THE COMMUNITY TRACKING STUDY SURVEYS OF PHYSICIANS

Frank Potter, Stephen Williams, and Nuria Diaz-Tena
Mathematica Policy Research, Inc. Princeton, New Jersey 08543-2393

Key words: Community Tracking Study; physician surveys; longitudinal surveys

I. INTRODUCTION

The Community Tracking Study (CTS) is designed to provide a sound information base for decision making by health leaders. The study is conducted by the Center for Studying Health System Change (HSC) with funding from the Robert Wood Johnson Foundation (RWJF). The CTS collects information on how the health system is evolving in the United States, and on the effects of those changes on people. The CTS, operational since 1996, is a longitudinal project that relies on periodic site visits and surveys of households, physicians, and employers. This paper describes the physician (MD) survey component and discusses implementation issues related to national longitudinal surveys of MDs.

The CTS surveys consist of two samples with coverage of the 48 contiguous states: the site sample, and the supplemental sample. The site sample is a multi-stage survey that includes 60 randomly selected locations (sites): 48 large Metropolitan Statistical Areas (MSAs), 3 small MSAs, and 9 nonmetropolitan areas. The site sample is designed for national estimates, but precision levels are specified for site-level estimates, which decrease the efficiency of the site sample for national estimates. The supplement sample is designed to be used in combination with the site sample to enhance the precision of national estimates. The supplement sample also covers the 48 contiguous states stratified into ten regions (Potter et al. (2002)).

In the third round of the CTS Physician Survey (as in the previous two rounds), Mathematica Policy Research (MPR) was responsible for sample design and selection and survey management. Physician interviews were conducted by The Gallup Organization with MPR responsible for sample weighting and estimation. Social and Scientific Systems, Inc. performed the data processing and file preparation. More information about the Community Tracking Study is at http://www.hschange.com/.

II. PURPOSE OF THE SURVEY

The physician survey is designed to document changes physicians (allopathic physicians, MDs and osteopathic physicians, DOs) are experiencing in the health care system and to learn how these changes are affecting physicians, their practices and the way they deliver medical care to their patients. The goal is to provide information to public and private leaders that will enable them to make better policy decisions.

Some of the analytic areas include:

- Impact of managed care participation on physician behavior, perceptions of quality of care provided and physician satisfaction.
- Effects of physician practice arrangements, ownership and risk bearing on the practice of medicine.
- Relationships between the distribution of practice revenue and physician practice style and satisfaction.
- Effects of socio-demographic or market factors on physicians' practice revenues or income.
- Impact of federal, state and local policies affecting physician practice (including Medicare and Medicaid policy) on physician behaviors and perceptions of their impact on quality of care.

The survey is a nationally representative telephone survey of non-federal, patient care physicians. Each round of the Physician Survey contains observations from more than 12,000 physicians who spend at least 20 hours a week in direct patient care. Data were mostly collected from physicians practicing in 60 randomly selected communities, allowing analyses to be conducted at both the national and community level. Twelve sites were selected randomly for more in-depth community level analyses (called the high-intensity sites) from the 48 larger MSA sites (MSAs with 200,000 persons or more).

Together, the high-intensity and low-intensity sites account for about 90 percent of all survey respondents. The site sample can be used to make national estimates and also may be used to make site-specific estimates for the high-intensity sites.

III. TARGET POPULATION

The target population was based on information provided on the American Medical Association (AMA) Masterfile and on the American Osteopathic Association (AOA) membership file. Eligible physicians had to have completed their medical training, practice in a state within the 48 contiguous states, and provide direct patient care for at least 20 hours per week. Physicians who were excluded were: residents, interns, and fellows; inactive or retired physicians; physicians who were not office- or

hospital-based (such as teachers, administrators, and researchers); and non-patient care specialties such as pathologists, anesthesiologists and radiologists[1]. The physicians designated as ineligible for this survey also included federal employees and graduates of foreign medical schools who are licensed only temporarily to practice in the US.

