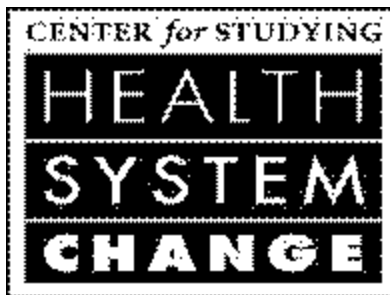# Community Tracking Study

Impact Of Shifting The CTS Physician Survey From A CATI To A Mail Survey

Richard Strouse

COMMUNITY TRACKING STUDY (CTS)

The CTS includes a periodic national survey of physicians. The survey samples are concentrated in 60 communities that were randomly selected to provide a representative profile of change across the U.S. Among these communities, 48 are "large" metropolitan areas (with populations greater than 200,000), from which 12 communities were randomly selected to be studied in depth. Those 12 communities generally have larger survey samples and also comprise the communities used for the site visits.

CTS PHYSICIAN SURVEY

The CTS Physician Survey is a nationally representative telephone survey of non-federal, patient care physicians who spend at least 20 hours a week in direct patient care. Each of the first three surveys includes responses from approximately 12,000 physicians. The most recent survey (2004-05) consists of approximately 6,600 physicians. The survey is conducted by The Gallup Organization. Physician Survey questions cover a range of topics, including financial incentives, care management, acceptance of new patients, provision of charity care, practice characteristics, income and career satisfaction.

The Physician Survey has been conducted in 1996-97 (Round One), 1998-99 (Round Two), 2000-01 (Round Three) and 2004-05 (Round Four).

ADDITIONAL INFORMATION

For more information on the CTS Physician Survey and related HSC Technical Publications, please visit the HSC web site (www.hschange.org).

AUTHOR

Richard Strouse, Independent Consultant

ACKNOWLEDGEMENTS

EXECUTIVE SUMMARY

As a result of increasing cost and declining response rates to the Community Tracking Study's (CTS) CATI Physician Survey, it is likely that a future Center for Studying Health System Change (HSC) physician survey will be a self-administered questionnaire delivered by mail. However, changing data collection modes could affect HSC's ability to track questions collected by telephone in prior rounds, since there is considerable evidence that mode affects responses to many survey questions, particularly attitudinal items which are affected by interviewer mediation. On the other hand, there is interest in re-designing a future CTS physician survey to respond to emerging policy issues, so fewer tracking questions may be retained than for past surveys.

To better understand the tradeoffs in changing data collection modes, HSC pilot tested a mail survey that included tracking questions from the round four CATI survey. The main purpose of the mail survey, which was conducted several months after completion of the round four CATI survey, was to determine how a shift from telephone to mail data collection will affect estimates for tracking variables. In addition, the mail survey provides guidance on expected response rates, optimal incentive level, completeness of individual items, representativeness of the sample, and survey costs.

The fourth round of the CTS CATI Survey was selected from 60 nationally representative communities; primary care physicians were sampled at a higher rate than specialists. The round four sample included physicians who had been sampled in round three (overlap sample) to provide better estimates of change between rounds and to maintain a panel component. Among physicians selected for the overlap sample, those who completed interviews in round three were sampled at a higher rate than were non-completers. A total of 6,628 CATI interviews were completed from June 2004 through July 2005, and the final weighted response rate was 52.4 percent. However, the weighted response rate was significantly higher for the overlap sample of prior round interviews and noninterviews (54.8 percent) than for physicians sampled for the first time (45.7 percent). Assuming that a future national survey of physicians is based on a fresh sample, the lower response rate is a better reflection of the CATI response rate in 2004-2005. Several weights were computed for the CATI survey. For the comparison with the mail survey, we used the weight that is designed for national estimates.

The mail survey sample, which was designed to be representative of the lower 48 states, was a single-stage, unclustered sample of physicians selected with equal probability from the 2003 AMA and AOA Master File, the same frame used for the CATI survey. To assess the impact of a $20 versus $25 incentive on response rates, we randomized the sample between the two incentive levels. A $25 incentive had been used for the CTS CATI survey, but a Mathematica mail survey of physicians conducted at the same time achieved a comparable response rate with a $20 incentive. The CTS mail survey was conducted from August through December, 2005. A total of 709 mail questionnaires were returned by sampled eligible physicians, of whom 338 received the $20 incentive and 371 the $25 incentive. Weighting procedures were largely based on the round four CTS CATI Survey to control for differences in sample composition, although

i

there were some differences in non-response adjustments and trimming. Key findings are summarized below.

Response Rate and Mail Survey Incentive Level. The mail survey response rate (which was computed using the CTS CATI algorithm) for physicians receiving a $20 pre-paid honoraria was 47 percent, but was 53 percent for those receiving $25 pre-paid honoraria; the difference in the response rate was statistically significant. Overall, response rates are higher than the Round Four CATI survey's new sample (46 percent), at less than half the cost per interview. It is possible that a $50 incentive could yield substantially higher response rates. However, given the wide variation in response rates to mailed surveys of physicians by sponsor and topic, it would be useful to test the impact of a $50 incentive in a pilot test for a future survey.

Data Collection Approach for a Future Mail Survey. Following the Dillman Tailored Design Method for mail surveys, the survey included an initial mailing and post card reminder delivered to all sampled physicians, a second mailing to non-respondents, and telephone reminder calls to encourage returns by non-respondents after the two mailings. The telephone reminder calls only added one percentage point to the response rate. Given the relatively high cost of adding a telephone survey operation to the scope of work, we recommend dropping reminder calls from future pre-paid mail surveys. If the response rate is insufficient after two mailings, we recommend a third mailing rather than reminder calls.

Mode Effects. The major goal of the study was to determine whether tracking questions included in prior rounds of the CATI survey can be compared with future mail versions of these questions, without having to adjust for large differences related to data collection mode. Since many tracking comparisons are weighted but not regression adjusted, we first compared unadjusted weighted estimates of CATI and mail versions of tracking items. Next, we ran both unweighted and weighted regressions to show the effect of data collection mode on tracking variables, after controlling for practice and demographic characteristics. Design effects for the weighted regressions were much larger than for the unweighted regressions, likely reflecting the variance in weights. There were differences in the weighting procedures used for the CATI and mail surveys, in particular the decision to trim the CATI weights but not the mail survey weights. Comparing the unweighted and weighted mode survey coefficients indicated where weighting affected mode. Since item non-response affects the size of useable samples and, if large, may bias estimates, we also compared item non-response rates for tracking items. Item nonresponse rates were unweighted because they are largely a function of instrumentation and we wanted to assess the likely impact of mode change on future item nonresponse rates regardless of weighting methods used in subsequent surveys.

Mode clearly affects estimates for many tracking measures, which calls for statistical adjustments in a transition from a CATI to self-administered survey. In the following sections, we summarize mode effects by topic.

