Community Tracking Study

Physician Survey Methodology Report 2004-05 (Round Four)

Statistical Design and Tracing for the Community Tracking Study Physician Survey

Providing Insights that Contribute to Better Health Policy

600 Maryland Avenue, SW
Suite 550
Washington, DC 20024
www.hschange.org

Technical Publication No.

70

October 2006

AUTHORS

Steve Williams, Mathematica Policy Research, Inc.

Frank Potter, Mathematica Policy Research, Inc.

Nuria Diaz-Tena, Mathematica Policy Research, Inc.

Richard Strouse, Center for Studying Health System Change

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
www.mathematica-mpr.com

Center for Studying Health System Change
600 Maryland Avenue, SW, Suite 550
Washington, DC 20024
www.hschange.org

This is one in a series of technical documents that have been done as part of the Community Tracking Study being conducted by the Center for Studying Health System Change (HSC), which is funded primarily by The Robert Wood Johnson Foundation and is affiliated with Mathematica Policy Research, Inc.

Contract No.: 6007-03-004
MPR Reference No.: 6060-170

MATHEMATICA
Policy Research, Inc.

Statistical Design and Tracing for the Community Tracking Study Physician Survey

Final Report

August 31, 2006

Steve Williams
Frank Potter
Nuria Diaz-Tena
Richard Strouse

Submitted to:

The Center for Studying Health System Change
Suite 550, 600 Maryland Ave., SW
Washington, DC 20024

Project Officer:
Jim Reschovski

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
Frank Potter

CONTENTS

CONTENTS *(continued)*

Chapter **Page**

CONTENTS *(continued)*

Chapter **Page**

TABLES

TABLES (*continued)*

Table **Page**

FIGURE

I. OVERVIEW

A. OBJECTIVES OF THE COMMUNITY TRACKING STUDY

The Community Tracking Study (CTS) is the core research effort of the Center for Studying Health System Change (HSC), a nonpartisan policy research organization in Washington, DC, that is funded by the Robert Wood Johnson Foundation (RWJF). HSC's mission is to inform health care decision makers about changes in the health care system at the local and national levels, as well as about how such changes will affect people. HSC conducts surveys of those affected by changes in the health care system—households, physicians, and employers—and interviews with health care leaders in 12 communities.

The focus on markets is central to the design of the CTS. Understanding market changes requires studying local markets, including their culture and history, and public policies relating to health care. To track change across the United States, a random sample of 60 nationally representative communities (stratified by region, community size, and whether metropolitan or nonmetropolitan) was selected in 1996 and this sample formed the foundation for subsequent surveys (see Table I.1).[1]

The CTS examines 12 of the 60 communities in depth by conducting site visits and using survey samples large enough to draw conclusions about health system change in each community. The 12 communities make up a randomly selected subset of sites that are metropolitan areas with more than 200,000 people (as of July 1992). We refer to these as *high-intensity* sites.

[1]The CTS covers the contiguous 48 states and the District of Columbia. Alaska and Hawaii are not part of the study.

TABLE I.1

SITES SELECTED FOR THE COMMUNITY TRACKING STUDY

High-Intensity Sites[a]	Low-Intensity Sites[a]		
Metropolitan Areas >200,000 Population[b]	Metropolitan Areas >200,000 Population[b]	Metropolitan Areas <200,000 Population[b]	Nonmetropolitan Areas
01-Boston MA	13-Atlanta GA	49-Dothan AL	52-West Central Alabama
02-Cleveland OH	14-Augusta GA/SC	50-Terre Haute IN	53-Central Arkansas
03-Greenville SC	15-Baltimore MD	51-Wilmington NC	54-Northern Georgia
04-Indianapolis IN	16-Bridgeport CT		55-Northeastern Illinois
05-Lansing MI	17-Chicago IL		56-Northeastern Indiana
06-Little Rock AR	18-Columbus OH		57-Eastern Maine
07-Miami FL	19-Denver CO		58-Eastern North Carolina
08-Newark NJ	20-Detroit MI		59-Northern Utah
09-Orange County CA	21-Greensboro NC		60-Northwestern Washington
10-Phoenix AZ	22-Houston TX		
11-Seattle WA	23-Huntington WV/KY/OH		
12-Syracuse NY	24-Killeen TX		
	25-Knoxville TN		
	26-Las Vegas NV/AZ		
	27-Los Angeles CA		
	28-Middlesex NJ		
	29-Milwaukee WI		
	30-Minneapolis MN/WI		
	31-Modesto CA		
	32-Nassau NY		
	33-New York City NY		
	34-Philadelphia PA/NJ		
	35-Pittsburgh PA		
	36-Portland OR/WA		
	37-Riverside CA		
	38-Rochester NY		
	39-San Antonio TX		
	40-San Francisco CA		
	41-Santa Rosa CA		
	42-Shreveport LA		
	43-St. Louis MO/IL		
	44-Tampa FL		
	45-Tulsa OK		
	46-Washington DC/MD/VA		
	47-West Palm Beach FL		
	48-Worcester MA		

NOTE: Numbers correspond to coding of the site identification variable in the survey.

[a]The designation "High" and "Low" intensity sites pertains only to survey rounds one through three, not to Round Four. The definition of the sites and the selection procedures are in Metcalf et al. (1996).

[b]Based on estimated population counts for 1992 from the Bureau of the Census.

B. ANALYTIC COMPONENTS OF THE CTS

The CTS has qualitative and quantitative components, which we describe here:

- ***Site Visits.*** To examine the forces affecting health care organizations and how these organizations are responding, researchers interview 40 to 60 health care leaders in each of the 12 high-intensity sites. HSC conducts and manages the site visits, with help from outside researchers.

- ***Household Survey.*** The main focus of this survey is assessing whether consumer access to the health care system is increasing or declining. We surveyed about 60,000 people in 33,000 families for each of the first three rounds of the survey. For Round Four (2003), the sample was made up of about 47,000 people in 25,000 families. Areas of inquiry include access, satisfaction, use of services, and insurance coverage. The survey also collects information on health status and sociodemographic characteristics. To enhance the reliability of information on health plans, we obtain selected information on plan characteristics from linked surveys of insurers. HSC provides technical direction and oversight, and Mathematica Policy Research, Inc. (MPR) is responsible for sample design, data collection, sample weights, and variance estimation for the household and followback surveys. MPR and Social and Scientific Systems Inc. (SSS) collaborated with HSC to prepare the documentation for the public and restricted use files.

- ***Employer Survey.*** For the first round of the CTS (1996–1997), 22,000 public and private employers were interviewed to understand how the American population can access the health system nationally and locally. These employers, which span size and industry sector, were asked about the choice of plans they offer, how much their employees contribute to paying for their coverage, whether they participate in a purchasing alliance, and whether they provide high-quality information to their employees. HSC collaborated with RAND on the employer survey, which was not conducted for subsequent rounds.

- ***Physician Survey.*** For each round, a sample of practicing physicians across the country offers perspective on how health care delivery is changing. More than 12,000 physicians were interviewed for each of the first three rounds, but the number of interviews was reduced to approximately 6,600 physicians for Round Four to reduce survey costs. For Round Four, the sample was allocated across the 60 sites differently than the prior rounds to offset the reduction in sample size. The reduction of 40 percent in the sample size resulted in less than a 10 percent decrease in precision for key estimates at the national level. Physicians respond to questions on whether they can provide needed services for patients, how they are compensated, what effect care management strategies have on their practices, and their practice arrangements. MPR was responsible for the sample design, sample weights, variance estimation, and tracing of physicians who could not be located and the Gallup Organization conducted the interviewing for the physician survey. MPR and SSS collaborated with HSC to prepare the documentation for the public and restricted use files.

Additional background on CTS is available at HSC's website (www.hschange.com).

C. THE ROUND FOUR PHYSICIAN SURVEY

This report describes the design and conduct of the fourth round of the physician survey. The survey was completed by telephone, through computer-assisted telephone interviewing (CATI). MPR developed the sample frame by combining lists of physicians from the American Medical Association and the American Osteopathic Association. Interviews were completed with 12,385 physicians in Round One, 12,304 in Round Two, and 12,406 in Round Three. The sample size was reduced for the Round Four Survey, and 6,628 interviews were completed.

Reports describing the first three rounds of the physician survey are included in Technical Publications #9, #32, and #38 (www.hschange.com). In this report, we discuss the design of the Round Four sample, including sample size reductions (Chapter II), survey design and preparation (Chapter III), data collection (Chapter IV), and sample weighting (Chapter V). The survey instrument, advance materials mailed to physicians, and cognitive interviewing protocols are shown in Appendix A. Additional detail on the equations used to compute the weights is included in Appendix B, and an analysis of nonresponse in Appendix C.

II. SAMPLE DESIGN

For the first three rounds of the Physician Survey, interviews were administered to a stratified random sample of physicians in the 60 CTS sites and to an independent national sample of physicians, referred to as the "national supplement." (For a discussion of the sample designs used on prior rounds, see Technical Publications 09, 32, and 38 available at www.hschange.org). To reduce the cost of the Round Four 2004-2005 Physician Survey, we eliminated the national supplement and reduced the sample size. To compensate for the reduced sample size, the sample was re-allocated among the 60 sites to obtain a more efficient proportional national sample of physicians[1]. Finally, we reduced the extent of oversampling of primary care physicians (PCPs) to achieve approximately equal samples of PCPs and specialists. Otherwise, the design of the Round Four sample was similar to prior rounds, retaining the same sites in the sample for a nationally representative 60-site sample design, to facilitate the estimation of changes across the survey rounds.

Because the CTS Physician Survey has a longitudinal component (physicians sampled for Round Three were oversampled in Round Four), survey precision is affected by the amount of sample overlap between successive rounds. Therefore, a key design decision for each round has been the amount of overlap between rounds. In addition, there are differences between sample frame and interview classifications of physicians as PCPs or specialists and between the two classifications of physicians' practice location. Procedures developed in prior rounds for

[1]In the first three rounds, target sample sizes were assigned to each CTS site to support site-level estimates (approximately 400 physicians in each of the twelve high-intensity sites and approximately 100 physicians in each of the other 48 sites). In Round Four, the target sample sizes for each site were assigned in approximate proportion to the weighted number of physicians in the site. The allocation of the target sample size is statistically more efficient (smaller sample size can obtain comparable standard errors for national estimates by reducing the variation in the sampling weights) than the allocation for the prior rounds.

identifying and adjusting for errors in specialty assignment and geographic misclassification were applied in the Round Four sample selection.

In the following sections, we describe:

- Site selection;
- The target population;
- Key design issues, including our approach to sample overlap, specialty assignment, and geographic misclassification, and sample design changes for Round Four;
- Stratification; and
- Sample selection procedures.

A. SITE SELECTION

The primary goal of the CTS is to track health system change and its effects on people at the local level. Determining which communities (sites) to study was therefore the first step in designing the CTS sample. Sites were selected in 1996 and held constant for the duration of the CTS. Site selection involved three activities: (1) defining sites, (2) determining how many would be studied, and (3) selecting the sites. Metcalf et al. (1996) provides additional detail on these activities (see Technical Report 01 available at www.hschange.org).

1. Definition of Sites

The sites were intended to encompass the range of existing local health care markets. Although these markets have no set boundaries, the intent was to define areas such that residents used health care providers located predominantly in the same area, and providers served mostly area residents. To this end, sites were defined to be the Office of Management and Budget Metropolitan Statistical Areas (MSAs) or, in the case of nonmetropolitan sites, to be Bureau of Economic Analysis economic areas (BEAEAs).

2. Number of Sites

The next step in creating the site sample was to determine the number of high-intensity sites. The trade-offs between the cost of conducting case studies and surveys and the research benefits of a large sample of sites were considered. The research benefits include a greater ability to examine empirically the relationship between system change and its effect on care delivery and consumers and increased "generalizability" of the study findings to the nation as a whole.

Despite the cost advantages of conducting intensive case studies in fewer sites, focusing on a smaller number of communities would have made it more difficult to distinguish between changes of general importance and changes or characteristics unique to a community. Solving this problem by increasing the number of case study sites would have increased the cost of data collection and analysis prohibitively. Therefore, 12 sites were selected from the 48 large metropolitan sites for intensive study and combined with the remaining sample of 48 sites that would be studied less intensively. The 60 sites are primary sampling units (PSUs) and form the *site sample* (see Table I.1 in Chapter I).

3. Site Selection

After the number of sites for the sample was determined, the next step was to select the actual sites. Sixty sites were selected at the first stage of sampling. The sites were stratified geographically by region within three metropolitan status strata and then selected randomly, with probability proportional to their July 1992 population. Hence, The CTS sites (PSUs) were selected independently in three strata, based on metropolitan status and size:

1. MSAs with 200,000 or more people (large MSAs)[2]

[2]Some sites were defined as primary metropolitan statistical areas (PMSAs) or consolidated metropolitan statistical areas (CMSAs).

2. MSAs with fewer than 200,000 people (small MSAs)
3. Nonmetropolitan areas

In each of these strata, CTS sites were selected with probability proportional to the size of the civilian population (as of July 1992). For eight sites in the large MSA stratum, the population was sufficiently large that the site was selected with certainty. These eight sites were Boston (MA portion); Philadelphia, PA-NJ PMSA; Washington/Hagerstown PMSA; New York City; Detroit, MI PMSA; Chicago/Kenosha/Kankakee PMSA; Houston-Galveston-Brazoria, TX CMSA; and Los Angeles-Long Beach, CA PMSA. A ninth site (Baltimore, MD PMSA) was selected with certainty in the sample to complete coverage of the major cities of the Northeast Corridor.

In addition to the nine certainty selections, 39 sites were selected with probability proportional to their population size, using a sequential selection algorithm with selection controlled by geographic region. This allocation ensured that (1) all MSAs had a chance to be selected, (2) larger MSAs had a greater chance than smaller MSAs of being selected, and (3) the site sample would have an approximately proportional allocation across geographic regions.

For the small MSAs, three sites were selected with probability proportional to size, again using a sequential selection algorithm controlled by geographic region. For the nonmetropolitan areas, the first stage of selection was the state.[3] Again, the sequential selection algorithm (controlled again by geographic region) was used to select nine states with probability proportional to the size of their nonmetropolitan population. Based on nonmetropolitan county

[3]Washington, DC, and New Jersey were excluded from this stratum because they do not have any nonmetropolitan areas. Alaska and Hawaii were excluded by the CTS study design.

groups used by the BEA, one county group was selected within each of these states with probability proportional to the population in these county groups.

Of the 60 sites in the CTS sample, 48 were selected in large MSAs, 3 in small MSAs, and 9 in nonmetropolitan areas. The 12 high-intensity sites were randomly selected from the 48 large MSA sites. The site sample can be used to make national estimates, and to create site-level estimates for use as covariates in models that require market-level variables. However, site samples in Round Four are too small for most site-level analyses as a result of sample size reductions.

B. TARGET POPULATION

The target population was physicians who had to have completed their medical training, practice within the 48 contiguous states and the District of Columbia, provide direct patient care for at least 20 hours per week, and were not federal employees. To meet the initial eligibility criteria for sampling, we used information provided on the AMA Masterfile (which includes both AMA members and nonmembers) and on the AOA membership file (files were received in January 2004).[4] Residents, interns, and fellows were considered to be still in training and were excluded from the sample frame. The direct patient care criterion resulted in the exclusion of inactive physicians and physicians who were not office- or hospital-based (such as teachers, administrators, and researchers). The following types of physicians were designated as ineligible for this survey and were removed from the frame:

[4]The AMA Masterfile includes licensed allopathic physicians and osteopathic physicians who obtained graduate training in allopathic medical schools or were identified on state licensing boards. The AOA membership file includes graduates of osteopathic medical schools. In addition, the AOA file often has, for osteopathic physicians, current addresses that may not be on the AMA Masterfile. The AMA Masterfile contains approximately 85 percent of osteopathic physicians listed in the AOA file.

- Specialists in fields that do not focus primarily on direct patient care; Tables II.1 and II.2 list the specialties excluded from the frame.
- Federal employees
- Graduates of foreign medical schools who are licensed to practice in the United States only temporarily

Eligible physicians were then classified as either PCP or specialist. PCPs were defined as physicians with a primary specialty of family practice, general practice, general internal medicine, internal medicine/pediatrics, or general pediatrics. All others with survey-eligible specialties were classified as specialists.

The interviewer verified physician eligibility before continuing with the survey. The attributes that were verified during the interview included whether the physician (1) was a federal employee, (2) was a resident or fellow, and (3) provided patient care for less than 20 hours a week. Physicians who were eligible based on the AMA or AOA Masterfile data, but were ineligible at the time of the interview, were classified as ineligible and were excluded from further data collection.

C. DESIGN ISSUES

The key design issue for Round Four was to meet a cost constraint by reducing sample size, while achieving the best possible precision for national estimates and meeting minimal precision targets for site-level estimates for use as covariates in models that require market-level estimates. Other features of the design developed in prior rounds were retained. First, we explain how we chose the amount of sample overlap between surveys. Because this study has a longitudinal component, survey precision is influenced by the amount of sample (respondents) overlap across survey rounds. Since physician specialty and practice location could be defined differently in the sample frame (AMA and AOA files) and in the interview, we discuss procedures for identifying

TABLE II.1

SPECIALTIES EXCLUDED FROM THE AMA FILES

Allergy and Immunology/Clinical Laboratory	Epidemiology	Pain Management
Aerospace	Forensic Pathology	Pathology
Anatomic/Clinical Pathology	Forensic Psychiatry	Pediatric Anesthesiology
Anesthesiology	Hematology/Pathology	Pediatric Radiology
Bloodbanking/Transfusion Medicine	Musculoskeletal Radiology	Public Health and General Preventive Medicine
Chemical Pathology	Medical Management	Radiology
Clinical Biochemical Gene	Medical Microbiology	Underseas Medicine
Clinical Pharmacology	Medical Toxicology	Vascular and Interventional
Cytopathology	Neuropathology	Radiology
	Neuroradiology	
	Nuclear Medicine	

TABLE II.2

SPECIALTIES EXCLUDED FROM THE AOA FILES

Allergy/Diagnostic Lab Immunology	Forensic Psychiatry	Epidemiology
Anatomic/Clinical Pathology	Hematology/Pathology	Public Health
Anesthesiology/Pain Management	Neuroradiology	Radiation Oncology
Bloodbanking/Transfusion Medicine	Nuclear Medicine	Vascular and Interventional
Clinical Pathology Dermatopathology	Nuclear Radiology	Radiology
Forensic Pathology	Pediatric Anesthesiology	
	Aerospace Medicine	
	Preventive Medicine	

and adjusting for errors in specialty assignment and geographic misclassification in the sample design in the next two sections.[5] In the final section, we discuss options that were considered to reduce sample sizes and how we arrived at the sample allocation for Round Four.

1. Sample Overlap

A common feature of surveys with a longitudinal component is the selection of sampling units (in this case, physicians) in one round of a survey for participation in the next round. Precision may be increased, perhaps substantially, for change estimates and, to a lesser extent, for cross-sectional estimates when a portion of the physicians who responded to Round Three is included in the Round Four sample. At the same time, some proportion of the Round Three sample should be replaced to represent physicians who had no chance of being selected in prior rounds to ensure complete population coverage in Round Four and to minimize respondent burden and conditioning (because repeated contacts may influence survey responses).

We considered several factors when determining the optimum level of sample replacement, including coverage bias, the precision of cross-sectional and change estimates, and possible correlations between rounds that will improve survey estimates. Our analysis based on costs and response rates in Rounds Two and Three implied that a reinterview rate of 60 to 70 percent is advantageous both for cost and for precision reasons (see Technical Report 38). (The reinterview rate is defined as the percentage of physicians who responded in the prior round who responded again in the current round.) Based on an expected eligibility and response rate for reinterviewed physicians of 67 percent, the sample overlap for Round Three was set at near 100 percent for

[5]For some physicians, only a home address was available in the AMA or the AOA file, but the practice location was important for analytic purposes.