Eligible physicians were then classified as either a primary care physician (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. 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.

IV. DESIGN ISSUES

The precision requirements for cross-sectional site and national estimates, shown in Table 1, were the same for all rounds. However, because this study is longitudinal, survey precision is influenced by the amount of respondent overlap among survey rounds. In addition, physician specialty and practice location could be reported differently in the sample frame (AMA and AOA files) and in the interview. This section also discusses the processes for handling misclassification errors.

A. Sample Overlap

A common feature of longitudinal surveys is the selection of sampling units in one round of a survey for participation in the next round. Including a portion of the physicians who responded to Round Two in the Round Three sample may increase precision substantially for change estimates and, to a lesser extent, for cross-sectional estimates. At the same time, to ensure complete population coverage in Round Three and to minimize respondent burden and conditioning, some proportion of the Round Two sample should be replaced to represent physicians who had no chance of selection in prior rounds.

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 Round Two costs and response rates supported a re-interview rate of 60 to 70 percent, for PCP and specialist, respectively. Based on an expected eligibility and response rate for reinterviewed physicians of 67 percent, the Round Three sample overlap was set at 100 percent for Round Two interviews and 80 percent of the Round Two non-interviews.

[1]Tables listing the specialties excluded from the frame are available from the authors.

1. Benefits and Drawbacks of Increasing Overlap

Increasing the degree of sample overlap between rounds increases the precision of change estimates; however, the potential for gains in precision depends on the degree of correlation between rounds. Increasing the overlap too much can lead to coverage bias. If the overlap portion of the sample includes the entire sample from the previous survey, the new round will have less opportunity to represent physicians who had no chance of selection 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. As the overlap is increased, the weights of new sample members become relatively larger given a fixed overall sample size.

2. Optimum Overlap

A key question for Rounds Two and Three was the optimal overlap between rounds. Because no information was available about the level of correlation between rounds for key study variables in Round Two, we reviewed the sensitivity of optimum overlap at different levels of correlation. Figure 1 shows that 40 to 50 percent overlap is desirable for a range of the most likely levels of correlation. For the Round Three overlap, however, we had information about relative costs and response rates for the various categories of physicians on the sampling frame. This new information prompted the increase in overlap sampling for Round Three compared to that used in Round Two.

For change-estimates between rounds, the optimum level of overlap is 100 percent. For regression-type estimates of Round Three statistics, the optimum level depends on the amount of correlation between observations obtained for both rounds. The form of the regression estimates for Round Three is

$$(1)\quad \overline{y}' = \phi\, \overline{y}'_{2u} + (1-\phi)\,\overline{y}'_{2m},$$

where:

ϕ = a function of reciprocal variances

$\overline{y}'_{2u} = \overline{y}_{2u}$

$\overline{y}'_{2m} = \overline{y}_{2m} + b(\overline{y}_1 - \overline{y}_{1m})$,

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

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 (ϕ) involving ratios of reciprocal variances (Cochran 1963). From Figure 1, 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 completed interviews used for Round Two was 46 percent, while the target re-interview rates were increased to 61 percent and 73 percent, for PCP and specialist, respectively. This increase reflects both the information from Round Two costs and response rates and also robustness of the cross-section estimates as shown in Figure 2. That is, to investigate the robustness of such estimators, we examine the relative efficiency for different levels of overlap (Figure 2). We are interested in optimum levels of overlap and loss of potential gain as we move away from that optimum. Four values for the between-round correlation coefficient (rho) are presented. Clearly, little is gained from these estimators for values of rho of less than 0.5. We can also see 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.

B. 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. During the interview, physicians verified their primary specialties. In some cases, they cited a specialty other than the one listed for them in the AMA or AOA file, necessitating 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 relatively small. In Round Two, seven (7) percent of physicians classified in the sample frames as PCPs responded as specialists, and four (4) percent classified in the sample frames as specialists responded as PCPs. Because PCPs and specialists comprised separate strata with sample size targets, we had to predict switching in the sample allocation to maintain the desired precision.