- *Career Satisfaction.* Although physicians responding to the CATI survey were more likely to say they were very satisfied with their careers than were those completing mail questionnaires (the socially desirable response), there was no

difference by mode for dissatisfied in career, which is the measure that has been tracked. This pattern indicates that while mode effects may occur for extreme values (very satisfied), they may not affect measures that combine moderate (somewhat dissatisfied) and extreme (very dissatisfied) measures.

- *Ownership.* Practice ownership is unaffected by data collection mode for weighted unadjusted and regression adjusted estimates. However, mail survey respondents were more likely than CATI respondents to own their practices in unweighted regressions, indicating that weights help to correct for differences in ownership response patterns.

- *Financial Incentives.* Physicians responding to the CATI survey were more likely than mail survey respondents to say that financial incentives do not affect the provision of services, which may be the more socially desirable response since it indicates that financial incentives do not affect behavior. Physicians responding to the mail survey were more likely to say that financial incentives expand services. The effect of mode on the impact of financial incentives in reducing services is more ambiguous, with mail survey respondents more likely to say that financial incentives reduce services for the unweighted regression, but not for the weighted regression.

- *Utilization of Time and Patient Mix*. Most items have mode effects, with possible social desirability bias. CATI respondents report more hours worked, more hours in charity care, and say they treat more patients from minority groups and with chronic conditions. Given the importance of these measures, some or all items may be retained; however, it is likely that self-administered measures will be lower than CATI.

- *Information Technology and Care Management*. Most items either had no mode effects or the mail survey measures were larger, which is likely given the temporal gap between the two surveys and the increasing use of information technology by medical practices. In addition, item nonresponse was much lower for the mail survey medical error item.

- *Perceptions of Quality.* Weighted differences for the tracking measures (adequate time, freedom to make clinical decisions, and quality care) were not significant. However, the unweighted coefficient for the percentage disagreeing that they have the freedom to make clinical decisions was positive and significant.

- *Cost Sharing.* Physicians responding to the mail survey were more likely to usually or always consider out-of-pocket costs. The results are contrary to expectations. One would have expected that the socially desirable response would be to consider out-of-pocket costs, which was reported more frequently for the mail than CATI survey.

- *Ability to Obtain Services.* Mail survey respondents were more likely to say they were unable to obtain specialist referrals and less likely to be unable to obtain non-emergency hospital admissions than CATI respondents. Item nonresponse rates also were high for the mail survey rating questions. Given the complexity of the CATI wording and logic, it will be difficult to track these questions in a mail survey.

- *Practice Acceptance of New Patients.* The key tracking measures are acceptance of some or all new patients for each major payor and the uninsured. Mail survey respondents were more likely than CATI respondents to limit access to Medicare, privately insured, and uninsured patients. Respondents to the CATI survey also were more likely to say they accept all new patients for all payors and uninsured patients. These results indicate that respondents to the CATI survey were more likely than respondents to the self administered survey to provide the socially desirable response (acceptance of all new patients) and less likely to acknowledge that their practices limit access to new patients. Differences for most items rating reasons why the practice limits access to Medicare and Medicaid were not significant. Item nonresponse rates for rating items were slightly higher for the mail survey, indicating that respondents to the self-administered survey may have had slightly more difficulty following the questionnaire logic.

- *Factors Affecting Quality of Care.* These questions had large mode effects and will be difficult to track if HSC shifts to a mail survey. Respondents to the mail survey were more likely to cite practice barriers (inadequate time with patients), financial limitations (patients inability to pay), and interference by insurers (rejection of care decisions) as major problems than CATI respondents. Telephone survey respondents were more likely to mention physician or hospital related issues (timely reports and medical errors in hospitals) or language barriers as problems. Relatively few physicians in either survey cited lack of qualified specialists as a major problem and there was no difference by mode.

- *Practice Revenue.* Mode for all three practice revenue questions (percent capitated, percent Medicare, and percent Medicaid revenue) altered weighted estimates by several percentage points. (However, only percent capitated revenue was significant for the unweighted regressions.) Item non-response rates were lower for the mail survey for all items. Given the importance of the Medicare and Medicaid revenue items, inconsistency between the weighed and unweighted regressions, and the absence of a theoretical reason for social desirability biases, these items could be retained with the recognition that mode could confound tracking in round five.

- *Factors Determining Compensation.* CATI survey respondents were more likely to say their practices consider each factor in determining compensation. Since CATI respondents also were more likely to be salaried, and salaried physicians may be more likely to cite these factors, we included salary as a control variable; however, the sign and size of the mail survey dummies were unaffected. Since

iv

these factors are factual items, it is not clear that social desirability was behind the mode difference. On the other hand, the regression model controlled for demographic and practice characteristics (including practice type and whether salaried), and the questions were worded and scaled the same. A possible explanation is that some CATI respondents may have been fatigued at the end of the survey and acquiesced when the interviewer asked the questions. Some studies have shown greater acquiescence on personal or telephone interviews than on self-administered instruments (Dillman, 2000; 2004). In any case, tracking compensation factors between a CATI and self-administered survey will be difficult.

- *Income.* Estimates for the income question were not affected by mode, although item nonresponse was higher for the mail survey. However, shifting to a single interval-based income question may reduce item non-response in a self-administered survey.

Continuing the CTS physician survey by mail would dramatically reduce costs, sustain or increase response rates compared with a telephone survey, but would result in mode effects that are pervasive and often sizable. A key goal for the next round of the physician survey is evaluating statistical methods to determine whether mode effects can be controlled to permit continued tracking of key CTS measures.

I. BACKGROUND AND SURVEY DESIGN

As a result of increasing cost and declining response rates to the Community Tracking Study's (CTS) CATI Physician Survey, it is likely that a future Center for Studying Health System Change (HSC) physician survey will be a self-administered questionnaire delivered by mail. The weighted response rate for the CATI survey has declined from 65 percent in round one (1995-1996) to 52 percent in round four (2004-2005), despite increasing levels of tracing efforts each round. The cost of the round four survey was approximately three million dollars to complete 6,628 interviews averaging 21.3 minutes. Results of recent self-administered physician surveys (all mail or mail with web option) suggest that comparable or higher response rates could be obtained at considerably lower cost. However, changing data collection modes could affect HSC's ability to track questions collected by telephone in the first four rounds, since there is considerable evidence that data collection mode affects responses to many survey questions, particularly attitudinal items which are affected by interviewer mediation. On the other hand, there is interest in re-designing a future CTS physician survey to respond to emerging policy issues, so the number of tracking questions retained in the future may be reduced.

To better understand the tradeoffs in changing data collection modes, HSC pilot tested a mail survey that included tracking questions from the round four CATI survey. The main purpose of the mail survey, which was conducted several months after completion of the round four CATI survey, was to determine how a shift from telephone to mail data collection would affect estimates for tracking variables. In addition, the mail survey provides guidance on expected response rates, completeness of individual items, incentive level, representativeness of the sample, and survey costs.