Round Two completed interviews and 80 percent of the Round Two noninterviews.[6] A substantial portion of the overlap sample was contained in the high-intensity sites in Round Three, but with the reduced sample allocation to most of those sites in Round Four, many of the Round Three completed interviews could not be used for the overlap sample.

Sampling rates for prior round interviews selected for Round Four changed because of the reduction in total sample size and the re-allocation of the sample. While the Round Three low intensity sites had approximately the same sampling rates in Round Four for specialists, other rates were substantially lower. The average sampling rate for PCPs in Round Four was only 45 percent of the Round Three completed interviews and 30 percent of the Round Three noninterviews (see Table II.3). Obviously the same sampling rates would not have achieved the required sample reductions or the representation of new physicians. On the other hand, the expected proportion of Round Four completed interviews that are reinterviews is still approximately two-thirds, near optimum. This provides a reinterview sample sufficient for longitudinal analyses and improved precision for most estimates. In the next section, we discuss the benefits and drawbacks of increasing the degree of overlap between rounds and discuss briefly how we arrived at the optimum level of overlap.

a. Benefits and Drawbacks of Increasing Overlap

Increasing the degree of sample overlap between rounds increases the precision of change estimates. The optimal overlap for estimates of change for any variables with positive correlations between rounds is 100 percent; however, the potential for gains in precision depends on the degree of correlation between rounds. Increasing the overlap too much can lead to

[6]Noninterviews include physicians in the sample who could not be located, who refused or who were ineligible at the time of data collection.

TABLE II.3

ROUND FOUR SAMPLING RATES AND REINTERVIEW PERCENT OF NOMINAL SAMPLE

Design Component	Influencing Design Change	Round Four Sampling Rates		Round Four Sample: Reinterview Percent Expected
		Percent Of Round Three Interviews	Percent of Round Three Noninterviews[a]	
All Physicians	Reduced Sample, Especially in High Intensity Sites; Equal Allocation for PCP/Specialists	54	34	65
PCP	Reduced Sample, Equal Allocation for PCP/Specialists	45	30	67
Specialists	Reduced Sample, Equal Allocation for PCP/Specialists	70	45	62
Round Three High Intensity Sites				
PCP	Reduced sample	21	14	70
Specialist	Reduced sample	38	24	75
Round Three Low Intensity Sites				
PCP	Equal allocation for PCP and specialists	66	44	67
Specialist	Minimal	100	66	59

[a]Noninterviews include physicians in the Round Three sample who could not be located, who refused or who were ineligible at the time of data collection.

coverage bias for cross-sectional estimates (that is the cross-sectional estimates would be based on less than 100 percent of the current population). If the overlap portion of the sample includes the entire sample from the previous survey, the new sample for the round will have little or no opportunity to represent physicians who were not in the sampling frame in the previous round.

A high degree of overlap also can be less than optimal for certain cross-sectional estimators. That is, the degree of overlap can affect the precision of cross-sectional estimates if it increases the design effect due to unequal weighting. Since the overall respondent sample size is fixed, as the overlap is increased, the sample size available to represent the physicians not in the previous

sampling frame is decreased and the weights for sample members representing these nonrespondents become relatively larger.

b. Optimal Overlap

A key question for Rounds Two through Four was what overlap between rounds was optimal for cross-sectional as well as longitudinal estimates. Because no information was available about the level of correlation between rounds for key study variables in Round Two, we reviewed the sensitivity of optimal overlap at different levels of correlation. Approximately 40 to 50 percent overlap is desirable for a range of the most likely levels of correlation; see analytical details in Technical Publication #38 (www.hschange.com).

For the Round Three and Four overlaps, however, we had information about relative costs and response rates for the various categories of physicians on the sampling frame. The response rates were higher and interviewing costs lower for physicians sampled in prior rounds compared with those sampled for the first time in the current round. We used this information to justify an increase in the size of the overlap sample for Rounds Three and Four compared with Round Two.

We noted above that for change estimates between rounds, the optimal level of overlap is 100 percent. For regression-type estimates, however, the optimal level depends on the amount of correlation between observations obtained for the current and prior rounds. The form of the regression estimates being considered here is:

(1) $\bar{y}' = f_2 \bar{y}'_{2u} + (1 - f_2)\bar{y}'_{2m}$,

where:

f_2 = afunctionof reciprocalvariances
$\bar{y}'_{2u} = \bar{y}_{2u}$
$\bar{y}'_{2m} = \bar{y}_{2m} + b(\bar{y}_1 - \bar{y}_{1m})$,

and b is a constant (for example = 1) or is estimated from data.

The subscripts in equation (1) denote the data collection round (that is 1 or 2) and the overlap (that is matched ["m"] and unmatched ["u"]).

In this form, the means without the prime are the simple means for the matched and unmatched portions of the sample. The primed means, estimated from regression-type equations, are then combined using a parameter (f) involving ratios of reciprocal variances (Cochran 1965).

We note that the maximum optimum overlap for these estimators does not exceed 50 percent and, for most typical correlations, is in the range of 40 to 50 percent. The target overlap for Round One respondents who would be respondents in Round Two was 46 percent; for Round Three, the target overlap rates of respondents to both rounds were increased to 61 percent and 73 percent of the Round Two completed interviews, for PCPs and specialists, respectively; and for Round Four, the target rates were 67 percent for PCPs and 62 percent for specialists, respectively. We used information from prior round costs and response rates and robustness of the cross-sectional estimates to support the rate of overlap.

To investigate the robustness of the cross-section regression estimators, we examined the relative efficiency for different levels of overlap (see Technical Publication #38). We are interested in optimal levels of overlap and loss of potential gain as we move away from that optimum. A range of values for the between-round correlation coefficient (rho, which differs among response variables) was investigated using Round Three data. Little is gained from these estimators for rho values less than 0.5. We also noted that, as rho increases, the optimum percentage overlap decreases. Finally, except for very large correlations, fairly large departures from optimum overlap do not seriously reduce the gain in precision.

2. Errors in Specialty Assignment

In preparing the sample frame, physicians were classified as PCPs or specialists, based on the primary specialty in the AMA and AOA files (as defined in Section B). During the interview, physicians were asked to verify their primary specialties. In some cases, they cited a specialty other than the one listed for them in the AMA or AOA file, requiring a change in classification. These physicians, whom we describe as *switchers,* were reclassified for some analyses, but their selection probabilities remained unchanged. Some unequal weighting resulted from the reclassification, but the number of switchers was small. In Round Two, 7 percent of physicians classified in the sample frame as PCPs responded as specialists, and 4 percent classified in the sample frame as specialists responded as PCPs. In Round Three, 5 percent of physicians classified in the sample frame as PCPs responded as specialists, and 4 percent classified in the sample frame as specialists responded as PCPs. Because PCPs and specialists comprised separate strata with sample size targets within each site, we needed to predict switching in the sample allocation to maintain the desired precision.

3. Geographic Misclassification

A goal of the sample design was to assign physicians to a site based on the location of their main practice. Operationally, we classified physicians listed in the AMA or AOA sample frame by the county of their "preferred mailing address," as that address was the most current on the files. However, as AMA staff indicated, many of these are home addresses rather than main practice locations. In other cases, physicians had moved their practices since the last file update. Nevertheless, even if the actual current practice location did not match the preferred mailing address on the AMA or AOA file, the two addresses usually were within the same site.

Therefore, in response to the survey question about practice location, some physicians gave a different address than the "preferred mailing address" used to assign the physician to a site. As

a result, some of them moved from one survey site to another for analysis purposes. Others were classified as being outside the boundaries of any of the 60 sites. These cases are known as *movers*, even though many of the preferred mailing addresses simply may have been home addresses located in other than the main practice site.

For sampling purposes, physicians remained in the site from which they were originally selected. For example, a physician selected in site A in Round One who actually practiced outside the site (a mover) was considered to be in site A for sample selection purposes in subsequent rounds of the survey. Also, physicians in the Round Three site sample who had a practice address outside the 60 sites for the survey were kept in the sampling frame for Round Four so that they could be included in national estimates. Maintaining the original site assignment enhanced the survey's coverage of physicians in the 48 contiguous states and the District of Columbia. If we had not retained these physicians, we would have progressively lost cases with each round of the survey.

For site-level estimates, physicians for the site sample were linked to the site in which they practiced according to their survey response, rather than to the site from which they originally were sampled. For example, some physicians were selected from a site that did not contain their practice. If the practice was outside the 60 sites, we did not use them in site-level estimates. We also did not use them in national estimates that used site-level independent variables. However, if we selected them from a site other than the one in which they practiced, we included them in the site sample for site-level estimates and for all national estimates. We considered a mover to be a member of the site sample for site-level estimates and some national estimates only if both the original address (based on the preferred mailing address) and the interview location were in the site sample. The probability that both locations would be in the site sample is referred to as the *joint inclusion probability.* Joint inclusion can result in large sampling weights, which can

result in large sampling variances; the issue of weight trimming to account for extreme sampling weights is discussed in Chapter V.

Because some preferred mailing addresses were the same as the home addresses, suburban sites tended to lose more physicians and the more urbanized areas tended to gain them. We adjusted the sample sizes for individual sites for the Round Four allocation to account for anticipated gains or losses caused by these movers.

Movers represented 11.2 percent of the site sample in Round Two and 15.4 percent in Round Three. An increase is expected because the initial site assignment for an individual physician (that is the site assigned when the sampling frame first included the physician) was maintained even though subsequent survey information may indicate a different site for their practice. The movers were a particular problem in the Orange County and Newark high-intensity sites, where the moving rates were 21 and 20 percent, respectively for Round Two, and 22 and 27 percent for Round Three. The trend reflects the cumulative effect of linking physician location to the original frame. Our experience with movers in Round Three is shown in Table II.4.

4. Sample Design Changes for Round 4

Given the reduced budget for Round Four, we analyzed several sample design options that varied assumptions about precision and sample size requirements, national and site design, PCP and specialist sampling rates, and the panel component. The Pearson coefficient of variation was used as the measure of comparison because it is independent of the variate scale. HSC staff selected 21 variables for the analysis; in addition, some difference estimators (Round Three estimate minus Round Two estimate) were examined. The analysis used Round Three

TABLE II.4

MOVERS EXPERIENCE IN ROUND THREE
PHYSICIAN SURVEY

Site	Started in Site	Stayed in Site	Moved out of Site	Moved into Site
Total	**11,238**	**9,509**	**1,729**	**627**
01-Boston MA	532	463	69	23
02-Cleveland OH	482	407	75	6
03-Greenville SC	387	339	48	9
04-Indianapolis IN	454	395	59	6
05-Lansing MI	332	267	65	3
06-Little Rock AR	353	280	73	5
07-Miami FL	492	431	61	11
08-Newark NJ	493	361	132	16
09-Orange County CA	404	315	89	14
10-Phoenix AZ	491	442	49	11
11-Seattle WA	509	455	54	24
12-Syracuse NY	370	315	55	1
13-Atlanta GA	155	141	14	17
14-Augusta GA	134	110	24	2
15-Baltimore MD	139	114	25	12
16-Ridgeport CT	144	118	26	4
17-Chicago IL	135	118	17	21
18-Columbus OH	135	120	15	4
19-Denver CO	143	123	20	11
20-Detroit MI	141	124	17	15
21-Greensboro NC	152	131	21	4
22-Houston TX	135	123	12	11
23-Huntington WV	112	91	21	0
24-Killeen TX	102	78	24	1
25-Knoxville TN	121	97	24	4
26-Las Vegas NV	113	107	6	8
27-Los Angeles CA	129	115	14	65
28-Middlesex NJ	126	97	29	30
29-Milwaukee WI	127	114	13	5
30-Minneapolis MN	136	122	14	8
31-Modesto CA	111	96	15	1
32-Nassau NY	119	84	35	9
33-New York NY	133	115	18	82
34-Philadelphia PA	137	118	19	16
35-Pittsburgh PA	135	126	9	7
36-Portland OR	133	125	8	14
37-Riverside CA	129	108	21	18
38-Rochester NY	135	119	16	4
39-San Antonio TX	135	111	24	4
40-San Francisco CA	158	127	31	12
41-Santa Rosa CA	126	109	17	4
42-Shreveport LA	132	102	30	0
43-St Louis MO	138	127	11	5
44-Tampa FL	124	112	12	7
45-Tulsa OK	133	113	20	1
46-Washington DC	128	115	13	33
47-West Palm Beach FL	130	106	24	5

TABLE II.4 (*continued*)

Site	Started in Site	Stayed in Site	Moved out of Site	Moved into Site
48-Worcester MA	147	108	39	11
49-Dothan AL	73	62	11	2
50-Terre Haute IN	73	61	12	1
51-Wilmington NC	104	87	17	2
52-W-Cen Alabama	33	26	7	0
53-Cen Arkansas	127	115	12	19
54-N Georgia	117	96	21	9
55-NE Illinois	91	73	18	1
56-NE Indiana	81	66	15	2
57-E Maine	128	104	24	1
58-E North Carolina	112	94	18	3
59-N Utah	136	106	30	1
60-NW Washington	102	85	17	2

coefficients of variation as benchmarks and evaluated the effects of stratification, clustering, unequal weighting due to different sampling rates, and unequal weighting due to nonresponse adjustments on design options. All survey estimates and variances were computed taking into account the complex sampling design.

The results indicated that we would lose relatively little precision for national estimates by:

- Retaining a nominal site sample of 7,000 completed interviews
- Oversampling panel members,
- Equal allocation between PCPs and specialists, and
- Elimination of the national supplement.

Compared with Round Three, standard errors for national estimates were projected to increase an average of eight percent. For national estimates that track changes between rounds of the CTS Physician Survey, omitting the national supplement increased standard errors by up to eight percent, but averaged less than five percent. These losses were considered acceptable by HSC researchers; see Table II.5 for a summary.

We reduced oversampling of PCPs because of changing research objectives. Although a proportional allocation of the sample to PCPs and specialists is more efficient for national estimates, we chose an equal allocation to better support comparative analyses between PCPs and specialists. The overlap sample was retained because it permits continued panel analyses and, based on prior experience, was expected to result in a higher response rate. Continued use of the site sample was essential to preserve analytic options, such as use of site means as ecological variables in analyses. Moreover, it preserves the basic design of the CTS surveys, which is important for the validity of tracking estimates. The over-sampling in the 12 high-intensity sites was replaced with a sample allocation to the sites that was more in proportion to

TABLE II.5

SURVEY PRECISION REQUIREMENTS FOR THE PHYSICIAN SURVEY:
ROUNDS ONE THROUGH FOUR

		Effective Sample Sizes[a]			Estimated Sampling Error for a Proportion Near 0.5		
Survey	Estimation Category	PCP	Specialist	Combined	PCP	Specialist	Combined
R1-R3 Site	High-intensity site	400	200	433	0.025	0.035	0.024
R1-R3 Site	Low-intensity site	100	50	114	0.050	0.071	0.047
R1-R3 Site[b]	National	3,450	2,645	4,285	0.009	0.010	0.008
R4 Site	Minimum Site	35	35	70	0.088	0.088	0.060
R4 Site[b]	National	2,590	2,690	4,665	0.010	0.010	0.007

Source: Technical publication #38 and MPR computations.

PCP = primary care physician.

[a]*Effective sample size* is the sample size needed in simple random sampling to achieve the specified precision; this sample size times the design effect is the nominal sample needed to achieve this precision in the survey. Design effect is defined as the variance of an estimate from the survey, divided by the variance of the estimate if a simple random sample of the same size were used, and is a measure of the effect of stratification and differential sampling rates.

[b]No specified constraints were adopted for national-level estimates from the site sample; numbers in this case are approximated by average design effects.

the physicians represented by the site, thus restricting or eliminating the ability to report on site-level estimates.

The first three rounds of the survey included an independent stratified national random sample of physicians (the national supplement) to improve national estimates. However, because design effects in prior rounds were lower than projected, it was possible to achieve precision requirements in the original design with smaller nominal samples. Eliminating the national supplement did not reduce the range of analytic questions that could be addressed by the survey, although for some there will be slightly lower statistical power to determine significant relationships. Many analyses involve models that use market-level variables for the 60 sites; these analyses do not require the national supplement. Eliminating the supplement also simplifies the process of developing weights and makes the data somewhat easier to use because

there are fewer analysis weights. The computation of the sampling and analysis weights is discussed in Chapter V. The key design decisions are summarized below:

- Reducing the sample from over 12,000 to 7,000 completed interviews
- Reducing the extent of PCP over-sampling to achieve approximately equal samples of PCPs and specialists
- Retaining the 60-site sample, but allocating the sample among sites so that (1) the sample sizes in the 9 certainty sites were a compromise between an equal allocation (optimum for sites) and a proportional allocation (optimum for national estimates); and (2) the allocation should result in approximately equal effective sample sizes for the noncertainty sites and equal effective sample sizes for PCP and specialists within each site.[7]
- Eliminating the national supplement
- Sampling physicians new to the survey at approximately the same rate as those representing the Round Three population.

D. IMPLEMENTATION

1. Sampling Frame

The sampling frame was developed from physician records maintained by the AMA and AOA, as in previous rounds. These files contained the most current information available from the two organizations as of November, 2003, just prior to the date used to select the Round Four sample. The data fields on records in the file included names, telephone numbers, addresses, dates of birth, specialties, and other information useful for sampling, data collection, and weight computations. Because of the longitudinal nature of this survey, we also used information from the Round Three frame and response status for each physician in the Round Three sample in the frame development.

[7]The effective sample size is the sample size adjusted for the loss in sample efficiency from the sample design (also called design effect). The design effect is caused by stratification and different sampling rates across strata and sites.

