally unable/incompetent	80	177	256
Language problem	243	619	862
<b>Unknown eligibility, non-interview (Category 3)</b>			
Always busy	45	393	438
No answer	4,416	11,123	15,539
Answering machine-don't know if household	2,066	0	2,066
Call blocking	6	233	239
No screener completed	2,605	93	2,698
<b>Not eligible (Category 4)</b>			
Fax/data line	46	7,440	7,486
Non-working number	8396	78,585	86,981
Business, government office, other organizations	1,034	10,352	11,386
No eligible respondent	2,102	7,012	9,113
Quota filled	0	2,938	2,938
<b>TOTAL</b>	<b>21,561</b>	<b>148,374</b>	<b>169,935</b>
RR3	27.3%	36.3%	34.7%

## 5. Imputation

Imputation was used to correct variable non-response due to respondents not answering individual survey items. Any variable with more than 5 percent missing data was imputed through multiple regression imputation software developed for the SAS system (IVWare, University of Michigan, 2002). For ease of use, only single regression imputation was implemented.

The COHS data were generated from 10,090 completed telephone interviews. Consequently, there was a very low rate of variable non-response in the COHS data. Only the “Don’t know” or “Refused” answers to survey questions were considered as variable non-response, or missing data.

Several questions on the COHS were asked only if the respondent answered a preceding question appropriately or had a certain demographic characteristic. For example, questions regarding employment were asked only if a target was 16 years and older. Missing values generated by these skip patterns in the survey instrument were not considered as eligible missing values and were not subject to imputation. However, within a subgroup that could answer a skip pattern question, if the combined “Don’t know” or “Refused” answers accounted for over 5 percent of the data, those values were imputed.

“Don’t know” and “Refused” were separate response options on the survey; however, they were combined for each survey question in order to calculate the percent missing for the item. Table II lists the variables in the COHS data file that had over 5 percent missing values.

Table II. List of Variables for Potential Imputation

Variable Name	Percent Missing	Variable Label	Imputed
INS_VER_TYPE	31.6	H3a. If NO to verification, what type of insurance	no
NOINS_PRIORI	6.6	H9. What was the last type of health insurance the target had	yes
USOC_TYPE_H CTR	7.7	A2a. Specific type of health center used for health care issues	yes
OOP_RX	7.3	A10aa. In past 12 m, out-of-pocket expense for Rx meds (amount)	yes
OOP_OTH	5.9	A10ac. In past 12 m, out-of-pocket expense for other med (amount)	yes
INCOME_2007	17.9	IN3. Total family income for 2007	yes
INC_2007_GRP	7.2	Income based on quartiles of 2007 inc (created)	yes
INCOME_MON	16.4	IN6. Total family income for previous month	yes
INC_MON_GRP	8.4	Income based on quartiles of monthly inc (created)	yes
M_ESI_DEP	7.9	E12A. Target’s mother’s ESI includes coverage for dependents	yes
F_ESI_DEP	6.8	E12A. Target’s father’s ESI includes coverage for dependents	yes

The INS\_VER\_TYPE variable was not imputed because it was part of a two-stage skip pattern question series, which resulted in too few possible respondents answering the question (<20).

Both the income and out-of-pocket expense questions were each asked as a two-part question. In each case, the respondent was first asked to provide the total amount of income and out-of-pocket expenses for the family. If she or he refused, the interviewer prompted the respondent to answer the question based on categorical groupings of income and out-of-pocket expenses. If the respondent answered these questions according to the categories, the information was used to impute a value for total income and total out-of-pocket expenses. If the remaining missing values for income and out-of-pocket expense were still greater than 5 percent, those values were imputed using single regression imputation.