C. Geographic Misclassification

A goal of the sample design was to assign physicians to a site based on the location of their main practice. Operationally, physicians listed in the AMA or AOA sample frame were classified by the county of their "preferred mailing address," as that address was the most current on the files. AMA staff indicated that many of these addresses are home addresses, rather than main practice locations. In other cases, physicians' practices had moved 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 (MSA).

Some physicians gave a different address when asked in the survey about practice location. As a result, some of them "moved" from one survey site to another. 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 a site other than the main practice site.

For sampling purposes, physicians remained in the site that was originally assigned (i.e., physicians in the Round Two sample who had a practice address outside the 60 sites for the survey were kept in the sampling frame for Round Three). Maintaining the original site assignment enhanced the survey's coverage of physicians. 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, rather than to the site from which they originally were sampled. A mover was considered 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 variances that subsequently must be subjected to weight trimming.

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. The sample sizes for individual sites were adjusted for the Round Three allocation to account for anticipated gains or losses caused by these movers. Movers represented 11.2 percent of the site sample in Round Two. They were a particular problem in the Orange County and Newark high-intensity sites, where they represented 21 percent and 20 percent of the site samples, respectively.

V. IMPLEMENTATION

A. Sampling Frame

As in previous rounds, the sampling frame was developed from physician records maintained by the AMA and AOA. These files contained the most recent information available from the two organizations as of May 2000, just prior to the date

used to select the Round Three sample. The data fields for the full file included names, telephone numbers, addresses, dates of birth, specialties, and other information useful for sampling and data collection. We also used selected information from the Round Two frame and survey results in the frame development.

The three key steps used to construct the frame were:

- Matching the 2000 AMA and AOA files against each other and the Round Two sample to identify physicians added to the sample frames since Round Two
- Excluding ineligible physicians
- Classifying records by primary design strata and site and by the specialty and Round Two outcome secondary strata.

The complete list of physicians for the Round Two and Round Three sampling frames were obtained from the AMA and AOA. The records were then assigned to primary and secondary design strata, and the sample was allocated on the basis of these stratum counts.

After reviewing frequency counts for key items to ensure file accuracy and completeness, the AMA and AOA files were matched to identify physicians in each file, after which the combined AMA/AOA file was matched to the Round Two frame and sample. A computer match by AMA identification number was performed to determine which physicians were on both AMA and AOA files and which were new to the frame in 2000 sample. Two types of nonmatches resulted: 1) a few on the 1998 frame were not identified on the 2000 list—presumed to be mostly ineligible in 2000 and 2) a larger number were on the 2000 list but not on the 1998 frame—presumably these were mostly new and/or ineligible.

Because physicians are added to the AMA and AOA files on an ongoing basis, we had to identify physicians in the Round Three frame who were not in the Round Two frame. Next, each physician was linked to an appropriate site or stratum. For sampling purposes, the site designation and geographic stratum were based on the physician's preferred mailing address on the AMA and AOA files. Finally, each physician was classified as a PCP or specialist. This classification was based on the Round Two survey response (if available) or on the AMA or AOA specialty code.

B. Sampling Units and Stratification

The design for Round Three used stratification to improve precision and to ensure adequate representation by site, geographic region, population density, and physicians who were new to the frame. Stratification also was used to control precision for survey estimates of PCPs and specialists.

1. Site Sample

Within each site, the sample was stratified by PCPs and specialists (primary strata) and by the following four frame strata (secondary strata):

- Physicians who completed interviews in Round Two
- Physicians who were selected for Round Two but who did not complete interviews
- Physicians who were in the sample frame for Round Two but not selected in the sample
- Physicians who were new to the frame.