The five key steps used to construct the frame were:

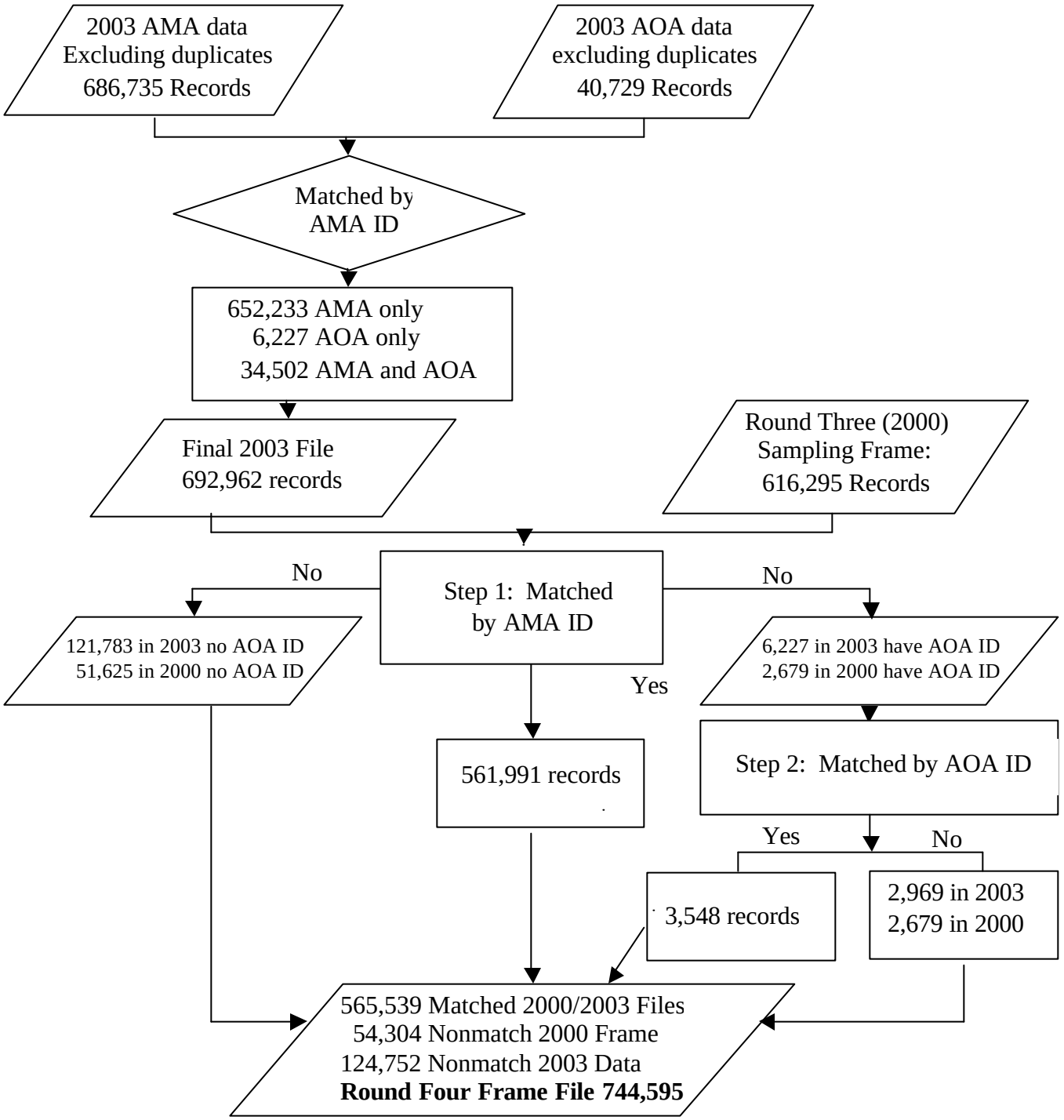
1. Specify file content and format for ordering the files
2. Verify file content after receiving the AMA and AOA files
3. Match the 2003 AMA and AOA files against each other and against the Round Three sample frame to identify physicians in both Round Three and Round Four files and physicians added to the sample frames since Round Three
4. Exclude ineligible physicians
5. Classify records by primary design strata and site and by the specialty and Round Three response status secondary strata

The complete list of physicians for the Round Four sampling frames was obtained from the AMA and AOA (physicians in the new master files combined with those on the Round Three frame). After reviewing frequency counts for key items to ensure file accuracy and completeness, we performed a series of processing steps (Figure II.1). We matched the AMA and AOA files to identify physicians in each file; then we matched the combined AMA/AOA file to the Round Three frame and sample. We performed a computer match by AMA identification number to determine which physicians were on both AMA and AOA files and which were new to the Round Four frame. Two types of nonmatches resulted: (1) physicians on the Round Three frame who were not identified on the Round Four list, and (2) physicians on the Round Four list but not on the Round Three frame.

Physicians on the Round Four list that were not in the Round Three frame were excluded as ineligible if their primary specialty was listed in Tables II.1 (AMA) or II.2 (AOA), if their major professional activity was administration, teaching, or research, or their preferred mailing address or practice location was not in one of the 48 contiguous states or the District of Columbia.

FIGURE II.1

SAMPLING FRAME CONSTRUCTION FOR PHYSICIAN SURVEY
2003 SOURCE FILES (AMA AND AOA) AND
2000 SAMPLING FRAME



Matched records and new records also were excluded from the frame if they were currently classified by either AMA or AOA as retired, deceased, or practicing in a foreign country.

Because physicians are added to the AMA and AOA files on an ongoing basis, we identified physicians in the Round Four frame who were not in the Round Three frame to define them as a separate stratum and to receive a specific sample allocation. The records were then assigned to primary design strata and Round Three response status secondary strata and the sample was allocated on the basis of the counts to these strata and Round Three response status secondary strata. (Section D.2 discusses primary design strata and Round Three response status secondary strata.)

Next, each physician was linked to an appropriate site. For sampling purposes, we based the site designation on the physician's preferred mailing address on the AMA and AOA files.

Finally, each physician was classified as either PCP or specialist. This classification was based on the Round Three survey response (if available) or on the specialty code from the AMA or AOA data files.

2. Sampling Units and Stratification

Stratification, a feature of most large-scale surveys, performs several important functions. Using strata containing populations expected to have similar responses can increase survey precision. Another key function of stratification is to ensure an adequate sample size for important study populations. Stratification also is a useful tool for optimum allocation in surveys in which some groups exhibit more variability in responses or are more costly to survey. The design for Round Four used stratification to improve precision and to ensure adequate representation by site, geographic region, population density, and physicians who were new to

the frame. We also used stratification to control precision for survey estimates of PCPs and specialists.[8]

The sample of sites was stratified geographically by region and population size and was selected with probability proportional to size (estimated population for July 1992). Within each site, we stratified the sample by PCPs and specialists (primary strata) and by the following four secondary strata (see Table II.6)[9]:

1. Physicians who completed interviews in Round Three
2. Physicians who were selected for Round Three but did not complete interviews (refusals, ineligible, unlocated)
3. Physicians who were in the AMA/AOA sample frames for Round Three but were not selected in the Round Three sample
4. Physicians who were not in the AMA/AOA sample frames for Round Three and were new to the frame for Round Four

The resulting frame counts are listed in Table II.7. The numbers in the two explicit strata, PCP and specialist, and for the four sampling strata within each of those two strata are presented for each site.

3. Sample Allocation

For Round Four, the goals of the sample allocation were to achieve the highest possible precision for national estimates and sufficient precision for site-level estimates to develop stable site-level covariates for national analyses. In addition, the Round Four design was based on

[8]We expect that some groups sampled for Round Four, such as physicians who could not be located or who refused in Round Three, will be more costly to survey or will have lower response rates. We used data from Round Three on interviewing costs and response rates to optimize sampling rates for different groups of Round Three respondents for Round Four.

[9]The first three secondary strata are partitions of the Round Three sampling frame with the first two comprising the Round Three sample. These two differ by survey outcome (Round Three response status). The fourth sampling class represents physicians new to the sampling frame.

TABLE II.6

STRATIFICATION AND SAMPLING ASSUMPTIONS AND SPECIFICATIONS FOR ROUND FOUR PHYSICIAN SURVEY

Sample	Primary Strata	Site Definition	Site Selection Classification[a]	Selection Assumptions	Primary Unit	Within-Site Stratification	Secondary Unit
Site Sample	1. MSAs with >200,000 population (1992 Census Bureau data)	MSAs	Certainty sites (9)	Equal probability with replacement sampling within sites	CTS site	PCP/specialist (2) with frame sampling classes (4)[b]	Physician
			Noncertainty sites (39)	PPS without replacement sampling of sites and equal probability with replacement sampling within sites	CTS site	PCP/specialist (2) with frame sampling classes (4)[b]	Physician
	2. MSAs with <200,000 population (1992)	MSAs	Noncertainty sites (3)	PPS without replacement sampling of sites and equal probability with replacement sampling within sites	CTS site	PCP/specialist (2) with frame sampling classes (4)[b]	Physician
	3. Nonmetropolitan areas	BEA county groups	Noncertainty sites (9)	PPS without replacement sampling of sites and equal probability with replacement sampling within sites	CTS site	PCP/specialist (2) with frame sampling classes (4)[b]	Physician

[a]Of the 48 MSAs with population >200,000, 9 were selected with certainty. Site selection procedures differed for these sites. (See Section II.A.3.)

[b]The four secondary frame sampling classes are (1) Round Three completes; (2) Round Three noninterviews (including nonrespondents, ineligible respondents, and unlocatable physicians); (3) physicians in the Round Three AMA or AOA frames who were not sampled for Round Three; and (4) physicians who were not in the Round Three AMA or AOA frames but who were new to the frames for Round Four.

TABLE II.7

FRAME COUNTS FOR THE PHYSICIAN SURVEY

Site	All	PCP				Specialist			
		Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame	Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame
Total	276,425	7,175	7,407	76,799	19,971	4,063	3,190	136,744	21,076
01-Boston MA	14,452	338	361	3,536	969	194	124	7,564	1,366
02-Cleveland OH	6,069	318	301	1,257	474	164	141	2,922	492
03-Greenville SC	1,724	241	213	147	130	146	95	657	95
04-Indianapolis IN	3,721	294	258	642	288	161	116	1,637	325
05-Lansing MI	1,131	177	188	76	90	155	121	248	76
06-Little Rock AR	1,657	175	230	61	120	178	126	618	149
07-Miami FL	5,679	319	489	1,160	305	173	184	2,688	361
08-Newark NJ	5,211	335	286	1,195	306	158	149	2,476	306
09-Orange County CA	6,221	268	333	1,676	424	136	144	2,891	349
10-Phoenix AZ	5,578	318	346	1,251	490	173	154	2,486	360
11-Seattle WA	5,837	343	255	1,458	402	166	143	2,612	458
12-Syracuse NY	1,668	217	156	190	84	153	116	651	101
13-Atlanta GA	7,485	102	82	2,272	484	53	23	3,890	579
14-Augusta GA	1,237	76	78	186	98	58	31	585	125
15-Baltimore MD	7,232	89	95	2,020	473	50	25	3,851	629
16-Bridgeport CT	2,411	93	101	566	128	51	40	1,288	144
17-Chicago IL	18,828	95	93	6,482	1,494	40	33	9,153	1,438
18-Columbus OH	3,402	84	97	923	324	51	33	1,618	272
19-Denver CO	5,637	91	91	1,707	410	52	30	2,831	425
20-Detroit MI	10,212	91	107	3,305	803	50	48	5,098	710
21-Greensboro NC	2,284	102	62	565	158	50	20	1,151	176
22-Houston TX	8,677	86	102	2,433	615	48	37	4,484	872
23-Huntington WV	688	70	85	89	64	42	29	267	42
24-Killeen TX	653	62	68	92	75	40	35	222	59
25-Knoxville TN	1,586	78	120	350	94	43	29	790	82
26-Las Vegas NV	2,174	73	102	566	197	40	51	973	172
27-Los Angeles CA	18,872	88	139	6,096	1,277	41	57	9,850	1,324
28-Middlesex NJ	4,094	85	70	1,244	329	41	33	2,005	287
29-Milwaukee WI	3,724	85	64	1,149	228	42	31	1,894	231
30-Minneapolis MN	5,754	97	85	2,004	421	39	39	2,661	408

TABLE II.7 *(continued)*

Site	All	PCP				Specialist			
		Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame	Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame
31-Modesto CA	635	78	67	117	57	33	25	228	30
32-Nassau NY	9,503	85	92	2,962	517	34	51	5,129	633
33-New York NY	23,230	87	110	6,502	1,915	46	58	12,251	2,261
34-Philadelphia PA	15,247	92	85	4,716	1,044	45	31	8,073	1,161
35-Pittsburgh PA	5,880	85	79	1,679	420	50	38	3,089	440
36-Portland OR	4,457	89	97	1,391	365	44	25	2,122	324
37-Riverside CA	3,882	84	128	1,291	276	45	39	1,806	213
38-Rochester NY	2,510	91	63	778	152	44	28	1,175	179
39-San Antonio TX	3,192	86	111	754	282	49	31	1,583	296
40-San Francisco CA	6,453	102	137	1,762	445	56	47	3,372	532
41-Santa Rosa CA	947	80	76	242	51	46	22	395	35
42-Shreveport LA	1,036	81	94	111	92	51	32	458	117
43-St Louis MO	5,672	87	91	1,578	395	51	35	2,973	462
44-Tampa FL	4,891	79	127	1,432	352	45	31	2,513	312
45-Tulsa OK	1,719	90	97	513	136	43	33	708	99
46-Washington DC	12,868	89	80	3,864	987	39	39	6,744	1,026
47-W Palm Beach FL	2,536	77	130	614	112	53	68	1,375	107
48-Worcester MA	1,855	94	65	543	155	53	19	811	115
49-Dothan AL	282	27	36	16	10	46	35	96	16
50-Terre Haute IN	266	40	47	17	16	33	28	66	19
51-Wilmington NC	545	59	69	39	42	45	32	215	44
52-W-Cen Alabama	66	29	15	8	6	4	4	0	0
53-Cen Arkansas	966	84	68	325	78	43	27	299	42
54-N Georgia	1,107	75	79	310	101	42	34	405	61
55-NE Illinois	362	55	63	49	29	36	23	97	10
56-NE Indiana	200	55	31	17	11	26	19	33	8
57-E Maine	715	87	39	193	65	41	27	230	33
58-E North Carolina	525	70	54	79	44	42	29	173	34
59-N Utah	416	82	65	46	26	54	18	95	30
60-NW Washington	564	66	55	153	36	36	25	169	24

Round Three Completed Interview = Round Three completed interviews; Round Three Noninterview = Round Three noninterview (including nonrespondents, ineligible respondents, and unlocatable physicians); remaining Round Three Frame = Round Three frame cases not in the Round Three sample; 2004 Added to Frame = physicians in the 2004 AMA and AOA frames but not in the 2000 frames.

equal sample allocation of PCPs and specialists, in contrast to prior rounds where PCPs were more heavily over sampled.

For the sample allocation, we first took account of the probability of selecting each site. Because the nine certainty sites would be assigned a large sample allocation in a strictly proportional allocation, the sample allocation for these sites was reduced to distribute more sample to the 51 other (non-certainty) sites. In this redistribution, we used the goal of approximately equal effective sample size[10] within each site and used the finite population correction factors[11] to guide the redistribution (see Table II.5 for the targeted effective sample sizes).

Next, we used the completion rates from Round Three to adjust the allocation. The completion rate is the number of completes divided by the sample size so it incorporates both the response and eligibility rates. Then we adjusted this allocation for site-level estimates, geographic misclassification (the "movers"), and changes in patient care classification from PCP to specialist or vice versa ("switchers"), based on Round Three experience. We calculated the adjustment factor as:

(2) $F = S/(S - L + G)$,

where the denominator is equal to the starting number S minus the loss L plus the gain G. For movers, we made site-specific adjustments. For switchers, we made site-specific adjustments for

[10] *Effective sample size* is the sample size needed in simple random sampling to achieve the specified precision.

[11] *Finite population correction factor* is the multiplier needed to reflect the fact that the variance of an estimate is a function of both the sample size and the population count—smaller sample sizes needed for smaller populations.

the sites that were high-intensity in Round Three, but used overall average adjustments for the remaining sites.

After these adjustments, we used the design effect related to unequal weights to further adjust the sample allocation. Finally, we decreased the sample sizes for PCPs to make them approximately equal to those for the specialists. Table II.8 presents the sample sizes adjusted to account for nonresponse and ineligibility; the resulting *base* sample is the expected sample needed to obtain the target number of completed interviews).

The numbers in Table II.8 reflects the nominal allocation adjusted for response and eligibility and then stochastically rounded to establish a fixed sample size of physicians and allocation (the "base sample"). We expected this sample would be necessary to complete 7,000 interviews, assuming eligibility and response rates for Round Four were the same as for Round Three. In practice, we expected that eligibility and response rates would vary by strata, with some achieving their targets with less than a 100 percent allocation and some requiring more than a 100 percent allocation. To control for uncertainty in response and eligibility rates, we selected an *augmented sample,* which included the base sample and a reserve sample equal in size to approximately 50 percent of the base sample. We then randomly partitioned the augmented sample into waves and released subsamples throughout the survey. The base sample (100 percent allocation) consisted of 14,334 physicians; the final released sample included 15,063 physicians (includes the non-contacts and some others that were released but not fielded—classified unlocated cases). The base sample counts are presented in Table II.8.

TABLE II.8

BASE SAMPLE COUNTS FOR THE PHYSICIAN SURVEY
(Base sample sizes: the anticipated number needed to obtain 7,000 completed interviews)

Site	All	Specialist							
		Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame	Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame
Total	14,334	3,231	2,198	111	1,575	2,836	1,426	1,822	1,135
01-Boston MA	439	111	77	--	58	110	43	--	40
02-Cleveland OH	229	55	33	--	32	57	32	--	20
03-Greenville SC	186	46	27	--	22	55	23	--	13
04-Indianapolis IN	204	51	28	--	26	54	25	--	20
05-Lansing MI	209	56	36	--	25	50	26	--	16
06-Little Rock AR	239	53	46	--	33	60	26	--	21
07-Miami FL	239	60	59	--	22	51	33	--	14
08-Newark NJ	247	64	35	--	24	67	41	--	16
09-Orange County CA	258	59	46	--	26	66	44	--	17
10-Phoenix AZ	208	48	33	--	28	54	30	--	15
11-Seattle WA	192	49	23	--	20	52	29	--	19
12-Syracuse NY	199	56	25	--	18	58	28	--	14
13-Atlanta GA	182	49	25	--	21	53	15	--	19
14-Augusta GA	219	49	33	--	32	58	24	--	23
15-Baltimore MD	315	67	46	--	33	50	17	76	26
16-Bridgeport CT	241	57	40	--	22	51	26	31	14
17-Chicago IL	592	95	60	48	56	40	22	229	42
18-Columbus OH	207	46	34	--	30	51	21	--	25
19-Denver CO	205	50	34	--	24	52	20	--	25
20-Detroit MI	360	73	56	--	39	50	32	86	24
21-Greensboro NC	208	56	21	--	26	50	13	24	18
22-Houston TX	323	65	50	--	36	48	25	69	30

TABLE II.8 *(continued)*

Site	All	Specialist							
		Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame	Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame
23-Huntington WV	230	48	38	--	30	42	20	37	15
24-Killeen TX	235	46	33	--	36	40	24	33	23
25-Knoxville TN	247	49	50	--	21	43	19	52	13
26-Las Vegas NV	199	40	34	--	26	40	37	--	22
27-Los Angeles CA	436	79	80	--	39	41	36	134	27
28-Middlesex NJ	193	47	23	--	25	41	21	21	15
29-Milwaukee WI	213	52	26	--	20	42	21	38	14
30-Minneapolis MN	210	49	28	--	21	39	25	31	17
31-Modesto CA	201	45	25	--	22	33	16	48	12
32-Nassau NY	306	65	44	--	24	34	35	85	19
33-New York City NY	501	87	71	--	64	46	40	151	42
34-Philadelphia PA	526	92	55	39	50	45	21	190	34
35-Pittsburgh PA	201	48	29	--	24	50	28	--	22
36-Portland OR	184	43	30	--	23	44	23	--	21
37-Riverside CA	203	43	42	--	20	45	29	--	24
38-Rochester NY	201	53	24	--	18	44	18	27	17
39-San Antonio TX	240	48	40	--	34	49	21	25	23
40-San Francisco CA	239	52	46	--	28	56	35	--	22
41-Santa Rosa CA	193	49	30	--	15	46	14	30	9
42-Shreveport LA	244	49	37	--	36	51	22	21	28
43-St. Louis MO	207	49	33	--	25	51	26	--	23
44-Tampa FL	221	45	46	--	23	45	20	27	15
45-Tulsa OK	225	52	36	--	23	43	21	34	16
46-Washington DC	376	76	45	--	41	39	25	123	27
47-W Palm Beach FL	254	52	58	--	18	53	52	--	21
48-Worcester MA	219	56	25	--	27	53	13	28	17
49-Dothan AL	161	27	36	16	10	46	15	--	11
50-Terre Haute IN	166	40	33	--	15	33	16	16	13
51-Wilmington NC	200	43	33	--	25	45	20	18	16

TABLE II.8 *(continued)*

Site	All	Specialist							
		Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame	Round Three Completed Interview	Round Three Noninterview	Remaining Round Three Frame	2004 Added to Frame
52-W-Cen Alabama	66	29	15	8	6	4	4	--	0
53-Cen Arkansas	138	40	20	--	15	43	10	--	10
54-N Georgia	180	44	31	--	22	42	20	9	12
55-NE Illinois	180	42	31	--	16	36	14	34	7
56-NE Indiana	138	43	15	--	9	26	12	26	7
57-E Maine	187	52	15	--	21	41	16	29	13
58-E North Carolina	166	44	22	--	22	42	17	6	13
59-N Utah	161	49	25	--	14	50	10	--	13
60-NW Washington	186	49	27	--	14	36	15	34	11

Round Three Completed Interview = Round Three completed interviews; Round Three Noninterview = Round Three noninterview (including nonrespondents, ineligible respondents, and unlocatable physicians); remaining Round Three Frame = Round Three frame cases not in the Round Three sample; 2004 Added to Frame = physicians in the 2004 AMA and AOA frames but not in the 2000 frames.