In the first chapter, we provide background on response rates and methods used on recent self-administered surveys of physicians, summarize recent literature on impacts of mode differences on survey estimates, and describe the designs of the CATI and mail surveys. In the next chapter, we discuss the results of the mail survey and findings from an experiment to test the impact of different pre-paid incentives on response rates. The CATI data collection methodology has been described in methods reports that are available on the HSC web site (www.hschange.org). In the third chapter, we discuss differences between CATI and mail survey estimates and item non-response rates for tracking variables. Differences in estimates are based on bi-variate comparisons and regression analyses that show the effect of data collection mode after controlling for demographic, geographic, and practice characteristics. Bi-variate comparisons of estimates are based on weighted data to control for differences in sample composition between the CATI and mail surveys.

The summary results for both weighted and unweighted regressions are presented. While regressions that incorporate CTS weights are useful for assessing mode effects on tracking variables, design effects were large and there were differences in weighting methods by mode. Nonresponse adjustments were based on somewhat different covariates of the propensity to be located and to respond for the CATI and mail modes. In addition, CATI weights were trimmed and mail survey weights were not.

1

Consequently, regressions also were run on a sample that was unweighted, except for adjustments for unequal probabilities of selection for the CATI sample. Item non-response rates are unweighted to reflect differences in data collection procedures.

The survey instruments and advance letters used on the CATI and mail surveys are included in Appendix A. Supplemental tables with detailed results of the unweighted and weighted regressions along with the item non-response rates can be found in Appendix B. Mail survey costs are discussed in a separate memorandum.

A. BACKGROUND

1. Recent Self-Administered Surveys of Physicians

Response rates to surveys are affected by the respondent's interest in the topic, his view of the study's sponsor, the burden imposed by the survey, and incentives to participate. Monetary incentives have been shown to increase response rates to many self-administered surveys, particularly if they accompany the questionnaire and advance letter (Armstrong 1975; Church 1993; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). Many recent studies have successfully combined Dillman's (2001) Tailored Design Method with financial incentives (James and Bolstein 1990; James and Bolstein 1992; Martinez-Ebers 1997; Singer, Van Hoewyk, and Maher 1998; Tambor et al. 1993; Warriner et al. 1996).

Recent self-administered physician surveys sponsored by non-profit organizations suggest that response rates that are comparable to the round four CTS survey could be obtained with pre-paid incentives that are roughly equal to the $25 post-paid incentive offered to CTS CATI respondents. The Commonwealth Fund sponsored a 2003 survey, completed by mail survey with a web option, on physician use of information technology. That survey, which was selected from the AMA Masterfile, offered a $50 honorarium, was completed in 10 weeks, and obtained a 54 percent response rate (Audet 2004). MPR conducted a 2003 mail survey for Harvard University's Institute for Health Policy on professionalism in medicine ($20 pre-paid incentive, also selected from the AMA Masterfile) and achieved a 56 percent response rate, and was in the field for about four months (Ballou, et.al. 2004).

Higher response rates have been achieved on government sponsored mail surveys with pre-paid incentives on clinical topics. For example, an 81 percent response rate (75 percent using the CTS response rate algorithm)[1] was obtained in a 2004 CDC sponsored survey conducted with 6,906 physicians and other clinicians on knowledge, attitudes, and practices about genital HPV infection and related conditions (Montano 2005). The sample was selected from the AMA Masterfile; each survey (24 pages) was delivered by Federal Express and included a CDC cover letter and a cash incentive of $50. The letter indicated that the survey findings would be used to develop or update clinical training

---

[1] Reported response rates were completed interviews divided by the sum of completed interviews and refused or no response; undeliverables, deceased, and ineligibles were excluded. The CTS response rate is the ratio of complete and ineligible interviews divided by the entire sample.

curricula, clinical decision support tools, and clinical materials physicians would use to counsel their patients. Repeat survey packets were sent to non-respondents 5, 8, 11, and 15 weeks after the initial mailing. Response rates varied by specialty (reported response rate/CTS response rate): family medicine (68/64 percent); general internal medicine (59/54 percent); adolescent medicine (79/75 percent); obstetrics and gynecology (81/70 percent); dermatology (80/75 percent); and urology (78/72 percent). Response rates were higher for specialists who treat HPV infections more often than they were for generalists, indicating the importance of study saliency.

During the design of the CDC study, Kasprzyk (2001) tested the impact of courier service versus first class mail and level of cash incentive (0, $15, $25) on response rates. There was a significant main effect for mode of delivery, with physicians who received the survey by courier service being more likely to respond than those receiving it by first class (by 61 percent to 53 percent). There also was a main effect for incentive (70 percent versus 27 percent); the two way interaction was not significant. Among clinicians provided a monetary incentive, courier service delivery resulted in a higher response than first class. The highest response rate (81 percent) was obtained from clinicians who received $25 with federal express delivery. Since the fielded survey was longer than the one that was pilot tested, a $50 incentive was used in the main study.[2]

2. Mode Effects

There is considerable evidence that the choice of data collection method affects respondents' answers to questions, even when they are worded the same (de Leeuw and Van Der Zowen 1988; Fowler, Roman and Di 1998; Dillman Sanster, Tarnia and Rockwood 1996; Dillman and Christian, 2003). Even when efforts are made to minimize differences in question construction across mode, the presence of an interviewer can affect responses. The main sources of these mode differences are social desirability, acquiescence, and recency effects (Dillman 2000; 2004).

- *Social desirability* is the tendency to offer responses based upon the perceived cultural expectations of others. The presence of an interviewer may increase social desirability compared with self-administration.

- *Acquiescence* is the tendency to agree when interacting with another person. Some studies have shown greater impacts on personal or telephone interviews than on self-administered instruments.

- *Recency and Primacy Effects.* Respondents to telephone interviews are more likely to choose the last offered choice, whereas respondents to self-administered questionnaires tend to choose among the first offered choices; these effects represent differences between visual versus aural cues. The potential bias from recency effects can be controlled in computer assisted interviews (web or CATI) by randomizing the order with which items in a sequence are read. Primacy effects also can be minimized by reversing the order with which items in a list are shown to a randomized half of a sample of mail survey respondents.

---

[2] We are planning on including courier service for a round five physician mail survey.

Two recent papers evaluated mode effects on health surveys conducted by mail and telephone. In a study evaluating the effect of mode of non-response follow-up (overnight delivery or telephone) on a survey initially mailed to Medicare beneficiaries (CAHPS fee-for-service survey), Pugh, et al (2002) did not find significant mode effects associated with the beneficiary's overall health care rating or with any of the 12 items that comprise CAHPS composites related to communication with a physician or nurse, getting needed care, and getting care quickly. However, after adjusting for demographic differences between the overnight delivery and telephone samples, the authors found that mode was a significant factor in Medicare ratings and other subjective questions. The authors speculated that mode (the desire to please on the telephone) affected ratings.