The number of physicians available in each site and stratum varied substantially among the sites. However, the CTS design specifies a larger effective sample in Lansing, MI or Little Rock, AR (which are high-intensity sites) than in New York, Los Angeles, and Chicago combined (each are low-intensity sites). The smaller pool of physicians and larger effective sample size for some of the high-intensity sites required the use of the finite population correction in the computation of the nominal sample size. The sample allocation process also had to account for stratification and geographic and specialty misclassification.

The sample size and allocation were based on the precision requirements, the frame counts, and the stratification.[2] Table 1 specifies the precision requirement (in terms of effective sample size) for each site for PCPs and specialists. The effective sample sizes were adjusted to compensate for design effects (especially the finite correction); switching among patient care classifications; geographic misclassification; and expected nonresponse from unlocatable, ineligible, or nonresponding physicians. For all sites, a constant design effect (deff) was used in addition to the site-specific finite population correction factor. The sample sizes were then adjusted for physicians who may have been geographically misclassified by practice location and for physicians who may have been incorrectly classified as PCPs or specialists.

The sample sizes also were adjusted for expected errors in specialty assignment (switchers) and geographic misclassification (movers), based on Round Two experience. The adjustment factor was calculated as:

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

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

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, site-specific adjustments were made for the high-intensity sites and overall average adjustments were made for low-intensity sites (1.06918 for PCP and 0.90294 for specialists). The sample sizes were then adjusted to accommodate sample losses resulting from ineligibility nonresponse, and inability to locate some physicians. These numbers, which are referred to as the *base sample*, were allocated to the secondary frame strata. The projected response rates for each frame stratum were used to check that the allocation met the target values in each cell.

The allocation rule was to assign to the frame cells 100 percent of the Round Two completes, 80 percent of the Round Two noninterviews (except physicians who were deceased, retired, or out of the country were excluded), and a proportional number of new cases (physicians new to the frame in 2000). We wanted to proportionally allocate as much sample as possible to control the variation in weights. To obtain a minimum of five interviews in each frame stratum, we permitted some departures from this ideal.

The expected results were obtained by adjusting for an anticipated completion rate (that is, the number of Round Two completed interviews divided by the number fielded in each site, where the fielded sample included completes, nonrespondents, ineligible respondents, and unlocated physicians). The Round Two site-specific completion rates averaged 45.2 percent for PCPs and 53.8 percent for specialists and were used to select samples from the pool of physicians in the Round Two frames who were part of the Round Two sample and from the pool of physicians who were new to the frame since Round Two. For all sites, the projected completion rate was 70.2 percent for the Round Two completes and was 21.7 percent for the Round Two noninterviews.

To control for possible changes in response and eligibility, we selected an *augmented sample*. A substantial proportion of the augmented sample was ultimately fielded in order to approach the target nominal sample sizes.

2. Supplemental Sample

The supplemental sample was a stratified simple random sample of physicians. As with the site sample, four frame categories (secondary strata) were used in two primary strata (PCPs and specialists). In addition, 10 geographic strata were also used.

The basic allocation of the four frame categories again assigned a sample of 95 percent of the Round Two completes and nearly 80 percent of the Round Two noninterviews (except for deceased, retired, and foreign practice) to the two strata for the Round Three sample. A proportional number was then assigned to the stratum of physicians who were new to the Round Three frame; the intent was to include physicians new to the Round Two frame at approximately the same rate as those included from the Round Two frame. Finally, in order to reach the target total, part of the sample was assigned to the stratum of physicians who were in the Round Two frame but who were not selected in Round Two. Some exceptions had to be made when the frame counts would not permit this allocation. This occurred when fewer physicians were available in a stratum than had been allocated to the stratum and when the allocation would have resulted in fewer than five interviews without adjustments.

We began with the target effective sample and then, to determine the nominal sample size, adjusted that sample on the basis of the Round Two design effect. The nominal sample size was then adjusted to account for specialty misclassification and other attrition. The misclassification counts were apportioned by region and stochastically rounded.