III. SURVEY DESIGN AND PREPARATION

A. SCHEDULE

Survey design, preparation, and data collection for the Round Four Physician Survey were conducted from March 2003 through July 2005 and sample weights and analysis files were prepared from July 2005 through March 2006. The survey schedule and organizations responsible for each activity are shown in Table III.1.

TABLE III.1

ROUND FOUR PHYSICIAN SURVEY SCHEDULE

Dates	Activities	Responsibility
3/03–10/03	Design questionnaire	HSC
6/03-2/04	Design sample	MPR
11/03-2/04	Conduct cognitive testing and revise instrument	GALLUP AND HSC
3/04-4/04	Program instrument and test program	GALLUP
3/04–4/04	Prepare and obtain approvals for study endorsements and advance letter	GALLUP AND HSC
3/04–5/04	Prepare sample for field	MPR
5/04-5/05	Release sample (four releases)	MPR
5/04	Develop interviewer training materials and train interviewers	GALLUP
5/04-4/05	Trace telephone numbers and addresses	MPR
6/04–7/05	Mail advance letters and conduct interviewing	GALLUP
8/04	Deliver first 400 cases to SSS for data review	GALLUP AND SSS
9/04	Deliver interim data and mover file	GALLUP
7/05	Deliver final data file	GALLUP
7/05-3/06	Weight data and prepare analysis files	MPR and SSS

B. INSTRUMENT DEVELOPMENT

Several topics were added to the survey after an assessment of the policy value of each topic, review of the literature, and an assessment of added burden. Additional topics and their impact on policy are shown in Table III.2. To provide time for the new questions, we eliminated a number of questions throughout the survey that proved to have only limited analytic value. These included items on number of practices, practice ownership other than by respondent, Internet access, specialist scope of care, communications between specialists and PCPs, and risk adjusted profiling. Questions on ability to obtain services and income from bonuses were modified. A summary of all questions deleted from the Round Three Survey and added for the Round Four Survey is shown below in Table III.3.

TABLE III.2

NEW SURVEY TOPICS AND POLICY VALUE

Topic	Policy Value
Rank importance of reasons why physicians are not accepting Medicaid and/or Medicare patients	Allows enhanced analysis of key access issues.
Relationship between patient cost sharing and clinical decision making	Given the expected rise in patient cost-sharing and the results of a study that showed physicians generally do not talk with patients about cost sharing, these questions assess the extent to which physicians take cost sharing into account in deciding on treatment.
Coordination of care and information technology	Adds questions on use of information technology, including e-prescribing and hospitalists.
Use of CPOE in the hospital and access to hospital system for anonymously reporting medical errors	Adds key patient safety measures.
Threats to quality of care	Measures relative importance of various threats to quality of care.
Case mix – chronic conditions, race/ethnicity, language problems	Information on case mix permits analysis of the extent to which physicians with more vulnerable patients are different and have different experiences.
New scales for questions regarding inability to obtain needed services	There was insufficient variation from scale used in prior round.
Location of charity care	Provides an understanding of whether physicians provide charity care in their offices or in other settings such as clinics or the ER.
Patient encounters in in-patient and out-patient settings (PCPs)	Adds productivity measures to survey.
Degree to which financial incentives affect compensation	Assesses the relative importance of performance measures such as profiling, quality and patient satisfaction in determining physician income.

TABLE III.3

CHANGES MADE TO THE ROUND FOUR PHYSICIAN SURVEY

	Items Added to Round Four		Items Deleted From Round Three	
Section and Topic	Item #	Topic	Item #	Topic
A. Introduction and screening			A3a, A3b, A3c	Status of new interviewees: as of previous round
			A4, A4a	Number of practices
			A12, A14, A16, A18	Board eligibility
B. Time, productivity, info brought by pts, case mix	B5a A-D	Number of patient visits in four different settings; asked only of PCPs	B7 through B11	Information brought by patients
	B6a	Location of charity care		
	B12, B14, B15 \(No B13)	Case mix: chronic conditions, race/ethnic group, language, communications problems		
C. Type and size of practice	C8a	Level of nursing support compared to 3 yrs ago	C4, C5,A-D, C6	Other owners of practice
	CX,CY	Effect of personal financial incentives (All asked here; half had been in Section H in R3—H10b, H10b-1)	C8	Number nurse practitioners, etc.
			C10, C11	Practice acquired in last two years
			C12A-D	Practice preferences
	CZ	Competition Q (H10c) moved here from Section H		

TABLE III.3 *(continued)*

	Items Added to Round Four		Items Deleted From Round Three	
Section and Topic	Item #	Topic	Item #	Topic
D. Care management, IT, hospital safety, scope of care	D1F1	New IT item on clinical data exchange w hospitals and labs	D2	Internet access
			D4B+ B1 D4C+ C1	Effect of profiling, patient satisfaction surveys on practice of medicine
Note: There is no Section E	D1H1	New IT item on drug interactions		
	D2aa	Percent prescriptions written electronically (docs who e-prescribe)	D5B-E	Effect of CM tools on ability to provide efficient and high quality care
	D6a, D6b	CPOE, medical errors (asked of specialists and also PCPs with hospital visits)	D7	PCP change in scope of care
	D7 (reuse of Q #)	Percent hospitalized pts with hospitalist	D11, D12, D13	Specialist scope of care Qs
F. Quality, ability to obtain services, cost sharing, new patients	F8b A-D F8c A-C	Whether unable to obtain specific services; Reasons for inability to obtain services	F1F/G	PPQ: level of communication with specialists/PCPs
	F8d A-C	Cost sharing	F8A-G, F8a (all)	Ability to obtain services base question revised; now F8b, F8c
	F11 F12	Reasons not accepting new Medicare patients Reasons not accepting new Medicaid patients	F10	Whether accepting new capitated patients
G. Practice revenue		NONE		NONE
H. Compensation, PPQ follow up	General Notes	Change in skip patterns to determine compensation type; renumbering of income and race questions.	H6/H8	Whether profiles are risk-adjusted
			H9, H9a	Percent income from bonuses; if none, eligible for bonuses?
	H4A	End of year adjustments (asked only of those not eligible to receive bonuses)		
	H5E, H7E	New compensation factor: overall financial performance of practice		
	H7a A-E	Importance of compensation factors		
	H20A-F, H	Importance of factors that may limit ability to provide high quality care		

C. COGNITIVE AND PILOT TESTING

1. Cognitive Testing

Cognitive testing was divided into two groups of questions because there were too many new items for a single testing session. Group A included questions regarding access (except for charity care) and compensation. Group B included questions regarding charity care, productivity, information technology, cost sharing, case mix, coordination and patient safety, and threat to quality. Because of complicated skips and fill-ins, Group A was programmed into Computer-Assisted Telephone Interviewing (CATI) to ease the interviewers' ability to navigate the script. The skip patterns included in Group B were not as complex as those in Group A and did not require CATI programming.

Gallup purchased a nationally representative sample of physicians from Medical Marketing Service, Inc. (MMS) that provided an appropriate distribution of primary care physicians and specialists. The sample excluded full-time federal employees, residents or fellows, physicians who perform less than 20 hours of direct patient care during a typical week, physicians who practice outside of the continental United States, and specialties excluded from previous rounds of the CTS Physician Survey. A screener was used to ensure both primary care physicians and specialists were recruited. Screener questions were included to ensure that participants met CTS eligibility requirements. Physicians were offered $100 honoraria for completing the cognitive interviews.

Two executive interviewers who had worked on prior rounds of the survey and had experience conducting cognitive interviews conducted the cognitive interviews. Gallup's Cognitive Testing Director listened to audiotapes of each interviewer's first few completed interviews and Gallup's Co-Project Director reviewed the written transcripts. Detailed feedback

was then provided to the interviewers and their supervisor before any further interviews were attempted. Each interview was tape-recorded and transcribed.

A total of 20 interviews were completed with Group A and 15 with Group B; the cognitive interviewing protocol and report are included in Appendix A. Following review of the results by HSC and Gallup staff, a draft instrument was programmed and prepared for pilot testing.

2. Pilot Testing

The objective of the pilot test was to assess skip patterns, verify that the CATI program did not contain any errors, and evaluate the time required to administer the interview. The pilot survey sample was selected from physicians who had not previously participated in the CTS Physician Survey and for whom telephone numbers were available for the pilot test. The sample excluded full-time federal employees, residents or fellows, physicians who perform less than 20 hours of direct patient care during a typical week, and physicians who practice outside of the continental United States. A team of ten executive interviewers, who had worked on prior rounds of the survey and would be assigned to Round Four, conducted the pilot test interviews.

Pilot test interviewing was conducted from March 8 to March 20, 2004. Twenty-seven interviews were conducted with primary care physicians and 23 with specialists. No substantive problems were encountered during the data collection. However, the mean time to conduct the interview was 28 minutes, considerably longer than projected. To reduce the length of the interview, we prioritized the policy value of each topic and item and dropped several questions; the final changes from the Round Four instrument are shown above in Table III.3. The mean length of the Round Four interview was 21.3 minutes. The final instrument is included in Appendix A.

D. PREPARATION OF ADVANCE LETTERS

As in previous rounds, we prepared and mailed an advance letter to sampled respondents one week before the release of each sample. Because endorsement by medical societies generally increases response rates, we asked societies that endorsed prior rounds to provide their endorsement for Round Four. Medical societies endorsing the study included the American Medical Association, the American Osteopathic Association, the American Academy of Family Physicians, the American Academy of Pediatrics, the American College of Physicians—American Society of Internal Medicine, the American Psychiatric Association, and the American College of Surgeons. In addition to the letter describing the survey and requesting the physician's participation, initial mailings included a brief description of HSC and a list of recent articles accessible on HSC's Web site (see Appendix A for advance letters and HSC factsheet).

We used different versions of the advance letter depending upon whether the physician had participated in the survey before; if he or she had, the letter thanked the respondent for prior participation. We also conducted an experiment in Round Four to determine whether pre-paying physicians who were being sampled for the first time would increase the likelihood of participation in the current round and reduce the number of calls to complete an interview. (Results of pre-payment experiments conducted in this and prior rounds are discussed in Chapter IV.) Additional copies of the letters were mailed to physicians who said they had not received them.

E. CATI SYSTEM, TELEPHONE MANAGEMENT SYSTEM, AND CASE MANAGEMENT SYSTEM

The CATI instrument was programmed on the SURVENT system. SURVENT interfaces with Gallup's Telephone Management System (TMS), which distributes telephone numbers to each interviewer. For Round Four (as with prior rounds), the sample was divided into waves,

comprising random sub-samples of the total sample. Additional waves were released throughout the field period until target numbers of completed interviews were achieved. The system maintains call histories on every released case to support reports on survey progress and disposition and measures of interviewer productivity. Data on call histories from the TMS and data on tracing efforts were combined in a case management system (CMS) that was programmed in Microsoft Access. We used the CMS, which permitted greater flexibility in ad hoc reporting and sample reconciliation, for weekly progress reporting, analyzing interviewer productivity, and tracking sample across various categories throughout the field period (response categories are shown in Chapter IV).

F. INTERVIEWER SELECTION

The CTS Physician Survey was an "executive ownership" study, which means that it was conducted by executive interviewers who specialize in interviewing physicians, and other health professionals and business executives. Executive ownership also means that the interviewers "owned" their cases. Interviewers were responsible for setting and keeping their own callback appointments. They therefore had ample opportunity to establish rapport with office workers, as well as with the physicians themselves.

All of the interviewers had worked on prior rounds of the survey. Although virtually all physicians speak English, some practice receptionists or other office staff prefer Spanish, especially in the Miami site. As in previous rounds, a bilingual interviewer communicated with Spanish-speaking receptionists and other practice staff; however, all interviews with physicians were conducted in English.

G. INTERVIEWER TRAINING

We updated the content of training materials for Round Four to reflect questionnaire modifications, new information, and feedback from pretest interviews. Interviewers received the following documents, which they kept in their carrels when making calls:

1. Physician specialty lists (alpha and numeric for cross-reference)
2. Copies of each of the advance letters
3. HSC Overview
4. 9" × 12" flat outgoing RWJ envelope
5. Interviewer's manual

Interviewer training was conducted in May 2004. The training session on the survey instrument was designed to provide background information on the study, summarize the sample and sample release procedures, review the instrument, and highlight issues that had been discovered during pretesting. Since all of the interviewers had worked on prior rounds of the study, the review focused on changes since Round Three. A review of sample design and release procedures also was provided.

Following the reviews, interviewers participated in practice interviews that presented various scenarios. After a final debriefing and discussion at the end of the training session, interviewers conducted additional mock interviews until they were comfortable with the instrument and the information provided during training.

IV. DATA COLLECTION

In this chapter, we describe data collection activities, including telephone center supervision and monitoring procedures, tracing activities to locate physicians, efforts to increase response rates by attempting to convert refusals and by using monetary incentives, response rate calculations, and data preparation tasks. Overall, we completed 6,628 interviews with eligible physicians. The unweighted response rate was 55.2 percent (compared with 60.5 percent in Round Three) and the weighted response rate was 52.3 percent (compared with 58.6 percent in Round Three).

A. TELEPHONE CENTER SUPERVISION AND MONITORING

The supervisors monitored interviews, reviewed and resolved problem cases, produced reports, and communicated interviewing problems to HSC staff. In addition, all of the interviewers assigned to Round Four had worked on one or more of the previous rounds. The same monitoring procedures were used in Rounds Two, Three, and Four. A total of 15 percent of the interviewers' work was monitored by supervisors, who listened to a sample of interview attempts, refusal conversion calls, and full interviews. For full interviews, the supervisors used a standard evaluation form which scored interviewers on explaining the survey, reading questions verbatim, recording responses accurately, using objective probing techniques, courtesy, voice quality, and diction. An abbreviated scoring system was used to evaluate interview attempts and refusal conversion attempts. A perfect evaluation score was 50 points, and interviewers were expected to maintain a score of at least 48.

B. LENGTH OF INTERVIEW

The average length of the Round Four interview was 21.3 minutes, slightly longer than Round Three (20.8 minutes) and Round Two (19.1 minutes). The average interview length for PCPs during Round Four was 22.3 minutes, while the average for specialists was 20.2 minutes.

C. SPANISH-SPEAKING PHYSICIANS

In sites with sizable Hispanic populations, bilingual interviewers occasionally had to make appointments with Spanish-speaking office workers. However, as in previous rounds, the actual interviews were conducted in English.

D. TRACING

MPR conducted two types of tracing activities. In the first phase, we sent cases with missing telephone numbers to a vendor that used directory assistance and telephone matching software to obtain new numbers. Some of these telephone numbers were incorrect, as were some numbers obtained for physicians sampled in prior rounds. In the second phase, we made an intensive effort to obtain telephone numbers to replace the incorrect ones, as well as current numbers of physicians who had changed practices.

During the second phase, several websites were used to update physicians' addresses and telephone numbers since no single web site provided comprehensive tracing information. The web sites used for the Round Four survey are discussed below in the sequence with which they were used.

1. **http://www.choicetrust.com**. By entering the doctor's last name and state, we were able to retrieve a history of the doctor's business addresses complete with dates and specialty. In many instances, telephone numbers were provided, which was helpful in tracking interstate address changes.

2. **http://www.medicare.gov**. After entering the doctor's last name and state, we were able to obtain a list of the doctor's practicing locations and specialty. Since telephone numbers were not provided at this website, we used **Accurint** (http://www.accurint.com), a subscription service that accesses various public record databases and provides addresses and, in many cases, telephone number updates as people update their credit file and other public records. In order to obtain the most recent business address and/or telephone number, Accurint's "People at Work" and "Business Search" functions were used. In cases where we were unable to obtain a business number, we used the "Person Search" function. In addition, Accurint provided birth dates for some cases where they were previously unavailable. This proved to be helpful in confirming that a particular physician had either retired or passed away.

3. **http://www.docboard.org**. This site allowed us to access state licensing boards. By clicking on the state of the doctor's practice, we were often able to discover the address of the doctor's practice. Once an address was obtained, we contacted Accurint to obtain a telephone number for the address. This website also has a feature that does a combined search for all states on the "Administrators in Medicine" server, which was helpful in tracking interstate searches.

4. **http://www.google.com**. In cases where the doctor had a somewhat unusual name, we employed the Google search engine. Entering the doctor's name with quotation marks often provided links that aided in tracking the doctor through publications, clubs, newspaper articles, educational institutions etc. This was especially helpful in international searches.

5. **http://www.doctordirectory.com**. This website provided a list of hospitals, along with an address and telephone number, by city, for every state in the United States. In cases where the doctor lived in a small city, we used this site to track the doctor through his or her hospital affiliation.

6. **http://www.switchboard.com**. Switchboard is an online telephone database. Using the "Find a Business "option, we entered the physician's name, city, and state. When we were unable to find a business address, we searched using the "Find a Person" option by entering the first initial, last name and state. This source was particularly helpful for locating physicians with unusual names.

7. **http://www.healthgrades.com**. This site was used to update business addresses when other sources were ineffective. Since telephone numbers were not listed at this site, it was necessary Accurint was used to obtain telephone numbers.

In addition to these websites, which were the main sources used, we occasionally used the websites listed below:

- http://www.ama-assn.org

- http://www.ashd.com
- http://www.searchsystems.net
- www.dr-411.com
- www.abms.org
- http://www.webmd.com
- http://www.permanente.net

After locating a physician, tracing staff called the telephone number to verify it. They asked to speak to the physician or someone who could verify the physician's full name and primary specialty. In some cases, we were able to confirm reasons for ineligibility (such as deceased, retired and not practicing, federal employee, or resident). Tracing cases were updated bi-weekly throughout the date collection period.

E. REFUSAL CONVERSION

The demanding schedules of physicians often make it difficult to schedule and conduct interviews with them. Because efforts to persuade reluctant physicians to participate in surveys can reduce nonresponse and the risk of nonresponse bias, interviewers were trained to coax these "soft refusals" into reconsidering and participating. A physician who was too busy to be interviewed at the time of the initial call, or a receptionist who said that the physician does not participate in surveys, was coded as a soft refusal. Soft refusals often were coded by the interviewers as callbacks rather than refusals, and were retained by the original interviewer who owned the case. In addition, a team of highly skilled "refusal converters" interviewed physicians who were more adamant—those coded as *hard* refusals. A call was coded as a hard refusal when the physician or office worker became hostile and the interviewer believed that a refusal

conversion specialist might be more successful. A second soft refusal also was assigned to a refusal converter.

If the physician was too busy during the initial call, the interviewer would emphasize that an interview would be rescheduled at the physician's convenience. If the physician could not be contacted, the interviewer would put the case aside for a few weeks and then try again. If a receptionist or other staff member acted as a gatekeeper, the interviewer would call again when that person was likely to be out of the office. In those cases, a different office worker might answer and transmit the call to the physician, or the physician might answer personally and be able to complete the interview.