Stringfellow et al (2001) compared the effects of telephone versus self-administration for similar items in the CAHPS and Experience of Care and Health Outcomes (ECHO) instruments. For both surveys, respondents were randomly assigned to mail surveys and telephone interviews. The study focused on 26 items designed to gather reports and ratings of health care. After controlling for health or mental health status and utilization of services, mode had a significant effect on only three items. The authors concluded that the effects of mode were primarily due to differences in self-selection rather than to the effect of mode on how people answered the questions.

Results from a mixed mail/telephone survey of health professionals (physicians, dentists, nurses, and pharmacists) that MPR conducted for RWJF indicated that there were mode effects. Providers (physicians and others combined) responding by mail were more likely to be male and over 45 years of age. We observed mode effects for key dependent variables. In separate logistic regressions run for each specialty that controlled for demographics, region, practice setting, and tobacco use, physicians completing mail questionnaires were less likely to be engaged in tobacco assessment and treatment activities and more likely to agree that patients would resist their advice to quit than those completing by telephone (Strouse 2004).

B. SURVEY DESIGN

1. Instrument Design

Ideally, multi-modal surveys are designed to minimize differences across different modes. Dillman (2001) developed "unimode" guidelines to minimize mode effects. These include:

- Make all response options the same across modes and incorporate them into the stem of the survey question

- Avoid changing the question structure across modes in ways that change the stimulus to the respondent

- Reduce the number of response categories to achieve mode similarity

4

- Use the same descriptive labels for response categories instead of depending upon people's vision to convey the nature of a scale concept

- Develop equivalent instructions for skip patterns that are determined by answers to several widely separate items

- Avoid question structures that unfold (e.g. avoid separate items for direction and intensity)

- Control order effects by reversing order for half the sample

In designing a mail survey version of the CTS CATI instrument, we were unable to follow some of these guidelines because the CATI instrument evolved over four rounds of data collection without consideration of a self-administered mail version. We modified question structures and instructions for several mail survey questions and could not avoid order effects for some other items. The design differences between the CATI and mail surveys are summarized below and differences for each question are shown in the crosswalk between the two versions (Table 1); the CATI and mail survey instruments are included in Appendix A:

- *Access to Frame and Prior Responses.* The CATI survey accessed last reported specialty, board certification, and practice address from the AMA and AOA Masterfiles, allowing the interviewer to anchor questions in the sample frame and update responses for items that have changed. In addition, the CATI program anchored some questions in prior responses. For example, the CATI program first asked for hours in all medically related activities and then asked how many of those hours were in direct patient care. The mail survey provided space for the respondent to record hours in direct patient care, hours in administrative tasks and professional activities, and total hours in medically related activities.

- *Optional Probes.* The CATI survey included many definitions and probes to be read by the interviewer if she believed that the respondent required assistance. Although this approach reduced the reliability of CATI questions, it was considered necessary to reduce interview length. All definitions and probes were displayed in the mail survey.

- *Conversational transitions.* The CATI survey included conversational transitions between sections to improve interview flow; these verbal aides were replaced by section headers in the mail survey. Since this difference was simply designed to transition sections, it was unlikely to affect responses to questions.

- *Order Effects.* The CATI survey rotated question series to minimize order effects; although this could have been replicated in the mail survey by reversing question order for half the sample, it was not done due to budget and time constraints.

- *Question structures.* The CATI question structures could not be replicated for some items (ability to obtain services, why limited Medicare and Medicaid,

5

compensation). Differences in question structures are described in Table 1 and are shown in the survey instruments.

- *New Information Technology (IT) questions.* Since we used the mail survey to pilot test IT questions borrowed from another survey, the IT questions that had been included in the R4 CATI instrument were embedded in the longer series on the mail survey.

- *Logic tests.* A few CATI questions were preceded by logic tests that could not easily be replicated in the mail version. For those items where the mail survey universe included categories of respondents excluded from CATI, we adjusted the mail survey sample during data editing to be consistent with CATI. Two CATI questions (D6a on hospital computer systems to order tests and D6b on medications and medical errors) screened out physicians who never saw patients on hospital rounds; the mail survey did not have a question on hospital rounds, although mail survey respondents were given the option of checking a box and skipping over the questions if they did not admit patients. Therefore, universes may differ slightly on these questions.

Because it was not possible to fully replicate CATI question structures, probes, logic tests, and order effects in the mail survey, differences in estimates between the two versions could have been affected by instrument design as well as the presence of the interviewer. In the discussion of survey findings, we will discuss the potential impact of design differences and interviewer mediation; however, we are not able to analytically differentiate between instrument design and interviewer effects.

TABLE1. Cross Walk Between the CATI and Mail Surveys

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
CAREER SATISFACTION AND DEMOGRAPHICS				
RCARSAT	Satisfaction with overall career in medicine	1	A19	The CATI version includes a conversational introduction: Many of the remaining questions are about your practice and your relationships with patients. Before we begin those questions, let me ask you…
YRBGN	Year began medical practice	2	A6	An optional CATI probe is displayed on the mail survey: A residency or fellowship would be considered graduate medical training.
Specialty (SPEC6CAT)	Primary specialty	3	A7-8	CATI first verifies the specialty on the M asterfile; if different, the R. is asked for current primary specialty. CATI also has access to all specialties on Masterfile in coding responses. The mail survey provides a checklist of eight specialites and "other" option to code specialty if it doesn't fit.
BDCTPSP	Board certification	4	A13	CATI asks: Are you board-certified in recorded specialty; mail survey asks if R. is board certified in primary specialty.
APRCTYPE	Main practice setting	5	C2-7	CATI uses a series of questions, whereas the mail survey uses a check box with an "other" specify option, to get at the main setting.
XOWNPR	Ownership	6	c1	The mail survey includes a check box for an employee or independent contractor; this is a probe in CATI.
NPHYS_CAT	Number of physicians in practice	7	c7	CATI truncates at 997 or more; mail survey truncates at more than 50 (used only as categorical variable in analysis).

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
FINANCIAL INCENTIVES				
XINCENT	Financial incentives	8	cx	The conversational CATI version was: How would you describe your overall personal financial incentives in your practice? On balance, do these incentives favor reducing services to individual patients, favor expanding services to individual patients, or favor neither? The mail version: On balance, do the overall personal financial incentives in your practice favor reducing services to individual patients, favor expanding services to individual patients, or favor neither?
LOCATION OF PRACTICE				
LOCATION	Location of practice	9	A5-5b	County and state verified from file accessed by CATI interviewer and then zip code requested; for mail Survey, respondent asked to record the county, state, and zip code where the practice is located. Both CATI and mail survey state that R should provide location for the practice where the physician spends the greatest amount of time in direct patient care.
UTILIZATION OF TIME AND PATIENT MIX				
XHRSPAT	Hours in direct patient care	10	B3	Questions worded the same. However, CATI first asks for hours in all medically related activities and then asks how many of those hours were in direct patient care. Mail survey provides space to record hours in direct patient care, hours in administrative tasks and professional activities, and total hours in medically-related activities .
HRSMED	Hours in medical activities	10	B2	
XHRFREE	Hours in charity care	11	B6	The CATI question first asks for hours spent providing charity care and then defines charity care. The wording was awkward and the mail survey defines charity care first and then asks for hours spent providing charity care. The CATI survey includes an optional probe explaining that bad debts and time spent under discounted fee for service contracts are not charity care; the mail survey displays the probe.