These region-specific samples were then allocated to the four frame strata according to two rules: (1) the regional sample was to include essentially all of the Round Two completes and 80 percent of the Round Two noninterviews, and (2) the remaining sample size was proportionally assigned to the physicians who were *new* to the frame and (if necessary) to physicians in the Round Two frame who were not selected for the Round Two sample.

Using projected completion rates based on experience of Round Two for the four strata, and the proportional adjustments made to the counts, we checked whether the allocation would satisfy the target nominal sample sizes (The completion rate is the number of completed eligible interviews divided by the total sample). If it would, the numbers were stochastically rounded to obtain the final base sample. As with the site sample, these numbers were increased to obtain an augmented sample that allowed for approximately a 50 percent reserve sample in each stratum.

References

Cochran, W. G. *Sampling Techniques.* New York: John Wiley and Sons, 1963.

Potter, F., M. Sinclair, and S. Williams (2001). "Examining Attrition in the Physicians Component of the Community Tracking Study". Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA; American Statistical Association.

TABLE 1. SURVEY PRECISION REQUIREMENTS

		Effective Sample Sizes			Sampling Error for $P = 0.5$		
Survey	Estimation Category	PCP	Specialist	Combined	PCP	Specialist	Combined
Site	High-intensity site	400	200	433	0.025	0.035	0.024
Site	Low-intensity site	100	50	114	0.050	0.071	0.047
Site [a]	National	3,450	2,645	4,285	0.009	0.010	0.008
Supplement	National	515	685	1,200	0.022	0.019	0.014

PCP = primary care physician.
[a] No specified constraint for national-level estimates from the site sample; numbers are approximated by average design effects.

FIGURE 1. OPTIMUM SAMPLE OVERLAP FOR DIFFERENT LEVELS OF CORRELATION BETWEEN SURVEY ESTIMATES IN CONSEQUENT ROUNDS

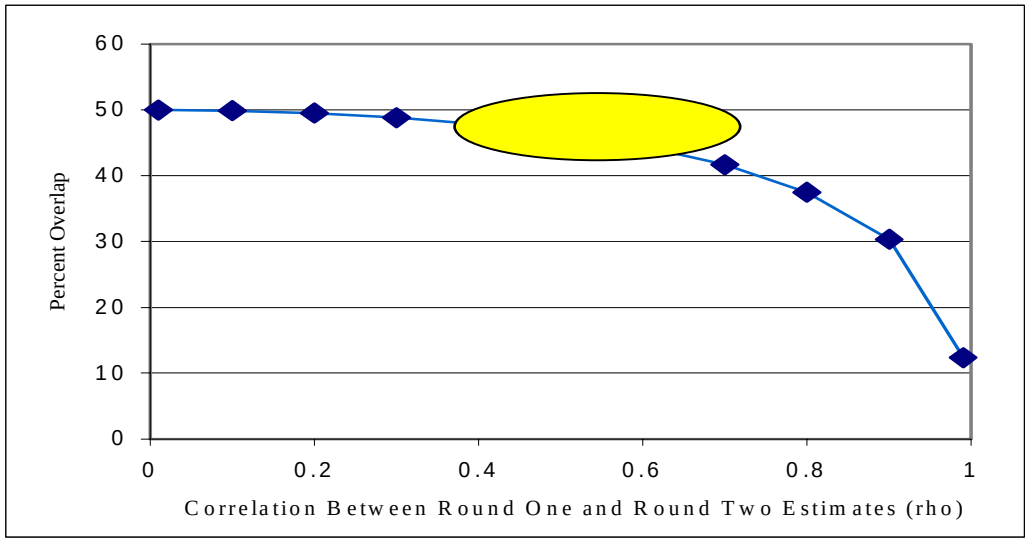

FIGURE 2. RELATIVE EFFICIENCY FOR DIFFERENT LEVELS OF OVERLAP