Often, receptionists or other office staff refuse for physicians, so the physician may not have been aware of the call. In cases where the physician may have refused because he or she was extremely busy at the moment, the refusal was allowed to age for several weeks. The refusal conversion specialist would then prepare for the interview by reviewing notes about prior interactions, which the original interviewer had recorded in the CATI system. The notes enabled the specialist to prepare responses to previously expressed concerns. To prepare for the refusal converter's approach, we mailed or faxed the physician another copy of the introductory letter but did not acknowledge the previous refusal.

Round Four rules used to determine whether a case would be assigned to the refusal conversion team and the level of effort expended on these cases were consistent with prior rounds. Our goal was to maintain a balance between efforts to reduce nonresponse and the need to complete the survey in a reasonable time, and to avoid harassing physicians who clearly did not wish to participate. Although no limit was placed on call attempts, we agreed that a case given to the refusal conversion team (in other words, a case that had received one hard or two

soft refusals) would result in a disposition of a final refusal after one additional physician refusal or two additional gatekeeper refusals.

The Gallup refusal conversion team that was assigned hard refusals and second soft refusals for Round Three consisted of executive interviewers skilled in coaxing receptionists and other gatekeepers to transfer calls to physicians and in fluently addressing physicians' concerns about survey participation, such as burden, sponsorship, study purpose, or data confidentiality.

F. RESPONDENT INCENTIVES

For the first two rounds of the physician survey, all eligible physicians were offered $25 honoraria for participation and mailed checks after completing the survey. The incentive was designed to demonstrate commitment rather than to compensate physicians for their time. We promised the honoraria to physicians who completed the survey rather than mailing checks prior to the initial call because of uncertainty concerning the benefits and costs of mailing checks prior to participation. Many physicians sampled for the first time have incorrect addresses or are ineligible.

We reconsidered this decision for the Round Three panel component because eligibility and participation were likely to be high for this group and we had information from the last survey on current addresses, which reduced financial risk. Our objective in testing prepayment was to increase response rates and reduce cost. Interviewing costs for prepaid physicians would be less than for physicians promised payment if fewer calls were needed to complete interviews. On the other hand, some physicians mailed checks prior to the interviewer's first call could cash them without completing an interview, increasing the cost of prepayment.

Physicians participating in the experiment were randomized to either prepayment or promised payment. Physicians in the prepayment group were mailed a separate letter that referred to the honorarium, and those in the promised payment group received the same letter as

other physicians offered the honoraria after completing the interview. Mailing a $25 check to physicians in the panel component of the CTS survey instead of promising payment had a minimal impact on both response rate and cost, but slightly increased the eligibility rate and the representation of physicians providing patient care on a part time basis. Since the physician survey is designed to track change over time, even a small change in sample composition resulting from a procedure that has a negligible impact on cost and response rate is not desirable. Consequently, we decided against prepaying physicians in the panel component for Round Four and mailed checks to physicians in this part of the sample after they completed interviews. (See HSC Technical Publication No. 45 for more details on the Round Three incentive payment experiment.)

For Round Four, we conducted a second experiment to test the impact of prepayment on response rates and cost for physicians sampled for the first time. The experimental group was mailed a check for $25 along with their advance letter, whereas the control group was promised $25 and mailed checks only after completing interviews. Physicians in the experimental group were only mailed checks if their names and addresses were verified by telephone calls to a member of the practice, typically a receptionist. The weighted response rate for the Round Four experimental group was 49.9 percent, versus 45.0 percent for the control group. Physicians in the new sample who were not selected for the Round Four experiment and all physicians in the overlap sample were mailed $25 checks after completing the interview.

G. SAMPLE DISPOSITION AND RESPONSE RATES

Table IV.1 shows the disposition of the Round Four sample. Overall, 39.8 percent of the weighted sample completed interviews, 12.5 percent were ineligible, 38.5 percent were located nonrespondents, and 9.2 percent could not be located. Compared with Round Three, the percentage of the sample completing interviews declined sharply and the percentage that were ineligible, nonrespondents, or unlocated increased. In Round Three, 47.9 percent of the weighted sample completed interviews, 10.7 percent were ineligible, 35.6 percent were located nonrespondents, and 5.8 percent could not be located (see HSC Technical Publication #38, Table IV.2). During the three years between Rounds Three and Four, these results suggest that more ineligible physicians were on the sample frame, that physicians were more difficult to locate, and that those who were located were more difficult to interview by telephone.

Among the 38.5 percent of located nonrespondents, slightly more than half refused to complete the interview (18.3 percent) or indicated on the AMA Masterfile that they did not wish to be interviewed (2.2 percent). The other major source of located nonresponse was physicians who had been contacted but had not been coded as a refusal by the end of the data collection (14.7 percent in the "end of study" category). Some of the physicians in the "end of the study" category may have been "soft refusals" that were not coded as refusals by interviewers. In any case, repeated callbacks and broken appointments often indicate lack of interest in survey participation and have the same effect as refusals. The remaining located nonrespondents were coded no contact or answering machine (1.7 percent), were ill or had language problems (0.1 percent), or received other codes (1.4 percent).

TABLE IV.1

RESPONSE RATE CALCULATIONS FOR ROUND FOUR
PHYSICIAN SURVEY

	Total Sample				PCP				Specialist			
Disposition	Released Sample	Unweighted Percent	Weighted Count	Weighted Percent	Released Sample	Unweighted Percent	Initial Weighted Count	Weighted Percent	Released Sample	Unweighted Percent	Weighted Count	Weighted Percent
Total Sample	**15,063**	**100.0**	**559,967**	**100.0**	**7,969**	**100.0**	**233,918**	**100.0**	**7,094**	**100.0**	**326,049**	**100.0**
Completed Eligible	6,628	44.0	222,961	39.8	3,426	43.0	88,279	37.7	3,202	45.1	134,683	41.3
Ineligible - Total	1,683	11.2	70,331	12.5	916	11.5	30,335	12.9	767	10.8	39,997	12.2
Retired	312	2.1	15,316	2.7	155	1.9	5,785	2.5	157	2.2	9,531	2.9
Deceased	94	0.6	3,996	0.7	53	0.7	1,984	0.8	41	0.6	2,012	0.6
Other ineligible[a]	1,277	8.5	51,019	9.1	708	8.9	22,566	9.6	569	8.0	28,454	8.7
Located Non-Respondent - Total	5,339	35.3	215,346	38.4	2,756	34.6	88,811	37.9	2,583	36.4	126,538	38.9
AMA refusal[b]	290	1.9	12,507	2.2	173	2.2	6,549	2.8	117	1.6	5,959	1.8
Study refusal	2,490	16.5	102,742	18.3	1,262	15.8	41,172	17.6	1,228	17.3	61,570	18.9
Illness/language barrier	18	0.1	805	0.1	9	0.1	244	0.1	9	0.1	562	0.2
No contact/ answering machine	227	1.5	9,624	1.7	129	1.6	4,534	1.9	98	1.4	5,091	1.6
End of study[c]	2,141	14.2	82,098	14.7	1,099	13.8	33,630	14.4	1,042	14.7	48,468	14.9
Other	173	1.1	7,570	1.4	84	1.1	2,682	1.1	89	1.3	4,888	1.5
Not Located	1,413	9.4	51,327	9.2	871	10.9	26,495	11.3	542	7.6	24,832	7.6

[a]Other ineligible includes federal employee, practicing less than 20 hours/week in patient care, resident or fellow, ineligible specialty, or no longer practicing in U.S.

[b]Physician notified AMA that he or she did not want to be contacted for any surveys; if sampled, their cases were not contacted but included as nonrespondents.

[c]Physician was contacted and was a hard refusal; no further attempts were made. This category may include "soft" refusals, which were coded as callbacks.

The response rate is the proportion of eligible cases providing completed interviews. Since we were not able to determine eligibility for nonrespondents and unlocated physicians, we assumed that the eligibility rate for these physicians was the same as for those who responded (coded complete or ineligible). Therefore, we computed the response rate as the ratio of the sum of completed eligible and ineligible physicians to the total released sample.

For Round Four, the unweighted response rate was 55.2 percent and the weighted response rate was 52.4 percent. For PCPs, the respective Round Three unweighted and weighted response rates were 54.5 percent and 50.7 percent, and for specialists, 55.9 percent and 53.5 percent. Compared with Round Three, weighted response rates declined for all physicians (from 58.6 to 52.4 percent), PCPs (from 57.4 to 50.7 percent), and specialists (from 59.3 to 53.5 percent).

Sample dispositions varied considerably by stratum (Table IV.2); key findings are summarized below.

1. ***Round Three Completed Interviews.*** Among PCPs, approximately three-fourths (75.9 percent) of the physicians who completed Round Three interviews and were sampled for Round Four responded (completed or ineligible interview) to Round Four; 21.6 percent did not respond, and only 2.5 percent could not be located. Results were similar for specialists sampled from Round Three interviews, as 77.6 percent responded, 20.6 percent were located nonresponses, and 1.9 percent could not be located.

2. ***Round Three Noninterviews.*** PCPs sampled from Round Three Noninterviews (refused, ineligible or not located on the last round) were less likely to respond (31.6 percent). About half (52.5 percent) were located but either refused or did not respond for other reasons, and 15.9 percent could not be located. Again, results were similar for specialists, as 32.8 percent responded, 55.6 percent were located but did not respond, and 11.2 percent were not located.

3. ***Round Three Frame Not Sampled.*** As a result of changes in the Round Four sample design, very few PCPs (n=103) who were on the Round Three sample frame were selected for the first time in Round Four. The response rate was very low for this group (35.2 percent), but the sample was too small from which to draw inferences. Among the larger group of specialists on the Round Three sample frame

TABLE IV.2

DISPOSITION OF ROUND FOUR PHYSICIAN SURVEY SAMPLE,
BY SAMPLE TYPE AND SAMPLING CLASSES

Sample Type and Stratum	Cases Released	Response (Complete or Ineligible)		Located Nonresponse		Not Located	
		Count (Unweighted)	Percent (Weighted)	Count (Unweighted)	Percent (Weighted)	Count (Unweighted)	Percent (Weighted)
PCPs							
From Round Three Frame							
Reinterview	3,607	2,723	75.9	789	21.6	95	2.5
Noninterview	2,339	703	31.6	1,228	52.5	408	15.9
Not Sampled	103	38	35.2	53	55.6	12	9.3
From Round Four New	1,920	878	46.3	686	35.4	356	18.3
Specialists							
From Round Three Frame							
Reinterview	2,897	2,232	77.6	613	20.6	52	1.9
Noninterview	1,545	510	32.8	845	56.0	190	11.2
Not Sampled	1,137	486	43.6	540	47.1	111	9.3
From Round Four New	1,515	741	48.7	585	38.4	189	12.9
Total	15,063	8,311	52.4	5,339	38.5	1,413	9.2

selected for the first time in Round Four, 43.6 percent responded, 47.1 percent did not respond, and 9.3 percent were not located.

4. ***Round Four New Frame.*** Both PCPs and specialists who were new to the frame for Round Four were difficult to locate. Among PCPs, 46.3 percent responded, 35.4 percent were located but didn't respond, and 18.3 percent could not be located. The response rate was similar for specialists (48.7 percent), although slightly more were located nonrespondents (38.4 percent) and slightly fewer were not located (12.9 percent).

H. DISTRIBUTION OF THE SAMPLE BY SITE AND TYPE OF PHYSICIAN

Table IV.3 shows the distribution of completed interviews by site of practice and type of physician.

I. DATA PREPARATION

Most of the data coding and cleaning was done by the CATI system. As the interviewers entered response option codes selected by the respondents, these numbers were written to a data file. The CATI system was programmed to conduct range and consistency checks, and to prompt the interviewer when an impossible or unlikely response was entered. The interviewer could then correct the data entry or ask the respondent to clarify the answer.

1. Range Checks

The ranges of most closed-ended items in a CATI survey are determined by codes for the available responses. For example, a "Yes/No" variable offers the following codes:

1 = Yes

2 = No

8 = Don't know

9 = Refused

TABLE IV.3

ROUND FOUR PHYSICIAN SURVEY COMPLETED INTERVIEWS, BY SITE OF PRACTICE AND PHYSICIAN CLASSIFICATION

Site	All Physicians	PCP	Specialist
Total	**6,628**	**3,426**	**3,202**
00-Moved out of the 60 CTS sites	855	401	454
01-Boston MA	219	106	113
02-Cleveland OH	107	54	53
03-Greenville SC	96	42	54
04-Indianapolis IN	101	50	51
05-Lansing MI	79	45	34
06-Little Rock AR	88	40	48
07-Miami FL	94	51	43
08-Newark NJ	106	57	49
09-Orange County CA	96	52	44
10-Phoenix AZ	99	52	47
11-Seattle WA	111	61	50
12-Syracuse NY	87	43	44
13-Atlanta GA	110	53	57
14-Augusta GA	75	38	37
15-Baltimore MD	123	64	59
16-Bridgeport CT	92	49	43
17-Chicago IL	201	94	107
18-Columbus OH	79	50	29
19-Denver CO	102	48	54
20-Detroit MI	134	72	62
21-Greensboro NC	93	53	40
22-Houston TX	120	65	55
23-Huntington WV	72	43	29
24-Killeen TX	69	44	25
25-Knoxville TN	84	45	39
26-Las Vegas NV	84	47	37
27-Los Angeles CA	173	81	92
28-Middlesex NJ	77	41	36
29-Milwaukee WI	86	42	44
30-Minneapolis MN	91	49	42
31-Modesto CA	75	40	35
32-Nassau NY	87	45	42
33-New York City NY	203	107	96
34-Philadelphia PA	197	102	95
35-Pittsburgh PA	87	45	42
36-Portland OR	88	45	43
37-Riverside CA	74	38	36
38-Rochester NY	92	46	46
39-San Antonio TX	96	57	39
40-San Francisco CA	92	46	46
41-Santa Rosa CA	82	40	42
42-Shreveport LA	78	43	35

TABLE IV.3 (continued)

Site	All Physicians	PCP	Specialist
43-St Louis MO	94	52	42
44-Tampa FL	87	51	36
45-Tulsa OK	79	42	37
46-Washington DC	168	96	72
47-W Palm Beach FL	89	44	45
48-Worcester MA	79	39	40
49-Dothan AL	62	28	34
50-Terre Haute IN	51	24	27
51-Wilmington NC	80	35	45
52-W-Cen Alabama	15	14	1
53-Cen Arkansas	68	32	36
54-N Georgia	76	42	34
55-NE Illinois	72	42	30
56-NE Indiana	58	34	24
57-E Maine	77	42	35
58-E North Carolina	80	41	39
59-N Utah	59	38	21
60-NW Washington	80	44	36

If the interviewer mistakenly attempts to enter a code of "3," the CATI system will reject it as an unacceptable code. The interviewer can then enter the correct one.

Some items, such as dates, number of hours worked, or percentages of revenue, do not have a set of preassigned response codes. Ranges are bounded by what is possible. For example, values greater than 100 percent are not accepted for questions requesting percentages of revenue.

2. Consistency Checks

Consistency or logic checks examine the relationships between two or more variables to be sure that the responses do not conflict with one another. A few such checks were contained in the CATI program. For example, question B2 asks the physician how many hours he or she spent in all medically related activities in the past week. Question B3 then asks how many hours were spent in direct patient care that week. If the responses to these two questions are equal, a verification question is asked to ascertain that all of the physician's time was spent in direct patient care. Alternatively, if the physician indicated having spent more hours in direct patient care than in all medically related activities (a logical impossibility), the physician was prompted to revise one or both of the answers to questions B2 and B3.

Section G of the questionnaire includes consistency checks related to practice revenue, which resulted in interviewer prompts. The checks are summarized here; any of the following conditions resulted in an error message to the interviewer:

1. The combined practice revenue from Medicare and Medicaid (and other state sponsored health plans) is greater than 100 percent.
2. The percentage of practice revenue from all managed care contracts is less than the percentage received on a capitated basis.
3. All the practice's managed care revenue is paid on a prepaid basis.

3. Data Cleaning

Additional data-cleaning steps must be completed after the survey leaves the field. Frequencies are examined and cross-tabulations are run to check for additional consistency checks that were not built into the survey. On the basis of these tabulations, data may be changed or flagged for further checking.

4. Coding

As in the first three rounds, only an extremely limited amount of post interview coding was conducted for Round Four. Four questions in Section C permitted entry of "other—list" responses (questions C2, C3c, and C6b) for which the interviewer was to type in any answer that was not provided as a coded response option. Open-ended responses obtained for these questions were examined to determine whether the responses fit any of the categories provided in the question. If they did not, no change was made. If they did, the response the interviewer entered was recoded to the correct response category.

5. Location Coding Review

Physicians were sampled as part of the population of a particular site, and each site was defined as containing a particular set of Federal Information Processing System (FIPS) codes. During the interview, every respondent was asked to confirm the county and state where his or her primary practice was located. Respondents whose practices were not in the county and state shown in the sample record were asked to provide their current county and state.

County and state names were matched against a list containing all the FIPS codes in the country to determine the FIPS code of each physician's current location. We then compared these new FIPS codes (called NEWFIPS in the following text) with the FIPS codes in the sample record to determine whether the physician's practice was located in the site assigned for sample

selection. The following variables were provided in a separate file to document the site locations of physicians who moved between the time of sampling and the time of the interview:

OLDSITE— The site where sampled.

NEWSITE— The site where the physician's practice was located when interviewed. To determine the NEWSITE, we converted the verbatim county and state information to FIPS codes (NEWFIPS) and then matched those against a file that identified whether the code fell into one of the 60 sites or was outside them. If outside the 60 sites, it was coded as site 00. We added codes 98 and 99 to indicate, respectively, "DK/Refused on the county question" (A5a) and "no match found on state/county when compared with the database."

OLDFIPS— The FIPS code of the preferred mailing address provided in the AMA or AOA Masterfiles at the time of sample selection.

NEWFIPS— The FIPS code of the county in which the physician's practice was located when interviewed. These codes were determined by matching the verbatim county and state responses against a file that contains all FIPS codes in the United States.

LOCCODE— 1 = Respondent's practice was in the same site as that assigned for sample selection (sites 1–60).

2 = Respondent was sampled in one site but the physician's practice was in a different site.

3 = Respondent was sampled in the site sample but the physician's practice was outside the 60 sites (site 00).

4 = Respondent was sampled in the site sample but the physician's practice was in a new location, which was unknown.

STCNTY— This field was added to the final Round Four locator database; it contains the two-letter state code linked with the county name that was given by the respondent as the location (state and county) of the physician's practice.

V. SAMPLING AND ANALYSIS WEIGHTS

A. OVERVIEW

The objectives of the study and planned analyses affect the calculation and use of the sampling and analysis weights. In this chapter, we discuss how competing research objectives framed our approach to weighting. We then describe procedures used to compute the various weights, including adjustments for unequal probabilities of selection, modeling to adjust for nonresponse, and post-stratification and ratio-type adjustments.

In the following discussions, we distinguish between sampling weights and analysis weights. *Sampling weights* are calculated from the selection probabilities. Sampling units at each sampling stage have known probabilities of being selected, and the sampling weights equal the reciprocal of the product of these probabilities. We could have used sampling weights alone for our analyses if all the frame definitions had been correct, and if every eligible physician in the sample had been located and had completed a survey questionnaire. However, some of the frame definitions (for example, geographic and physician specialty coding) were incorrect; some physicians could not be located, and others did not participate. We therefore had to modify the sampling weights to account for both errors in the sample frame and for nonresponse. To produce valid study results, we used modified weights, which we refer to as *analysis weights.* Furthermore, because we are interested in several different analysis objectives in each study round, we computed several sets of both the sampling and the analysis weights. Finally, while varying sampling rates by previous survey outcomes increases statistical and data collection efficiency, it also introduces a complexity for the construction of weights. The most problematic situation occurs when the prior round sample and the new physicians on the frame are not adequate to meet the sample size requirement; then, we must return to the prior-round frame to

select the additional physicians. This in turn increases the number of paths by which a physician could be selected for the Round Four sample, and hence the number of probability factors needed to calculate weights.