8

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
XLOCFREE	Charity care location	11A	B6a	Response categories were rotated in CATI but not for the mail survey. The
XBLCKPT	Percent patient black	12a	B14a	CATI survey asks about what percentage of patients are in each group; the
XHISPPT	Percent patient Hispanic	12b	B14b	mail survey asks about what percentage of patients belong to the following patient groups and lists them.
XASIAPT	Percent patient Asian	12c	B14c	
CHRNPT	Percent chronic condition	12F	12	
LANGPT	Percent speak different language	13	B15	No difference

INFORMATION TECHNOLOGY AND CARE MANAGEMENT

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
IT_TRT	Obtain info on treatment	14a_CP	D1A	There were several differences between the CATI and the mail survey IT
IT_FORM	Obtain info on formularies	14b_PD	D1B	questions. CATI asked only if computers and other forms of IT were used in
ITRMNDR	Reminder for preventive svcs	14C_CP	D1C	the physician's practice and rotated each activity. The mail survey included an
ITNOTES	Access patient notes	14a_PI	D1D	expanded set of activities, with he CATI items embedded. For the mail
ITPRESC	Write prescriptions	14d_PD	D1E	survey, physicians also were asked first if the technology was used in the
ITCLIN	Exchange clinical data/phys.	14b_PI	D1F	practice and, if yes, whether the physician personally used the technology.
ITHOSP	Exchange clinical data/hosp/lab.	14c_PI	D1F1	The additonal questions were taken from another survey and were designed to
ITCOMM	Communicate pat. by e-mail	14a_C	D1G	test an expanded set of IT questions for use in future surveys.
ITDRUG	Info on drug interactions	14a_PD	D1H	
EPRESC	Percent e-prescriptions	15	D2aa	The mail survey included the work "About" preceding the CATI question to compensate for interviewers asking for best estimates.
CPOEHSP	Hosp. has comp, systems	17	D6a	The question stems were identical. However, D6a and D6b were asked of all specialists and of all PCPs who said that they saw patients on hospital rounds. In the mail survey, we did not ask PCPs if they saw patients on hospital rounds, although we did allow physicians to select out if they checked the box
ERRREPT	Hosp. system for med errors	18	D6b	"I do not admit patients." The universes should be very close but it is likely that at least some mail survey PCPs who rarely admit and answered the questions would have skipped out.

9

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
XHSPLST	Percent patients hospitalists	19	D7	The question stems were identical. The mail survey includes a not applicable category for physicians who did not admit patients to the hospital in the last year or for hospitalists. CATI skips hospitalists over the question and relies on interviewers to code physicians who didn't admit as not applicable.

PERCEPTIONS OF QUALITY

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
RADQTIME	Adequacy of time with patient	20a	F1a	CATI items were rotated and included two statements that were dropped from the mail survey because they would not be tracked in future surveys (F1E and F1H).
RCLNFREE	Freedom for clinical decision	20b	F1c	
RHIGHCAR	Possible to provide quality care	20c	F1d	

COST SHARING

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
GENERIC	Prescribe generic over brand	21a	F8dA	CATI items rotated
DIAGCST	Consider OOP costs/tests	21b	F8dB	CATI items rotated
IOPTCST	Consider OOP costs/inpatient	21c	F8dC	CATI items rotated

ABILITY TO OBTAIN SERVICES

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
OBREFSR	Unable to get specialist. referral	22	F8bA	The CATI (F8b and F8c series) and mail survey (Q22-24) questions on the ability to obtain needed care for patients were structured differently. CATI first rotated the three services that appear in Q22-24. If the answer was yes, then CATI rotated the reasons why the R. was not able to obtain needed care for each service, asking the physician to rate each one. The mail survey first asked if the R. was unable to obtain a service and, if yes, asked the R to rate each of the reasons why he or she was not able to obtain that service. The different approach used on the mail survey was necessary because of the absence of CATI logical controls.
OBHOSPR	Unable to get hospital admit	23	F8bC	
OBOUTPTR	Unable to get outpatient	24	F8bE	
REFPRVR	Not enough providers in area	22a	F8cAa	
REFHPR	Administrative barriers	22b	F8cAb	
REFINSR	Inadequate insurance coverage	22c	F8cAc	
HSPRVR	Not enough providers in area	23a	F8cBa	
HSPHPR	Administrative barriers	23b	F8cBb	
HSPINSR	Inadequate insurance coverage	23c	F8cBc	
MHPROVR	Not enough providers in area	24a	F8cCa	For the mail survey, Q.24 and Q.24 a-c were asked of all physicians, although Q. 24 has a "does not apply" option. Since the CATI Survey limited these questions to PCPs, OBGs, psychiatrists, geriatric medicine, and adolescent medicine we recoded mail survey physicians in those specialties to "not applicable."
MHHPR	Administrative barriers	24b	F8cCb	
MHINSR	Inadequate insurance coverage	24c	F8cCc	

10

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
PRACTICE ACCEPTANCE OF NEW PATIENTS				
XNWMCARE	Accepting new Medicare	25	F9A	To make the surveys consistent, pediatricians were coded not applicable in the mail survey q.25. The CATI and mail survey questions on practice acceptance of new patients also were structured differently. CATI first rotated the questions on acceptance of new patients for each type of insurance plan (Medicare, Medicaid, private) and the uninsured.
XNWMCAID	Accepting new Medicaid	26	F9B	
XNWPRIV	Accepting new private	27	F9C	
XNWNPAY	Accepting new uninsured	28	F9G	
WHY LIMITS MEDICARE				
MRBILL	Billing	25a	F11a	For practices accepting some or no new Medicare or Medicaid patients, the CATI program displayed questions to rate the importance of the reasons why the practice is limiting or not accepting new Medicare or Medicaid patients, respectively, rotating the reasons. The mail survey first asked the question on acceptance of Medicare patients and, if the answer were some or none, then asked the R. to rate the importance of the reasons why the practice limits or is not accepting new patients (no rotation). The same approach was used for Medicaid, after which the R. was asked the questions on acceptance of privately insured and uninsured patients.
MRAUDIT	Audit	25b	F11b	
MRREIMB	Reimbursement	25c	F11c	
MRNUFPT	Enough patients	25d	F11d	
MRPTBUR	Clinical burden	25e	F11e	
WHY LIMITS MEDICAID				
MDBILL	Billing	26a	F12a	Same approach as above for Medicare
MDDELAY	Delayed reimbursement	26b	F12b	
MDREIMB	Inadq. Reimbursement	26c	F12c	
MDNUFPT	Enough patients	26d	F12d	
MDPTBUR	Clinical burden	26e	F12e	
FACTORS AFFECTING QUALITY OF CARE				
QNOTIME	Q-inadequate time with patients	29a	H20A	The CATI questions on rating of quality of care were rotated; the mail survey questions were not.
QPRBPAY	Q-patient inability to pay	29b	H20B	
QINSREJ	Q-insurance care dec. rejections	29c	H20C	