1. Competing Objectives

Several sets of analysis weights were developed for prior rounds, reflecting the study's analytic objectives (see Table V.1). The site sample in the high-intensity sites was used to support site-level analyses for high-intensity sites. Combined with the low-intensity sites, both sets together comprised a valid national sample. Different site sample weights were developed for site- and national-level analyses because the weights efficient for national analyses were not suitable for site analyses. In the first three rounds of the survey, the supplemental sample was used to develop more efficient national-level estimates, since the supplemental sample was a stratified simple random sample and the associated supplemental sample weights were not based on whether or not the physicians practice was within one of the survey sites.

For prior rounds, several sets of weights were designed to use the site and supplemental samples in combination to produce the most accurate estimates for the individual sites and nationally. All the weights were calculated separately for the two physician specialty categories (PCPs and specialists). Although the equations are the same, the sampling rates differed and reflected the need to oversample primary care physicians. As Table V.1 indicates, fewer sets of weights were needed for Round Four because the supplemental sample was dropped from the study. A set of national-level weights was computed based on only those physicians who practice in one of the 60 sites. This was obtained by dividing the 5,773 site-level weights for physicians who have a practice in the 60 sites by their site selection probabilities. This set of weights would be used for the same analyses as the "augmented site sample weights for national estimates" computed for the previous rounds. The augmented site sample weights referred to the

TABLE V.1

SUMMARY OF ANALYSIS WEIGHTS

Type of Estimate	Sample	Weight Names[a] Round Three	Number of Records With Completed Interviews[b] Round One	Round Two	Round Three	Round Four	Comments
Site-Specific	Site sample (practice in 60 sites)	*PHYWGT1*	10,881	10,434	10,136	5,773	Does not include additional cases from the supplemental sample
	Augmented site sample	*PHYWGT5 (WTPHY1)*	11,456	10,920	10,659	n.a	Best option for site-specific estimates, because site samples include additional cases from the supplemental sample
National	Site sample (all)	*PHYWGT2*	11,310	11,216	11,238	6,628	Does not include additional cases from the supplemental sample
	Supplemental sample	*PHYWGT4 (WTPHY3)*	1,218	1,088	1,168	n.a	Unclustered design, minimal design effect
	Augmented site sample	*PHYWGT7* (WTPHY5)	n.a.	10,920	10,659	5,773	Best option for national estimates when using site-level variables in analysis, because, except in Round Four, it includes additional cases from the supplemental sample
	Combined sample	*PHNATWT1 (WTPHY4)*	12,528	12,304	12,406	n.a	Best option for most national estimates, because it uses all cases from site and supplemental samples
National Panel	Combined sample	*PAN23WTC (WTPAN1)*	n.a.	7,092	8,527	n.a	Includes only those physicians interviewed in both the current and previous rounds
	Site sample (all)	*PAN23WT1 (WTPAN2)*	n.a.	6,569	7,723	4,428	

[a]Name in parentheses refers to variable name on the Public Use File and Restricted Use File.

[b]Some physicians were sampled for both the site and the supplemental samples and are included in each sample, although they were interviewed only once. There were 143 physicians included in both samples for Round One, 24 for Round Two, and 17 in Round Three.

site sample *augmented* by the supplement sample units that fell in one of the 60 sites; the supplemental sample was dropped for Round Four. The augmented site sample shown in Table V.1 excludes physicians practicing outside the 60 sample sites and should be used for national analyses using site-specific information.

Finally, panel weights were developed for longitudinal analyses. These weights were designed to permit analyses of individual changes for physicians who responded to both Rounds Three and Four. These longitudinal analyses can use a model such as the following:

$$(1) \quad Y_{ij} = B_C X_{i(j-1)} + B_L(X_{ij} - X_{i(j-1)}) + e_{ij}$$

where Y_{ij} denotes the observed data for the ith physician at time j, $X_{i(j-1)}$ denotes the value of the independent variable at time j-1 for the ith physician, B_C denotes the coefficient estimate at time j-1, X_{ij} denotes the value of the independent variable at time j for the ith physician, B_L denotes the coefficient estimate of the change between time j, and time j-1, and e_{ij} is the random error term. The first two terms on the right side of the equation are the cross-sectional and the longitudinal terms, respectively, for subject i at time j (Diggle et al. 1999).

2. Focus on Primary Care

PCPs were sampled at a slightly higher rate than specialists to produce approximately equal precision for these physicians and for all physicians who had patient contact (a change from the substantial oversampling of PCPs used in previous rounds). The different sampling rates for PCPs and specialists in the sample resulted in unequal weights and, hence, slightly reduced the survey precision for estimates for all physicians who had patient contact. Because of this disproportionate sampling, the two physician categories were designated as strata to control sample sizes and were used in the nonresponse adjustment models as covariates. Prior to sample selection and interviewing, physicians were classified as PCPs or specialists based on the

sampling frame information for physicians who had not previously been interviewed or based on the Round Three survey response for those who completed Round Three interviews. During the Round Four survey, some of the physicians were re-classified based on information provided by survey responses. However, sampling weights had to ensure that they retained their initial probability of selection, even if they changed specialty classification based on interview data. (See Chapter II for a more detailed discussion on the sampling frame)

3. Geographic Misclassification (Movers)

Physicians in the site sample were to be assigned to the site containing their practice. However, information available at the time of sample selection did not always identify whether the practice was in one of the 60 sites; in some cases, the information on the sample frame may have been the physician's home address. Because practice site was an important analysis domain, some physicians had to be reassigned to a site other than the one assigned at sample selection because the practice site was not known with certainty until the interview. These physicians were called *movers* although most of these physicians merely had a practice location that differed from the "preferred mailing address" on the source files (see Table V.2 for the counts and percentages of movers by site in Round Four and Chapter II for more detail on how this issue was handled in the sample design).

Reassigning practice sites resulted in unequal weighting and complicated the equations used to compute the weights because physicians selected from one sampled site who practiced in another sampled site must reflect probabilities associated with both sites (referred to as *joint*

TABLE V.2

ROUND FOUR PHYSICIAN SURVEY EXPERIENCE WITH MOVERS

Site	Classified in Site, by Frame	Stayed in Site	Moved out of Site	Weighted[a] Percent Moved out of Site	Moved into Site	Weighted[b] Percent Moved into Site
Total	**6,629**	**5,322**	**1,307**	**19.0**	**452**	**16.2**
01-Boston MA	233	197	36	15.9	22	12.1
02-Cleveland, OH	121	101	20	18.1	6	6.6
03-Greenville SC	101	92	9	8.1	4	20.0
04-Indianapolis, IN	116	95	21	17.2	6	9.6
05-Lansing MI	111	78	33	31.0	1	6.7
06-Little Rock AR	110	85	25	20.9	3	23.7
07-Miami FL	105	88	17	17.2	6	8.9
08-Newark NJ	129	93	36	31.2	13	33.2
09-Orange County CA	116	88	28	25.7	8	16.4
10-Phoenix AZ	104	91	13	13.7	8	13.0
11-Seattle WA	106	93	13	13.4	18	28.8
12-Syracuse NY	111	87	24	20.2	0	0.0
13-Atlanta GA	97	86	11	20.9	24	29.0
14-Augusta GA	100	71	29	31.3	4	31.3
15-Baltimore MD	146	110	36	28.1	13	15.0
16-Bridgeport CT	117	86	31	26.3	6	19.4
17-Chicago IL	213	188	25	12.3	14	8.2
18-Columbus OH	104	79	25	28.3	0	0.0
19-Denver CO	107	95	12	10.8	7	14.4
20-Detroit MI	147	126	21	13.8	8	10.0
21-Greensboro NC	109	88	21	25.0	5	16.6
22-Houston TX	127	108	19	15.6	12	14.0
23-Huntington WV	99	72	27	34.3	0	0.0
24-Killeen TX	105	67	38	34.8	2	55.8
25-Knoxville TN	101	84	17	19.1	0	0.0
26-Las Vegas NV	86	75	11	13.1	9	59.7
27-Los Angeles CA	168	144	24	13.8	29	13.9
28-Middlesex NJ	90	63	27	27.8	14	31.9
29-Milwaukee WI	100	84	16	21.4	2	5.7
30-Minneapolis MN	99	86	13	15.2	5	7.2
31-Modesto CA	89	74	15	20.4	1	32.3
32-Nassau NY	120	82	38	33.3	5	5.8
33-New York NY	177	141	36	22.7	62	32.1
34-Philadelphia PA	218	182	36	16.9	15	9.0
35-Pittsburgh PA	101	83	18	18.3	4	8.0
36-Portland OR	91	79	12	14.1	9	17.8
37-Riverside CA	82	65	17	24.1	9	29.4
38-Rochester NY	106	91	15	13.8	1	1.2
39-San Antonio TX	111	91	20	16.8	5	15.7
40-San Francisco CA	112	81	31	25.2	11	24.9
41-Santa Rosa CA	98	79	19	18.2	3	25.2
42-Shreveport LA	105	75	30	33.5	3	22.6
43-St Louis MO	103	90	13	19.8	4	7.7
44-Tampa FL	93	79	14	19.0	8	18.0
45-Tulsa OK	100	79	21	22.0	0	0.0
46-Washington DC	160	141	19	14.6	27	15.9
47-W Palm Beach FL	108	85	23	23.0	4	19.1
48-Worcester MA	104	73	31	28.5	6	29.4
49-Dothan AL	72	60	12	20.6	2	70.6
50-Terre Haute IN	70	51	19	25.6	0	0.0
51-Wilmington NC	101	80	21	19.6	0	0.0
52-W-Cen Alabama	24	15	9	39.2	0	0.0

TABLE V.2 (continued)

Site	Classified in Site, by Frame	Stayed in Site	Moved out of Site	Weighted[a] Percent Moved out of Site	Moved into Site	Weighted[b] Percent Moved into Site
53-Cen Arkansas	73	64	9	12.5	4	34.3
54-N Georgia	98	71	27	26.4	5	58.4
55-NE Illinois	83	67	16	19.1	5	115.0
56-NE Indiana	70	56	14	20.8	2	91.3
57-E Maine	95	75	20	21.6	2	20.3
58-E North Carolina	93	77	16	17.2	3	27.1
59-N Utah	97	59	38	42.6	0	0.0
60-NW Washington	97	77	20	19.3	3	30.0

[a]The weighted percentage of physicians who moved out of site is computed as the ratio of the sum of the weights of the physicians who moved out of site over the sum of the weights of physicians in that site in the frame.

[b]The weighted percentage of physicians who moved into a site is computed as the ratio of the sum of the weights of the physicians who moved into a site over the sum of the weights of physicians in that site (where the site is defined by the new interview data).

inclusion probabilities). The sampling weight for these cases, therefore, sometimes differed substantially from the weight for the other physicians practicing in the same site.[1]

4. Longitudinal Versus Cross-Sectional Estimates

Because the CTS Physician Survey is a repeated survey with a longitudinal component, the Round Four sample will be used to provide both cross-sectional and change estimates. As discussed in Chapter II, part of the sample was also interviewed in previous rounds to improve the precision of both change and cross-sectional (point-in-time) estimates.

Weighting for longitudinal surveys is complex because the inclusion probabilities are defined not only on the current conditional selection probabilities, but also on when the physician first entered the sampling frame and whether the physician was selected in one or more previous rounds (see Appendix B). Finally, panel weights for the reinterviewed physicians required adjustments so that the panel weights reflected the estimated population distribution of the eligible Round Three population.

5. Analysis Weights

Unbiased estimates are the goal of any well designed survey. However, some of the physicians sampled for the CTS Physician Survey could not be located, and others who were located refused to participate or did not respond after many calls. Using logistic regression models based on available data from the sampling frames (for all physicians) and from the prior round (for reinterviewed physicians), we developed adjustment factors for the sampling weights to reduce the potential for bias by compensating for the physicians who could not be located and for nonresponse among located physicians. We refer to these weights as the *analysis weights.*

[1]Extremely large weights may be trimmed to improve the precision for the site-level estimates. However, we minimized weight trimming to avoid introducing significant bias into the survey estimates.

Separate multivariate models were developed to adjust the weights for unlocated and nonresponding physicians in the sample.

6. Weights Used

The calculation of sampling and analysis weights was both influenced and complicated by limitations of the sampling frame (for example, missing or incorrect information from the AMA and AOA files used to create the frame) and the need to use unequal sampling rates. In addition, the analytic objectives required the calculation of several sets of analysis weights. The various weights include those needed for:

- National-level estimates for the full site sample
- National-level estimates based on the physicians who practice in one of the 60 sites
- Site-level estimates for computation of covariates for use in modeling
- National-level panel analyses (based on physicians who responded in both Round Three and Round Four).

Table V.1 summarizes the analytic uses of the weights and the final adjustments.

B. COMPUTATIONAL METHODS

1. Overview

The sampling and analysis weights had one component in common—the weight was calculated as the reciprocal of the inclusion probability of the physician. This weight was based on the site weight and one or more conditional weights (based on reciprocal selection probabilities). As Table V.3 shows, the sets of weights were computed to serve different analytic objectives. Because the equations for each weight were complex, only a few basic examples are presented here. The process for adjusting the sampling weights to account for unlocated physicians and nonresponse was complex and included nonresponse adjustments (including

TABLE V.3

POSTSTRATIFICATION AND RATIO-TYPE ADJUSTMENTS
FOR NATIONAL AND SITE ESTIMATES WEIGHTS:
ROUND FOUR PHYSICIAN SURVEY

Weight Name	Analytic Purpose of Weight	Poststratification and Ratio-Adjustment Methodology
PHYWGT1	Site-level estimates	Weights were ratio-adjusted to projected count of eligible physicians on a 120-cell basis. The 120 cells defined by combinations of PCP/specialist status and site membership and ratio-adjusted to agree with the 12 (age: less than 45, 45-64 years, 65 or older; primary care physician or specialist; gender) totals of PHYWGT2.[a]
PHYWGT2	National estimates, full sample, best weights for national level estimates	First level of poststratification is based on two cells defined by Round Three frame versus new physicians to Round Four frame; within the two initial cells adjustment groups were age (3), PCP/specialist (2), and gender (2)
PHYWGT3	National estimates from sample within the 60 sites, weights for national analyses using site-level information	PHYWGT1 divided by PSU (Site-selection) probabilities; agree with the12 totals of PHYWGT2 age (3), PCP/specialist (2), and gender (2).

[a]For consistency with the changes in sites, site membership was defined as the physicians' site membership reported during the interview, rather than as the site membership at the time of sample selection.

separate treatment of unlocated physicians and nonresponding physicians who were located), poststratification, and weight trimming.

2. Probability of Selection

Sampling weights were essential for calculating unbiased statistics from the survey data and for conducting valid analyses. To calculate the weights, the inclusion probabilities had to be calculated for each record on the data file.

As noted, the entire sample, including movers, was used to develop weights for national estimates. The sample was a two-stage probability sample drawn from the national frame (that is, from the population of all physicians in the defined target population). For national estimates, the calculation of the inclusion probability (P_i) for any sampled physician accounted for the selection probability of the site and the selection probability of the physician in the site.

To illustrate, for the Round One sample (the simple case), the sites were randomly assigned as high intensity sites and low intensity sites and the sample size for the high intensity sites was approximately 4 times the sample size for the low intensity sites (see Technical Publication #9). Therefore, the probability of selection (P_i) of a physician sampled within a site was calculated according to the following equation:

$$(2)\quad P_i = P(site)*P(i|site)$$

$$= P(site)*[P(HI)(n_{HI}/ N_s) + (1 - P(HI))(n_{LO}/N_s)],$$

where $P(site)$ was the probability that a site was selected, N_s was the sampling frame size, $P(HI)$ = 12/48=1/4 for the 48 large metropolitan sites and = 0 for the other sites, and n_{HI} (n_{LO}) was the sample size that would have been allocated to a site if it was chosen as a high- (low-)intensity

site.[2] More specifically, to use equation (2), we had to estimate the sample size that would have been released under our original sample allocation plan, treating each site first as a high-intensity site and then as a low-intensity site (since this was a stochastic assignment). The process was required for each of the four sampling strata used in Round One of the study (PCP or specialist by frame source [AMA or AOA]) within each of the 48 large metropolitan sites.

At this point, we ignore the issue of physicians whose geographic or patient care classification was misassigned by the frame. (This issue was discussed in Chapter II) In this example, we also ignore the fact that large MSA sites were randomly assigned as high- or low-intensity sites, in order to simplify the discussion. In Round Two and subsequent rounds, these calculations must also reflect probabilities and response status relating to previous points in time.

Consider that a physician could be selected for Round Four via several paths, which were used to develop four frame strata:

1. Physician was eligible and completed a Round Three interview. The paths include (a) a Round Three eligible *complete* that could have been or not been selected at Round Two; if selected at Round Two, they may have been either a noninterview or an interview, or (b) could have been new on the Round Three frame
2. Physician was selected in Round Three but did not complete the interview, for example, was ineligible, could not be located, or refused, (a Round Three *noninterview*; same alternate paths as described for stratum one)
3. Physician was not selected in Round Three but was in the Round Three frame (an *old* frame physician; same alternate paths as described for stratum one)
4. Physician was not in the Round Three frame (a *new* frame physician)

[2]In Round Four, the site classification of high or low intensity was not germane to the survey. However, because the sampling weights were computed from the selection probabilities, which were computed based on a physician's probability of selection in prior rounds, these selection probabilities were affected by the high intensity and low intensity site classification from the prior rounds. Therefore, the site classification used in the prior rounds is a component of the computation of the sampling weights in Round Four.

If we consider the chain of events for the Round Two physicians selected from the Round One population, we have two possible routes, *a* (was selected in the Round One sample) and *b* (was not selected in the Round One sample). For simplicity, we ignore at this point the allocation to high- and low-intensity study and frame misclassifications:

(3) $P(a) = P_{site} * P_{11} * P_{2j}$

and

(4) $P(b) = P_{site} * (1 - P_{11}) * P_{23}$,

where:

P_{site} = the (unconditional) probability of selecting the site.

The conditional probabilities are defined as P_{ij}, *i* relates to Round One (i=1), Round Two (i=2), or Round Three (i=3), and *j* relates to the frame sampling groups for Round One: 1 for primary care physicians and 2 for specialists and for Rounds Two through Four: 1 to 4 for primary care (and 5 to 8 for specialists, reflecting the different selection probabilities of PCP and specialists).

$P_{1j'}$ = the conditional probability of selecting the physician in Round One given the site was selected. For Round One, only two strata were used, primary care physicians (j' =1) and specialists (j'=2).

P_{2j} = the conditional probability of selecting the physician in Round Two given the physician was an eligible respondent in the Round One sample (*j* = 1 or 5), or the physician was selected but not an eligible respondent in round One sample (j=2 or 6), or the physician was not selected in Round One (but was in the Round One frame) (j=3 or 7).

The inclusion probability for Round Two, *P*, equals the sum of probabilities for occurrence in one or the other of two disjoint events. That is, $P = P_{site} * \{P_{1j'} * P_{2j} + (1 - P_{1j'}) * P_{23}\}$, *where j' = 1 or 2, and j=1, 2, 4 or 5*. The probability for a new physician (not on the Round One frame)

is simpler because of the shorter path. Conversely, while the basic process is the same, the probabilities for a physician selected in Rounds Three and Four, are more involved because the paths can be longer. Consider, for example, a physician that was selected in Round One, not selected in Round Two, not selected in Round Three, but selected again in Round Four. The probability for Round Four is a function of the current probabilities as well as those in all prior rounds. This path history was retained on file to facilitate the weight calculations at each subsequent round.

The full equations used to calculate the Round Four weights are in Appendix B.

C. LOGISTIC PROPENSITY MODELS FOR NONRESPONSE ADJUSTMENTS

The purpose of nonresponse adjustment to sampling weights is to reduce the potential for bias associated with nonresponse. If nonresponse to a survey is completely random, then weighted estimates of means would be unbiased and nonresponse adjustment would not be required. For estimating population totals, however, a single adjustment still would be needed to inflate a weighted total to account for the proportion of physicians who did not respond. However, nonresponse is rarely completely random, and it is possible to ascertain patterns about characteristics of sampled individuals, such as physicians, who do or do not respond. For the CTS Physician Survey, the concept underlying nonresponse adjustments is to develop two types of logistic regression models, which can predict (1) the probability of locating a physician (location propensity score) and (2) the probability that the located physician will complete the interview (response propensity score). Then, we computed an adjustment value for each physician who completed the interview. The adjusted weight for nonresponse (sometimes referred to as the non-response-adjusted analysis weight) is simply the multiplication of the inverse of the location propensity score, the inverse of the response propensity score, and the sampling weights (the inverse of the probability of selection).

A key determinant in developing the logistic regression models is the availability of information for both respondents and nonrespondents. In many surveys, limited information is available beyond that used for creating sampling strata. For the CTS Physician Survey, however, we have considerable information that can be used to adjust for nonresponse from the sampling frames and the Round Three survey. For nearly all sampled physicians, demographic and practice characteristics are available from the AMA and AOA files used as the sample frame. We also have an extensive array of variables from the Round Three survey for physicians who completed interviews in Round Three. In addition, for nonrespondents and unlocated physicians in Round Three selected for Round Four we have data on survey dispositions (that is whether the physician was located, refused, or ineligible) for both rounds.

Logistic propensity modeling has been used for surveys where information on the characteristics of both respondents and nonrespondents is available. For example, this approach was used for the National Survey of Family Growth (Potter et al. 1998), in surveys of military personnel (Iannacchione et al. 1991) and in surveys of Medicare and Medicaid populations for which demographic and economic data are available from federal or state administrative files (CyBulski et al. 1999).

The steps used in adjusting for nonresponse are (1) examining patterns of nonresponse and (2) developing adjustment factors that are assigned to each respondent to compensate for nonrespondents. The following sections describe how the models were developed; we then describe the weight adjustment procedures.

1. Examining Patterns of Nonresponse

First, we examined nonresponse patterns for sample members. For this survey, we had different levels of data for three subgroups based on their Round Three interview status.

1. ***Round Three Interviews (Reinterviews).*** Physicians who completed the Round Three interview
2. ***Round Three Noninterviews.*** Physicians selected for the Round Three sample but who did not complete the interview (including physicians who could not be located, who refused to be interviewed or who were ineligible)
3. ***New Sample.*** Physicians in the Round Four sampling frame who were not selected for the Round Three sample

Location and nonresponse patterns vary by strata—reinterviews, noninterviews, and new sample—(see Appendix C for more detail on response patterns). In addition, the level of information on physicians that could be used to model location and response patterns varied by strata. Reinterviewed physicians had information from the Round Four sampling frame, responses from the Round Three instrument, and information from Round Three survey dispositions (such as location and the record of calls for the Round Three interview). For Round Three noninterviews, we had information from the Round Four sampling frame and survey dispositions, such as location and response status and the record of calls for Round Three. Only information from the Round Four sampling frame was available for *new* sample. By separately modeling each stratum, we are better able to take advantage of differences in available information in explaining location and response patterns.

The location rate for the reinterview stratum is higher than for the other two strata because these physicians completed Round Three interviews and recent contact information was available for them. Information on addresses for physicians contacted for the first time or noninterviews from the prior round were less accurate. Less than two percent of the reinterview sample could not be located; approximately fifteen percent of the noninterview and new physician samples could not be located.

Among located physicians, the nonresponse adjustment factors were also much larger for the samples of physicians new to the sampling frame and noninterviews (which included physicians

who refused in the prior round) than for the reinterview physician sample. Among located physicians, approximately 76 percent of the reinterview sample responded, but only 54 percent of the new physician sample and 37 percent of the noninterview samples responded.

2. Developing Adjustment Factors

To estimate the adjustment factors for locating a physician and for responding among located physicians, we used unweighted logistic regression models to estimate a "response propensity" score for each physician. In general, the use of a logistic regression modeling approach to computing nonresponse adjustments can result in a few sample members being assigned large adjustment factor (Little 1986), relative to the classical weighting class approach[3]. Moreover, the use of unweighted response propensity modeling can result in more variation in the adjustment factors than the weighted logistic response propensity modeling used in Round Two and Round Three. However, the possibility of large adjustment factors can be reduced by using a restricted logistic regression model[4] or by trimming and compensating for adjustment factors from an unrestricted logistic regression model in a sample alignment or poststratification adjustment process. We used the latter approach.

The model-based nonresponse adjustments are predicted values (based on maximum likelihood estimators that are *consistent*, asymptotically *efficient*, asymptotically normal, and therefore, asymptotically *unbiased*) and were used in the computation of different sets of analysis weights. That is, the model-based propensity scores developed for the full sample were used to

[3]The logistic regression modeling can take advantage of the statistical measures to evaluate the fit of the model to the observed pattern of response, whereas the classical weighting class method does not have a series of comparable measures.

[4]The coefficients of the model are estimated based on restrictions on the size of the adjustment factor.

account for the inability to locate a physician and for physician nonresponse in the computation of weights for site, national and panel estimates.

After computing adjustment factors for the inability to locate a physician and for nonresponse among located physicians, various sets of weights were computed. These adjusted weights were then checked for consistency with known (or estimated) population counts of eligible physicians and were post-stratified to control totals. We evaluated the few extreme weights, which could have seriously decreased the precision of the survey estimates and analysis, and trimmed some of them to improve precision while potentially introducing some bias into the estimates. The following section describes weight adjustment procedures and construction of analysis weights in more detail.

D. RESPONSE PROPENSITY MODELS

1. General Model Development

We prepared two sets of unweighted logistic regression models to adjust the survey weights to compensate for the inability to locate physicians and the inability to obtain a response (either a completed interview or ineligibility determination) among the located cases. We considered separate models for location and response for physicians for (1) Round Three completed interviews among eligible physicians, (2) Round Three noninterviews, and (3) physicians not in the Round Three sample (new). In total, we developed 5 models—separate location and response adjustments for Round Three noninterviews and the new sample and a combined location/response model for reinterviews. We used one model for reinterviewed physicians because 98 percent of the reinterview sample was located.

Each model was used to estimate a propensity score for location or for response among located cases as a function of physician characteristics represented by a series of indicator variables. The sampling weights were used as independent variables in the location regression

models and the sampling weights adjusted for not locating a physician were then used as covariates in the response regression models.

Unlike previous rounds, the logistic models were unweighted, with the sampling weight used as an independent variable in the respective model. This change in methods was motivated by the need to estimate conditional probabilities at each stage subsequent to the initial stage. That is, for location, we need an estimate of probabilities conditional on the sample that was selected, rather than an estimate of probabilities conditional on the entire population. In a paper by Williams et al (2004), unweighted models were shown to be preferable to weighted models for a prior round of the CTS survey. Although we used a different methodology, Little and Vartivarian (2003) showed that the use of unweighted model estimation for the purposes of nonresponse adjustment was more effective than weighted estimation (that is smaller increase in variance and less potential for bias).[5] We expect that the change in methodology will reduce the magnitude of nonresponse bias and, hence, may introduce a slight shift in the estimated trend. However, relying on a larger bias to be consistent across rounds has the potential to be even more of an artifact because bias is a function of the response rate and the difference between respondents and nonrespondents in the value being estimated.

[5]While the article by Little and Vartivarian focuses primarily on weighting class adjustments, they extend the results to our propensity modeling problem to cover the situation when the number of classes is too large and/or have too few observations per class. This study is a simulation study using a "Monte Carlo" (re-sampling) technique but the results seem quite strong. The comparison of the weighted and unweighted logistic regression response propensity modeling results and weighting-classes results are compared to the classical statistical estimation method called maximum likelihood estimation (MLE). This is reasonable since the random numbers in their paper are generated for known distributions reflecting a range of different assumptions about relationships between the observations of interest and the design and weighting classes, under the specification of random sampling with unequal sampling rates. As expected from the classical statistical theory, the MLE method produces the best estimates when the correct distribution is involved, but for weighting classes and logistic models, the unweighted results are better than the weighted. Although a simulation study does not give rigorous proof, the authors offer the explanation that while the weighted logistic will produce unbiased estimates of response rates for the population, when we are really seeking unbiased estimates for the variables of interest.

After reviewing the results from our nonresponse analysis (see Appendix C), we began by including all of the characteristics in the location and nonresponse models (referred to as the full models). Many of them were multi-level categorical responses (for example, specialty), and we transformed these multi-level variables into a series of binary indicator variables. The categories were chosen depending on the number of observations in each category and the different location or response rates in each one. We collapsed categories with similar location and response rate patterns. For a few variables, we modified the indicator variable definitions depending on whether they were used for the locating models or for the response models. In addition, we combined variables with missing information (for example, unknown country of medical school) with other categories or created an indicator to denote a status of missing.

The variables used in the location and response logistic regression models included: age, board certified, country of medical school (in the US and Canada vs. all other locations), gender, specialty, present employment, income (for the reinterviewed physicians), Round Three disposition code (for physicians who were Round Three noninterviews), region, whether the preferred mailing addresses was in a urban or rural county, and design variables such as sampling weights. Besides these variables, we evaluated second and third order interactions and included these interactions if they were significant in the model to explain the variation in the response propensities. We developed "nested" models in the sense that if a third-order interaction was significant all second-order interactions within a third-order interaction were included in the model regardless of their significance.

To identify possible interactions among the variables, we used a tree algorithm method, the Chi-squared Automatic Interaction Detector (CHAID), to identify potential interaction terms involving the independent variables. The identified variables (main effects) and interactions among these variables were entered into the model using a forward and backward stepwise

logistic regression, rejecting only those main effects or interactions that have less than 0.30 significance. As noted previously, this model is then expanded to include any main effects or interactions that are involved in a higher order interaction but missing from the model. At this point the model contains a number of covariates that are marginally significant that have large coefficients and standard errors. These covariates can have a specious impact on some of the sampling weights, so the model was reduced to those covariates that were significant at the 0.10 level or less according to the Wald Chi-Square test (except for those belonging to a significant higher order interaction term).

In the following sections we discuss each of the models and some of their measures of goodness of fit and (propensity) predictive ability. One of the statistics we present for each model is the re-scaled R-squared value that is a measure of predictive power. We emphasize, however, that small R-squared values are the norm in weighted logistic regression, but these unweighted models for Round Four have relatively large R-squared values. Nevertheless, these values cannot be interpreted the same as those in linear regression (Hosmer and Lemeshow 2000).

The goodness-of-fit tests indicated mixed results for the models. These statistics assess whether or not the model contains the variables and interactions that should be in the model and that they are included in the correct functional form. The goodness-of-fit statistics reviewed include the Chi-square test for the Global null-hypothesis that none of model coefficients are significant, the Pearson Chi-square, the Deviance statistic, and the Hosmer-Lemeshow (H-L) statistic. These measures and tests are fundamentally variations of Chi-Squared tests. The reason for such mixed results for our models is that the tests are not valid if the number of unique covariate profiles is large compared to the number of observations. The large number of covariates in our models resulted from the multi-level covariates entered as binary indicator

variables. The H-L statistic was developed to solve the problem faced by the other goodness of fit statistics, but we concluded that even this statistic seems problematic when the range of the estimated propensity scores was very high in a specific range and very low in the remaining ranges for the scores (that is the distribution of the propensity scores has a very large kurtosis coefficient and results in a wide range of frequencies within some of the H-L deciles or cells).

In addition to the goodness of fit statistics, several statistics are used also to assess the predictive ability of the models. These statistics were very good for all of the models and are important measures for our purposes because we wish to accurately predict the propensity on the basis of the covariates. The R-squared values, a logistic regression analogue to the statistic in linear regression, were much larger than usually experienced for weighted logistic regression, ranging from 0.55 to 0.60 for the two location models (Round Three noninterview and new physicians) and for the three response models (noninterview, reinterview, and new). The other statistics that were reviewed to assess the predictive ability of the models are all based on these concordance and discordance values. The concordance and discordance statistics are obtained by looking at all possible pairs of observations excluding those that have the same value for the observation variable. For a pair of sample members, the pair is concordant if the respondent physician (coded a one) has a larger predicted value than the nonrespondent physician (coded zero), and if not, the pair is discordant. A good model has a considerably larger percent of concordant (C) than discordant (D) pairs. Other statistics of predictive power (Tau-a, Gamma, Somer's D, and c) are functions of concordance and discordance. All of our models were strong for predictive ability based on these statistics.

2. Location Weight Adjustments

The location models were used to produce a location propensity score. The weight adjusted for location is obtained by multiplying the sampling weight and the inverse of the location

propensity score. These adjustments inflate the weights of the located physicians to compensate for those physicians who were not located. The percentage of the sample that could not be located increased in Round Four (9 percent) compared with Round Three (5 percent)[6]. The overall unweighted response rate (accounting for both unlocated and nonresponse) was 55 percent and the weighted response rate was 52 percent (see Chapter IV for more detail on the disposition of the sample).

For the three categories of physicians (reinterviewed physicians, Round Three noninterviews and physicians new to the sampling frame in Round Four), physicians who were difficult to locate had one or more of the following characteristics.

- Females
- Young physicians (under 40)
- Not board certified
- Doctors of Osteopathy (DOs)
- Physicians with no address in the AMA or AOA files for Round Four
- Physicians sampled in the following metropolitan areas: Orange county, Los Angeles and Riverside in California, New York City, Detroit
- Physicians who requested AMA not to release their names for surveys (no-contact physicians)
- Physicians with no more than two calls, which included ineligible physicians and physicians with no contact information
- Cases receiving a final code of Wrong Number, 60 percent of which had a final disposition of unlocated
- Cases with no phone given in the AMA and AOA data files
- Physicians who were only in the AOA data file

[6]In Rounds Two and Round Three, AMA provided social security numbers for the individual physicians. Social Security Numbers were not made available for Round Four.

- Physicians who were in the sampling frame in prior rounds and not the current AMA or AOA data files
- Physicians who were noninterview cases in Round 3 or who were new to the sampling frame in Round Four (very few unlocated cases among the reinterview, 2 percent)

a. Reinterview Physicians

A separate model for locatability was not used for reinterview physicians because only two percent were not located. Rather, a single model was developed to adjust for both unlocated and nonrespondent physicians.

b. Round Three Noninterview Physicians

A total of 3,286 physicians out of 3,884 (86.5 percent) noninterview physicians were located. The noninterview physicians included physicians in the Round Three sample who could not be located, who refused, or who were ineligible at the time of data collection. The full model for location for the physician sample included main effects and interactions using an inclusion criterion of significance set at alpha =0.30 for variables: age, census region, gender, location of medical school (US/Canada vs. all other locations), number of calls attempted, Round 3 survey final disposition (unlocated, refused or ineligible), an urban/rural indicator, Round 4 initial sampling weights, site when sampled, and board certification. This full model had a "rescaled" R-squared (R^2) 0.60.[7] For the final reduced model, we reduced the criterion for including a variable to alpha = 0.05; this resulted in a model with 27 covariates, including 6 interactions. This "reduced" model had an R^2= 0.57. The R-squared of this model is very high because, in part, the interaction terms are very significant variables in the model. The goodness of fit

[7]The R-squared values presented are the "maximum-rescaled" values because the generalized R-square in logistic regression does not have an upper bound of 1 as in ordinary least squares analysis.

statistics were inconsistent, possibly because of a large number of unique covariate profiles[8], but all of the measures of predictive power were very strong (the percentage of concordant pairs was 91.9 percent compared to only 7.7 percent for discordant pairs).

c. New Physicians

A total of 4,007 out of 4,675 (87 percent) new physicians were located. The full model for propensity to locate physicians who were not in the Round 3 sample includes main effects: age, gender, location of medical school (US/Canada vs. all other locations), present employment, specialty (PCP vs. specialist), urban/rural indicator, length of time since released to the sample, number of calls attempted, physician type (MD vs. DO), and board certification, and second, and third order interactions. For this sample, the "rescaled" R^2 for this model was 0.58. The goodness of fit statistics and all of the measures of predictive power were very strong (in addition to the R-squared value, the percentage of concordant pairs was 92.3 percent and the percentage of discordant pairs was 7.4 percent).

3. Response Weight Adjustments

The response models provide the probability that a located physician completes the interview (response propensity score). The final weight adjusted for nonresponse is obtained by multiplying the weight adjusted for location and the inverse of the response propensity score. These adjustments inflate the weights of the physicians who completed the interview to compensate for those physicians with similar characteristics who did not complete the interview.

[8]Each sample physician is associated with a vector of covariate values, if these vectors are identical for many of the physicians (i.e., relatively few unique covariate profiles) then the number of unique profiles is much smaller than the sample size. If the number of unique profiles is large, some may be linked to only one or two physicians, causing the chi-squared tests to be invalid—as in contingency tables with few observations per cell.

In general, the characteristics of physicians who did not respond were:

- Physicians who graduated in a foreign county
- Physicians with no address in the 2003 AMA or AOA data files
- Physicians sampled in Chicago, Detroit, Houston, Knoxville, Las Vegas, Los Angeles, Nassau County, NY, New York City, Riverside CA, Shreveport, Tampa, Tulsa and West Palm Beach
- Physicians who were attempted for interview for five or more months
- Physicians who were DOs and were not listed in the AMA data file

a. Reinterview Physicians

A total of 4,995 out of 6,504 (76.9 percent) physicians who were interviewed in Round Three responded in Round Four. The initial full model for response included: location of medical school (US/Canada vs. all other locations), present employment, income in 2000 (income categories based on Round Three survey responses), board certification, geographic division of the country, site, and urban/rural indicator, and second and third order interactions. This initial model had 52 covariates and the "rescaled" R^2 for this model was 0.62.

The reduced model (using the criterion for including a variable set at alpha = 0.05) had an R^2= 0.61, and the model included 36 covariates including 3 interactions. This final logistic regression model for reinterview response showed that the higher response rates are among physicians who are 50 to 65 years old, who graduated from a U.S. or Canadian medical school, and who are board certified. The covariates reflecting the length of time the case was worked during the data collection period were important in the model both as covariates and interactions. Several sites were significant covariates, with San Antonio as a high response site and Miami, Orange County, CA, and NE Indiana having relatively low response rates.

b. Round Three Noninterview Physicians

A total of 1,213 physicians out of 3,286 (37.3 percent) physicians in the Round Three noninterview sample were located but did not complete the interview. A full model for response included main effects: age, board certification, Round Three survey disposition, the length of time that the sample members were worked during the data collection period, number of calls attempted, site, and the availability of a phone number from the sample frame sources, and second and third order interactions was fit with an $R^2 = 0.56$. The reduced model also with an $R^2 = 0.56$ and this model included 36 covariates including 9 interactions. All measures of model goodness of fit and predictive power were very good.