11

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
QNOSPEC	Q- lack of qualified specialists	29d	H20D	
QNOREPT	Q-not getting timely reports	29e	H20E	
QLANG	Q- difficult comm. due to lang.	29f	H20F	
QERRHSP	Q-medical errors in hospitals	29g	H20H	

PRACTICE REVENUE

XPMCARE	%revenue Medicare	30_1	G1A	CATI included optional probes telling the R. to answer for the main practice
XPMCAID	%revenue Medicaid	30-mcaid	G1B	and asking for best estimates if the R. was unsure; these probes were displayed on the mail survey. Also, CATI checked for responses of over 100%; the mail survey could not. CATI first asks the question and then
XPCAPREV	% capitated practice revenue	31	G3	provides the definition of capitation; the mail survey first provides the definition of capitation and then asks the question. The mail survey also includes the "best estimate" prompt which is at the interviewer's discretion in CATI.

COMPENSATION AND INCOME

SALPAID	Whether salaried physician	32	H1	No difference
SPROD	own productivity affects compensation	33	H5A,H7A	In CATI, the survey first rotates the factors used in determining
SSAT	patient satisfaction affects comp.	34	H5B,H7B	compensation; for each one answered yes, CATI then asks the R. to rate the importance of the factor in determining compensation. The mail survey asks
SQUAL	quality measures affect comp.	35	H5C,H7C	the R. whether a factor is considered in determining compensation; if the
SPROF	profiling results affect compensation	36	H5D,H7D	answer is yes, the R. is asked to rate the importance of the factor in
SPERF	practice financial perf. affects comp.	37	H5E,H7E	determining compensation. Then the survey goes on the next factor.

IMPPROD	importance of productivity	33A	H7aA	In addition, the CATI survey screened full owners of solo practices out of
IMPPSAT	importance of patient satisfaction	34A	H7ab	these questions; for the mail survey, they were asked the questions to keep the
IMPQUAL	importance of quality measures	35A	H7aC	logic simple and excluded during data editing to maintain comparability.
IMPPROF	importance of profiling results	36A	H7aD	
IMPPERF	importance of practice financial perf.	37A	H7aE	

12

TABLE1. Cross Walk Between the CATI and Mail Surveys (cont.)

Variable Name	Description	Mail	CATI	Differences Between Mail and CATI Questions
INC4CAT	Income	38/38A	H15a/H15b	The mail survey displayed a probe instructed Rs t o exclude investment income; this probe was that was optional in the CATI survey. Both the CATI and mail survey had follow-up questions asking Rs who could or would not report open ended income to respond in broad categories. Since Q38 (open ended question) had very high item non-response on the mail survey (35%), we only compared categorized income between the CATI and mail surveys.
ETHNICITY	**ETHNICITY**			
RACEREC	Ethnicity	39-40	H18-19	Both surveys included separate Hispanic and race questions that were combined for analysis.

13

2. Sample Design and Weighting

CATI Survey. The fourth round of the CTS CATI Survey was selected from 60 nationally representative communities; primary care physicians were sampled at a higher rate than specialists. The sample frame for physicians selected for the first time was the merged 2003 AMA and AOA Master File. The round four sample also included physicians who had been sampled in round three (overlap sample) to provide better estimates of change between rounds and to maintain a panel component. Among physicians selected for the overlap sample, those who completed interviews in round three were sampled at a higher rate than were non-completers.

Several weights were computed for the CATI survey. For the comparison with the mail survey, we used the weight that is designed for national estimates using the full sample. The first stage of weighting adjusted for unequal probabilities of site selection and over sampling of primary care physicians. Next, non-response adjustments were made to the sampling weights to reduce the likelihood of nonresponse bias. Logistic regression models were developed to predict the probability of locating a physician (location propensity score), and the probability of that physician's completing the interview (response propensity score). Models were developed from demographic and practice characteristics available on the sample frame and from the prior round of the survey. After the adjustments to the weights for unlocated physicians and for nonresponse among located physicians were applied, 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 interest. A ratio-type adjustment was applied so that the sum of the nonresponse-adjusted weights matched the frame counts. After the national population estimates were developed, the weights were trimmed to address the potential of extreme weights that can inflate the sampling variance of survey estimates (Round 4 CTS Physician Survey Methodology Report, Technical Document #70. Available at www.hschange.org.)

Mail Survey. The mail survey sample, which was designed to be representative of physicians practicing in the lower 48 states, was a single-stage, unclustered sample of 2,000 physicians selected with equal probability from the 2003 AMA and AOA Master File. Since the number of eligible completed interviews was designed to be much smaller than the CATI survey, we wanted to minimize design effects that would have occurred had we clustered the sample in the 60 CTS sites or over sampled primary care physicians. Moreover, we are comparing the mail survey to the CATI sample weighted for national estimates. Before releasing the sample for data collection, we compared practice and demographic characteristics (whether or not PCP, office versus hospital based, age, gender, and census region) between the mail survey sample distribution and the 2003 AMA and AOA frame and verified that the sample was representative. To assess the impact of a $20 versus $25 incentive on response rates, we randomized the sample between the two incentive levels.

The weights for the mail survey sample generally followed the procedures used for the Round Four CATI weighting procedures, except that there was no need to adjust for the probability of selection because mail survey physicians were selected with equal probability, significant covariates used in nonresponse adjustments differed between telephone and mail modes, and the weights were not trimmed. In the first step, we used logistic regression models to adjust for the likelihood of being located and, if located, the likelihood of determining eligibility by returning a questionnaire. Significant covariates for the likelihood of being located

included census region, whether the practice is office based versus hospital based or other undetermined category, whether the physician graduated from a U.S. versus a foreign medical school, gender, whether a reminder call was used, and age. The second step estimated the likelihood of having eligibility determined, conditioned on having been located. Significant covariates were whether the physician graduated from a US versus a foreign medical school, gender, whether a reminder call was used, whether a second mailing was required, and age. For each step, the adjustment was the inverse of the predicted probability of being located or having eligibility determined. Once the sample of completed and ineligible questionnaires was weighted, we performed ratio adjustments to the weighted totals (through post-stratification) from the round four CATI survey. The ratio adjustments were made iteratively by the following groups: primary care versus specialist, gender, and age groups (under 45, 45-64 and 65 plus).

The first ratio adjustment was made to the combined sample of completed and ineligible cases, since Round Four CATI post-stratification adjustments had been made to the same target group. However, when the ineligibles were dropped, the weighted totals for the mail survey completes did not match those from the CATI survey. Since the main purpose of creating the weights was to compare estimates between the completed mail and CATI surveys, a second set of ratio adjustments was made to bring those totals in line with each other. For the second set of ratio adjustments, we also used primary care versus specialist, gender, and age groups (under 45, 45-64 and 65 plus).

Design Effects. Although the mail survey was selected with equal probability, the average design effects across 110 weighted means was 2.74, nearly as high as the clustered round four CATI Survey average of 2.89 for the same set of variables. Since the mail survey was selected with equal probability, we had expected that average design effects would have been similar to the CTS CATI round three national supplement (1.568), which also was selected with equal probability. Since weighting the mail survey had relatively small impacts on demographic and practice characteristics (discussed below in Chapter III with supplementary tables in Appendix B), we conducted a more extensive review of the weighting procedures. Based on this review (by a statistician not involved in the mail survey design), the most likely explanation for the large design effects for the mail survey was the decision not to trim the mail survey weights. At the time the mail survey was completed, a decision was made not to trim the weights because the potential gains did not appear sufficient based on the difficulty of re-fitting the post-stratification. In addition, trimming would have added an additional source of bias. Subsequent analysis indicated that while the weights for most of the mail survey sample (602 cases) were in a narrow range, weights for the remaining 107 cases were quite high.

The relatively high design effects may have reduced the likelihood of detecting statistically significant differences due to mode. The combined impact of the mail and CATI survey design effects on the weighted regressions discussed below resulted in average design effects of 5.96 for the mode coefficient. Consequently, we also conducted unweighted regressions to assess the effect of the mode coefficient, after adjusting for unequal probabilities of selection (for the CATI mode) and controlling for the same set of practice and demographic characteristics used in the weighted regressions, but without the nonresponse, trimming, and post-stratification adjustments from weighting. The average unweighted design effect across all mode coefficients was 1.05. The results of unweighted and weighted regressions on mode coefficients are discussed in Chapter III, with detailed data provided in Appendix B. Generally

14

speaking, the weighted estimates are most relevant to the question of how large the mode effects are for tracking questions based on weighting designs used in prior rounds of the CTS. To adjust for the larger design effects for the weighted regressions, we used a significance level of $p<.10$, rather than $p<.05$).

3. Data Collection

a. *CATI Survey.* A total of 6,628 CATI interviews were completed from June 2004 through July 2005, and the final response rate was 52.4 percent. The response rate was computed as the ratio of completed plus ineligible interviews divided by the total sample released for interviewing attempts. This definition implicitly assumes that the eligibility rate for non-responding physicians is the same as for responding physicians. However, the weighted response rate was significantly higher for the overlap sample of prior round interviews and noninterviews (combined response rate of 54.8 percent) than for physicians sampled for the first time (45.7 percent). The lower response rate is a better predictor of the rate one would expect for a future survey based entirely on new sample.

Extensive efforts were made to trace physicians before contacting them. Advance letters signed by Dr. Lavizzo-Mouray, President and CEO of The Robert Wood Johnson Foundation, were mailed to sampled physicians on RWJF letterhead before they were called by interviewers. Different letters were mailed to physicians who had participated in prior rounds versus those who were being contacted for the first time (see Appendix A). Most physicians were offered a $25 honorarium for survey participation; a sub-sample were mailed $25 checks along with their advance letters to test the impact of pre-payment on response rates (there was no impact on response rates). Extensive follow-up calls and refusal conversion efforts were made to achieve as high a response rate as possible.

b. *Mail Survey.* A total of 709 mail questionnaires were returned by sampled physicians, of whom 338 received the $20 incentive and 371 the $25 incentive. The response rate (which was computed using the CTS CATI algorithm) for physicians receiving a $20 pre-paid honoraria was 47 percent, but was 53 percent for those receiving $25 pre-paid honoraria.

Prior to the start of the survey, tracing efforts based on the CATI survey were used to locate physicians' mailing addresses and telephone numbers (for reminder calls); if telephone numbers were available, physicians' practices were contacted by telephone to verify mailing addresses. Data collection followed Dillman's Tailored Design Method (2001); under this approach, an initial mailing is followed by a post card reminder, second mailing, and reminder calls, if necessary. The questionnaire, which was accompanied by the honorarium and advance letter, was mailed during the last week of August, 2005. The sample was randomized between $20 and $25 honoraria. A $25 honorarium was tested because that amount was used for the CTS Round 4 CATI survey. We also tested $20 because a recent survey of physicians of roughly equivalent burden that was conducted by MPR for Harvard University (Ballou, 2004) obtained a 56 percent response rate with a $20 pre-paid honorarium. We hypothesized that a $5 difference would not affect response rates and that we could save the difference on a future mail survey.

A post card reminder was mailed to all sample members two weeks after the initial questionnaire mailing. A second questionnaire and revised advance letter, referencing the first questionnaire and check, were mailed to nonrespondents during the first week of October.

15

Telephone reminder calls were made to physicians for whom we had telephone numbers, beginning October 27, asking them to complete and return the questionnaires. Interviewers making reminder calls did not complete any interviews. Physicians who said they misplaced or did not receive checks were mailed second checks. Data collection ended December 2, 2005. Copies of advance letters and the post card reminder are included in Appendix A.

II. MAIL SURVEY RESULTS

A. INCENTIVE

1. Impact on Response Rates

The distribution of mail survey responses by incentive is shown in Table 2 and differences in response rates by incentive in Table 3. The $25 incentive resulted in a significantly higher response rate (53 percent) than the $20 incentive (47 percent). The higher response rate was due to greater cooperation by physicians who could be located and consequently received questionnaires. Backing out sample members for whom there was insufficient contact information or were residents of the New Orleans area (the survey was mailed during the hurricane), the cooperation rate (completions plus ineligibles/ completions plus ineligibles plus located non-responses) for the $25 sample was 57 percent versus 51 percent for the $20 sample. Moreover, the percentage of responses that were ineligible was virtually the same for both incentive treatments (29.2 percent for $25 and 29.7 percent for $20). The incremental $5 payment increased physicians' propensity to respond.

A key question for HSC is whether a pre-paid incentive larger than $25 would result in a significantly higher response for a future mail survey of roughly comparable length, difficulty, and sponsorship. As noted above, a 2004 CDC survey reported an 81 percent response rate (75 percent using the CTS response rate algorithm). The CDC survey is not directly comparable to the CTS survey because the topic was clinical rather than health services and the sponsor was a government agency. The honorarium was a pre-paid cash payment of $50. On the other hand, the Commonwealth sponsored survey, also cited above, obtained only a 54 percent response rate with a $50 pre-paid incentive. Given the wide variation in response rates to mailed surveys of physicians by sponsor and topic, it would be useful to test the impact of larger incentives on response rates if HSC wants to consider offering more than $25 in the future.