The final logistic regression model for response showed higher response rates in Cleveland, Denver, and Rochester. The physicians who were located but did not respond to the Round Three survey were also less likely to respond in Round Four than were other Round Three nonterviews.

c. New Physicians

A total of 2,143 out of 4,007 (52.6 percent) of the located physicians who were new to the sample in 2004-05 completed an interview in Round Four. The full model for the propensity to respond contained 58 covariates, including 12 interactions and had a rescaled R^2 of 0.56. Main effects included: age, present employment, specialty, the availability of a phone number from the sample frame sources, number of calls attempted, geographic division, and site, and second and third order interactions. The reduced model had a rescaled R^2 of 0.55, and this model had 40 covariates including 6 interactions. The goodness of fit statistics were mixed but the predictive capability of the model is good based on the Hosmer-Lemeshow test statistic. For example, the percentage of concordant pairs was 88.5 percent and the percentage of discordant pairs was 11.3 percent.

E. FINAL COMPUTATION OF THE WEIGHTS

The objectives when computing the national weights are (1) to minimize the risk of introducing bias to the sample estimates, and (2) to reduce the design effect of the sample estimates. After applying the nonresponse adjustments to reduce the potential for bias resulting from sample attrition, post-stratification procedures adjust the weights so that the survey totals match the population totals of the Round Four frame. Next, some of the extreme weights are trimmed to reduce large design effects while minimizing the mean square error (which is a function of both bias and variance). Actually, multiple iterations of post-stratification and trimming were used because the post-stratification adjustment sometimes caused large weight values.

In the following sections, we describe our approach to post-stratification, trimming, site weights, and panel weights.

1. Post-Stratification and Ratio-Type Adjustments

After applying the adjustments to the weights for unlocated physicians and for nonresponse among located physicians, the weighted counts for physicians who completed the interviews or who were ineligible did not reproduce the Round Four frame totals for some of the primary analytic domains of PCP/specialists and sample source. Therefore, we computed a ratio-type adjustment so that the sum of the nonresponse-adjusted weights matched the frame counts, before adjusting for geographic misclassification. In general, these adjustments were the frame count for a group divided by the corresponding sum of the nonresponse-adjusted weights for the completed and ineligible interviews in the group. Table V.3 presents the cell definitions used to post stratify or ratio-adjust each type of survey weight.

Classification as a PCP or specialist was used in all the poststratification adjustments. For the national estimates, we post-stratified the weights to the frame counts generally using the

combination of PCP/specialist status and sample frame characteristic (physicians listed on the Round Three frame versus physicians only appearing in the Round Four frame). We used frame totals because they are known and because external totals of the target population of physicians eligible for the CTS do not exist. For the site-level weights, the post stratification adjustment was limited to site membership (as of the time of sample selection) and PCP/specialist status (120 cells).

To ensure weights for the completed interviews produced totals matching the frame totals, we also conducted a similar post-stratification adjustment after the weights were trimmed (see Section D.4) and made adjustments to the site estimates (see Section D.3).

2. Trimming the Weights Adjusted for Nonresponse

After the national population estimates were developed, the weights were trimmed to reduce extreme weights that have the potential to inflate the sampling variance of survey estimates. The statistical measure to quantify the impact of the trimming was based on the design effect attributable to the variation in the sampling weights (Potter 1990). The design effect attributable to weighting is a measure of the potential loss in precision due to the variation in the sampling weights relative to a sample of the same size with equal weights. Sampling weights were trimmed to reduce the design effect. A weight for national-level estimates was trimmed for 0.37 percent of the physicians in the Round Four sample compared to 0.14 percent of the physicians in the Round Three sample.

Table V.4 presents the range in the propensity score adjustments for each of the five models before post-stratification and trimming. The combined adjustments (location and response) ranged from slightly greater than 1 to a high of 62 with averages ranging from about 1.3 to 3.5. In Round Three the adjustments ranged from slightly less than 1 to approximately 27. A few of the adjustment factors in Round Four were larger than encountered previously for the physician

TABLE V.4

SUMMARY OF PROPENSITY SCORE ADJUSTMENTS, BY SAMPLE TYPE AND PANEL FOR ROUND FOUR PHYSICIAN SURVEY (2004-2005)

		Location			Response			Combined[a]		
Panel		Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
Reinterview	PCP		N/A			N/A		1.00	23.87	1.33
	Specialist		N/A			N/A		1.00	20.10	1.30
Noninterview	PCP	1.00	34.30	1.41	1.01	44.97	2.73	1.03	57.39	3.49
	Specialist	1.00	15.02	1.19	1.02	31.32	2.48	1.94	31.63	2.85
New Cases	PCP	1.00	13.91	1.17	1.01	33.96	1.79	1.03	45.44	2.11
	Specialist	1.00	19.84	1.18	1.01	33.96	1.89	1.03	61.59	2.26

[a]The combined adjustment factor is the product of the inverse of the propensity score for the location model and the inverse of the propensity score for the response model. For the reinterview physicians in the supplemental sample, the factor is the inverse of the propensity score from the combined single logistic model, which accounted for both location and response.

surveys, presumably because of the stronger models involving a wider range of covariates and interactions and were focused on the sample rather than the population (Round Three used weighted logistic regressions and Round Four used unweighted logistic regression with the sampling weights and other design-based information as covariates). Also, the magnitude of the adjustments could have been trimmed at each stage, but trimming was used only on the final post-stratified weights. The reason was that less trimming would be required because of offsetting adjustments and weights—and trimming itself can potentially introduce bias.

Table V.5 indicates the impact of the adjustments on the design effects based on the variability in the survey weights. As might be expected from the increased maximum adjustments shown in the previous table, the design effects are also slightly larger—2.99 compared to 2.40 after trimming. The design effects increase incrementally at about equal levels for each of the adjustment stages (about 50 percent increase each at location, response, and post-stratification stages—based on national level full sample weights). The design effect for the national level all-sample weights is reduced from 4.72 after post-stratification to 2.99 after trimming.

3. Site Estimate Adjustments

Site estimates were desired on the basis of the physician's practice, but the site assignment at the time of sample selection may have been based on the physician's "preferred mailing address" available in the AMA and AOA files, and was often the physician's home address. As noted previously (see Section A.3 and Table V.2 of this chapter), physicians who were misclassified were called *movers,* and we had to account for this misclassification in the physician's weights and for site estimates of the eligible physician population.

TABLE V.5

SUMMARY OF DESIGN EFFECTS FROM UNEQUAL WEIGHTS
FROM WEIGHT ADJUSTMENTS FOR ROUND FOUR
PHYSICIAN SURVEY (2004-2005)

		Design Effects By Adjustment Stage				
Panel	Type	Initial Weights	Location Adjusted	Response Adjusted	Post-Stratification	Weight Trimming
Reinterview	Site (all)	1.77	n/a	3.06		
	National (all)	1.28	n/a	2.25		
	Site (in-site)	2.00	n/a	3.39		
	National (in-site)	2.40	n/a	3.56		
Noninterview	Site (all)	1.70	5.27	5.30		
	National (all)	1.33	2.32	3.36		
	Site (in-site)	1.83	5.60	5.67		
	National (in-site)	1.56	2.65	3.74		
New	Site (all)	1.86	4.49	8.91		
	National (all)	1.28	2.25	4.76		
	Site (in-site)	1.97	4.73	9.38		
	National (in-site)	4.14	4.38	7.50		
All	Site (all)				n/a	n/a
	National (all)				4.72	2.99
	Site (in-site)				3.95	2.99
	National (in-site)				4.68	3.36

The basic weights from the sample adjusted for nonresponse and ratio-adjusted to site totals, as of sample selection, provided the basis in part for estimating the number of physicians in each site. Physicians who indicated during the interview that their office was located in a site other than the one recorded at the time of selection were classified as *out-movers*. Out-movers residing in one of the other 60 sites were defined as *in-movers* to that site. Out-movers who were not in one of the 60 sites were not used in the site estimates. In preparing initial site estimate totals, we excluded the out-movers and included the in-movers. Hence, in comparison with the weighted count in each site based on the sample frame (frame estimate), the omission of the out-movers deflated the value for the estimate based on the Round Three survey (survey estimate), and the in-movers increased the value.

Because in-movers had a potentially substantial impact on the survey estimate, we reviewed the estimate and adjusted the estimate to reduce the effect of in-movers. In general, in-movers had larger weights relative to nonmovers (physicians who were correctly assigned to the site), because the weights for the in-movers also included a component to account for the joint selection of the two sites involved.[9] We reviewed the changes in the site estimates as a function of the in-movers, and decided to smooth the changes when the impacts appeared to be excessive and were based on the weights for a few physicians.

We compared the site weights for the full sample and the site weights for in-site physicians, respectively, to assess the impact of in-movers on the survey site estimates. For each site and physician classification (PCP or specialist) combination (120 cells), we computed the percentage

[9]The in-movers usually have a larger weight relative to static site cases and out-movers because an in-mover must have had original (frame) and current (survey) site membership in two of the selected 60 sites. As such, we adjusted the probabilities of selection for these cases to account for the joint selection probabilities of the two sites involved (see Section IV.2).

of the total weight accounted for by the in-movers, and the average percentage of the total weight accounted for by each individual in-mover.

We also computed a trimming criterion value associated with the weights. The weight-trimming algorithm compared each weight with the square root of the average value of the squared weight (Potter 1990):

(5) *Trimming Criterion = SQRT [c * (Sum of squared weights)/ n]*,

where c = 10 and n is the size of the subgroup. This trimming criterion suggested a maximum weight value for the trimming class or sample subgroup.

This process introduced a small downward bias in estimated population totals because the truncated values were not redistributed, and the potential for bias is greatest in those sites with inmovers. So, this estimator is downward biased and the estimator based on untrimmed weights is unbiased but has large standard error, especially at the site level. Hence, we developed a number of other estimators, seeking estimates that would be more stable and conditionally unbiased (subject to assumptions discussed in Appendix B) and that would also be reasonably consistent with frame counts and eligible physician population estimates from prior survey rounds. An average of five such estimators and the unbiased, direct expansion estimator were obtained for each site. These values were then adjusted to be consistent to the national level weights using the full sample (PHYWGT2).

The five estimators were ratio estimators. These estimators use, in different forms, three different estimated percentages of eligible physicians by site and PCP/ specialists (as a percentage of frame counts):

- P_t, the overall percentage of eligible physicians,

- P_o, the percentage of eligible physicians among the physicians on the Round Three frame, and
- P_n, the percentage of eligible physicians among those who were new to the frame in Round Four.

The details of the estimators are presented in Appendix B.3. In addition to the empirical analysis, we investigated other sources of site-level counts of eligible physicians for potential post-stratification counts at the site level. We have not located a source that is sufficiently consistent with the CTS definitions to use for post-stratification, either at the national or site level.

For example, an essay by P. Kletke in the 2000 issue of *Physician Characteristics and Distribution in the U.S.* (AMA) projects the supply of physicians from 1998 to 2020 (Kletke 2000). The projections are model-based and produce three levels that vary according to different sets of assumptions (similar to different levels in Census projections of the U.S. population). This is essentially a labor force projection and as such is substantially higher (about 50 percent) than the CTS physician totals. Some key differences relative to the CTS physician survey is that Kletke limits the counts to allopathic physicians and uses the AMA Physicians Masterfile for the size and composition of the physician population. This is one of the studies that suggested an increase of 1.5 percent per year. This study projects a leveling and eventual decrease in the physician to population ratio, although, historically, the trend has been upward (annual population growth is about 1 percent).

The *Area Resource File* (ARF), which summarizes health-related statistics from myriad sources (U.S. Census being a major source) was used to summarize physician counts by the CTS sites. The summary, similar to CTS, excludes anesthesiologists, radiologists, pathologists, and residents. The counts are mostly about 30 percent higher than the CTS counts, but range from slightly lower to 90 percent higher (large net-inmover situation and difference in definitions of

eligible physician). We feel the levels, like those in the Kletke paper, are overly influenced by AMA Masterfile counts, which we have found to contain only about 70 to 80 percent eligibles for the CTS (varies considerably among sites and strata). Also, the ARF numbers presumably do not account for the movers because of the AMA Masterfile linkage. On the other hand, the estimated percentage eligible on the frame agrees fairly closely with the difference between the ARF and CTS numbers.

Finally, the weights were ratio-adjusted to estimates from the full sample because it was not influenced by in-movers or related site-level adjustments.

4. Site Weight Trimming

After the site population estimates were developed, a second round of trimming was conducted to address the potential of extreme weights that inflate the sampling variance of survey estimates. The following discussion summarizes the procedure for weight trimming to obtain the site population estimates.

The second round of weight trimming identified weights to be trimmed and distributed the trimmed excess among the weights that were not trimmed. The statistical measure of the impact of the trimming was based on the design effect attributable to the variation in the sampling weights. The design effect attributable to weighting is a measure of the potential loss in precision caused by the variation in the sampling weights relative to a sample of the same size with equal weights. Sampling weights were trimmed to reduce the design effect and to minimize the risk of introducing bias into the sample estimates (that is, trimming was limited to ensure a minimal effect on survey estimates). A weight for site-level estimates was trimmed for 2.56 percent of the physicians.

5. Panel Weights

Over half (60 percent) of the Round Four responding physicians responded to both Round Three and Round Four. The panel size is considerably smaller than in previous rounds because of the reduced overall sample size relative to the other rounds (total sample size of 12,500 completed interviews in the first three rounds whereas the total sample size was 7,000 completed interviews in Round Four) and reallocation. However, it still is a valid probability sample of physicians in 2000 when Round Three was conducted, with sufficient precision to analyze matched observations because a substantial portion of the responding Round Three physicians were selected for the Round Four sample, and a high percentage (76.4) of those selected responded in the fourth round). The inferential population was based on the Round Three population, so the Round Four physician weights were adjusted to account for Round Three sampling rates and were then adjusted by the Round Four response rates among these physicians. These adjusted weights were then ratio-adjusted, using a raking procedure to the Round Three totals for various factors.

Table V.6 presents the total eligible sum of the weights of the Round Four sampled physicians who completed a Round Three interview and the total sum of the weights of Round Four after each of the three adjustments used to finalize the panel weights. The first adjustment accounts for the number of physicians who completed the interview in Round Three, but were not interviewed in Round Four. This adjustment inflated 40 percent of the total sum of the weights in the sample and was primarily caused by the reduced sample size in Round Four. The second adjustment accounts for the physicians who were sampled, but were not located, or did not complete the interview. The second adjustment multiplies the weights already adjusted with the first adjustment with the inverse of the total propensity score for nonresponse. The total sum of the weights was inflated 31 percent for the sample. The third adjustment accounts for the

TABLE V.6

ROUND FOUR WEIGHTED COUNTS FOR PHYSICIANS PARTICIPATING IN ROUNDS THREE AND FOUR

Weight	Weighted Sample
Initial weights	179,434
Weights after sampling adjustment	250,452
Weights after nonresponse adjustment	327,187
Weights after raking adjustment	378,711

ineligible doctors. A weighted raking procedure was used for this last adjustment, which inflated 16 percent the total sum of the weights for the sample. In addition, the raking procedure adjusted the survey weights for the Round Four eligible completes so that the weighted distribution for a specified set of Round Three survey items would match the reported results from the Round Three analysis. This weighted raking procedure is iterated until the sum of the weights for each set are matched. Table V.7 presents the survey items used in the raking procedure for the sample.

TABLE V.7

ROUND THREE AND FOUR QUESTIONNAIRE ITEMS USED IN RAKING PROCEDURES
FOR ROUND FOUR PHYSICIAN SURVEY PANEL WEIGHTS

Item	Categories
IMGUSPR3: Foreign Medical School Graduate	2 (Yes/No)
GENDER	2 (Male, Female)
SPEC: Specialty and Subspecialty	7 Categories
CARSAT: Overall Career Satisfaction	5-Point Scale Rating/6 Categories (including unknown)
HRFREEC: Hours of Charity Care	4 Ranges
OWNPR: Ownership Status	3 Categories (Full/Part/Not an Owner)
PRACTYP: Practice Type	11 Categories 0 = Unknown 1 = Solo 2 = Partnership 3 = Small Group 4 = Medium Group 5 = Large Group 6 = HMO Group 7 = Medical School 8 = Hospital 9 = Local Government 10 = Freestanding Clinic
SALWAGE: Salary Compensation	2 Categories 1=Fixed Salary, Not Eligible for Bonus 2=Fixed Salary, Eligible for Bonus
PMCAREC: % Payment Medicare	5 Ranges
PMCAIDC: % Payment Medicaid	5 Ranges
PCAPREVC: % Revenue Pre-Pay Capitation	4 Categories
NMCCONC: # Managed Care Clinics	5 Categories
PMCC: % Revenue Managed Care	5 Categories
SSAT: Patient Satisfaction Affects Compensation	3 Categories (including unknown)
PCTINCNC: Income Category Includes Bonus	3 Categories
YRPRAC3: Years in Practice	4 Categories (including unknown)
INCOME3: Physician's Own Net Income from Medical Practice(s)	6 Categories (including unknown)
AMAPRIM: PCP Status on AMA/AOA files	2 Categories (PCP, specialist)
Total Constraints	**79 Unique Category Targets**

REFERENCES

Cochran, William G. *Sampling Techniques*. Second edition. New York: John Wiley & Sons, Inc., 1965.

CyBulski, K., M. Sinclair, F.B. Potter, and A.B. Ciemnecki. "Adjusting for Nonresponse Among Medicaid Households that Could Not Be Located or Were Located but Did Not Participate in the Minnesota Managed Care Survey." Paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, 1999.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. *Analysis of Longitudinal Data.* New York: Oxford University Press, 1999.

Hosmer, D.W. Jr., and S. Lemeshow. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc., 1989.

Iannacchione, V.G., J.G. Milne, and R.E. Folsom. "Response Probability Weight Adjustment Using Logistic Regression." *Proceedings of the American Statistical Association, Section on Survey Research Methods,* 1991, pp. 637-642.

Kiel, L., M. Chattopadhyay, F. Potter, and M. Reed. "Community Tracking Study Physician Survey: Round One Survey Methodology Report." Technical publication no. 9. Washington, DC: Center for Studying Health System Change, 1998

Kletke, P.R. "The Projected Supply of Physicians, 1998-2020. In: Pasko T., Seidman, B., S. Birkhead (eds). Physician Characteristics and Distribution in the U.S., 2000-2001. Chicago, IL; American Medical Association, 2000.

Little, R.J. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review,* vol. 54, no. 2, 1986, pp. 139-157.

Little, R.J. and Vartivarian, S. On Weighting the Rates in Non-response Weights, *Statistics in Medicine,* 2003; 22, 1589-1599.

Metcalf, C., P. Kemper, L. Kohn, and J. Pickreign. *Site Definition and Sample Design for the Community Tracking Study*. Technical Publication No. 1. Washington, DC: Center for Studying Health System Change, October 1996.

Potter, F. "A Study of Procedures to Identify and Trim Extreme Sampling Weights." *Proceedings of the American Statistical Association, Section on Survey Research Methods,* 1990, pp. 225-230.

Potter, F.J., V.G. Iannacchione, W.D. Mosher, R.E. Mason, and J.D. Kavee. "Sample Design, Sampling Weights, Imputation, and Variance Estimation in the 1995 National Survey of Family Growth." *Vital Health Statistics*, series 2, no. 124, 1998.

Díaz-Tena, N., F. Potter, R. Strouse, S. Williams, and M. Ellrich. "Community Tracking Study Physician Survey Methodology Report: 2000-2001 Round 3." Technical publication no. 38. Washington, DC: Center for Studying Health System Change, 2003

Potter, F., R. Strouse, M. Sinclair, S. Williams, M. Ellrich and R. Torangeau. "Community Tracking Study Physician Survey Methodology Report: 1998-1999 Round 2." Technical publication no. 32. Washington, DC: Center for Studying Health System Change, 2001

Williams, S., J. Hall, and R. Lu (2004). "Response Models in RDD Surveys." *2004 Proceedings of the American Statistical Association, Section on Survey Research Methods [CD-ROM]*. Alexandria, VA: American Statistical Association.