TABLE 2. Distribution of Mail Responses by Incentive

	$20 INCENTIVE		$25 INCENTIVE		TOTAL	
	Number	Percent	Number	Percent	Number	Percent
Complete	338	33.6	371	37.3	709	35.5
Ineligible	143	14.2	153	15.4	296	14.8
Located Non-response	456	45.4	396	39.8	852	42.6
Insufficient contact information	61	6.1	69	6.9	130	6.5
Hurricane Katrina sample	7	0.7	6	0.6	13	0.6
Total	1,005	100	995	100	2,000	100

TABLE 3. Differences in Unweighted Response Rates by Incentive

	$20 INCENTIVE		$25 INCENTIVE	
	Number	Percent	Number	Percent
Response**	481	48.2	524	53.0
Non-response	517	51.8	465	47.0
Total	998	100	989	100

Note: Response rate is defined as the sum of completed and ineligible interviews divided by the total sample minus the Hurricane Katrina sample, which was not attempted.

** The difference in response rates is significant at p<.05

2. Cost Effectiveness of Incentive

While pre-paid incentives have increased response rates for many studies (see background discussion above), it is important to assess the cost of the procedure since incentives are mailed to all sample members for whom addresses are located. Some eligible respondents who complete questionnaires will not cash checks and some ineligible or non-responding physicians will cash their checks. Table 4 shows the relationship of survey disposition to cashed incentive for each incentive treatment and overall. Nearly all (93.5 percent) physicians who completed questionnaires cashed their checks. However, approximately three-fourths of ineligible physicians also did so. The advance letter did not explicitly state that only eligible respondents should cash their checks since this language might have offended some potential respondents. In the future, we could reduce the fraction of ineligible physicians mailed questionnaires with honoraria by making greater efforts to identify ineligible physicians during screening calls (as long as the added screening cost to determine eligibility does not exceed the savings in incentives that are not cashed).

Nearly 12 percent of physicians whose addresses were confirmed by telephone and did not return questionnaires cashed checks. Sampled physicians in this subgroup mailed the $25 incentive were slightly more likely (12.4 percent) to cash their checks than those mailed $20 (9.0 percent). A small number (n=56) of physicians with unconfirmed addresses (not verified by telephone contact) were mailed questionnaires. Twelve of these 56 physicians cashed checks without responding. Of the 11 physicians in counties affected by Hurricane Katrina, two mailed back questionnaires before the flood. However, since most questionnaires could not be returned, we excluded the Katrina counties from the sample.

Overall, 847 physicians cashed checks in a survey that yielded 709 completed questionnaires, for a ratio of 1.19 checks cashed per completed and eligible questionnaire. For a survey of physicians (conducted by another contractor for a government agency) that used cash incentives of $50, the ratio of incentives expended divided by the number of useable questionnaires was 1.5 times the incentive level. Based on these two observations, the use of checks for pre-paid incentives in the $20 to $50 range is preferable to cash. For future survey budgets using pre-paid incentive checks, we should assume a cost ratio of about 1.2 the incentive level.

18

TABLE 4. Relationship of Survey Disposition and Whether Cashed Incentive

DISPOSTION	$20 INCENTIVE			$25 INCENTIVE			TOTAL SAMPLE		
	Cashed	Not Cashed	Percent Cashed	Cashed	Not Cashed	Percent Cashed	Cashed	Not Cashed	Percent Cashed
Complete	316	22	93.5	347	24	93.5	663	46	93.5
Ineligible - mailed questionnaire	37	11	77	33	10	76.7	70	21	76.9
Ineligible - not mailed questionnaire	0	95	0	0	110	0	0	205	0
Located-non-response	51	405	9.0[b]	49	347	9.0[b]	100	752	11.7
Unconfirmed address - non-response	6	28	17.7	6	28	17.7	12	56	17.6
Other-insufficient contact information	0	27	0	0	35	0	0	62	0
Hurricane Katrina	0	7	0	2[a]	4	33	2[a]	11	15.4
Total	410	595	40.8	437	558	43.9	847	1,153	42.4

[a] Mailed back before flood but questionnaires not used

[b] Difference between $20 and $25 treatments significant at .10 level.

B. IMPACT OF LEVEL OF EFFORT ON RESPONSE RATE

The survey design included an initial mailing and post card reminder delivered to all sampled physicians, a second mailing to non-respondents, and telephone reminder calls to encourage returns by non-respondents after the two mailings. In planning future HSC physician surveys, we can use the results of the mode effects survey to assess the effectiveness of follow-up efforts. Table 5 shows the impact of the level of effort on the sample. The response to the first mailing and post card reminder was productive, as 44 percent of the sample returned questionnaires (30.1 percent eligible and 13.9 percent ineligible). The response to the second mailing was less effective, as only 10.1 percent of the remaining sample responded (8.7 percent eligible and 1.4 percent ineligible). The reminder calls were limited to physicians who did not respond after two mailings and who had valid telephone numbers. Interviewers called physicians' offices to leave a reminder message with a receptionist or on the answering machine; they rarely spoke with physicians. Only 8.6 percent of the reminder calls resulted in eligible or ineligible questionnaires.

TABLE 5. Impact of Level of Effort on the Sample

	First Mailing and Post Card		Second Mailing		Reminder Call[b]	
	Number	Percent	Number	Percent	Number	Percent
Complete	602	30.1	91	8.7	16	6.9
Ineligible	277	13.9	15	1.4	4	1.7
Non-response	1,046[a]	52.3	871	83.3	211	91
Insufficient contact information	62	3.1	68	6.5	1	0.4
Hurricane Katrina sample	13	0.6	0	0	0	0
Total	2,000	100	1,045	100	232	100

[a] Interim status that resulted in second mailing, except for one case that was assigned a final status after the first mailing.

[b] Attempted if there was a good telephone number; effort truncated after a few weeks since results were so poor.

The incremental impact of increased efforts on the survey response rate by incentive is shown in Table 6. For the $20 incentive treatment, the response rate after the first mailing and reminder postcard was 41.5 percent; the second mailing increased it to 47.2 percent and the reminder calls to 48.2 percent. For the sample receiving the $25 incentive, the response rate after the first mailing and postcard reminder was higher (47.0 percent). The incremental impact of the second mailing (increasing the response rate to 52.0 percent) and reminder calls (to 53.0 percent) was similar to the $20 sample. Setting up a telephone reminder operation and conducting the calls was relatively costly; moreover, the gain in response rate was small because we did not have telephone numbers for many non-respondents and the return rate for those for whom we did have was poor. Consequently, we recommend dropping reminder calls for future mail surveys.

TABLE 6. Incremental Impact of Level of Effort on the Response Rate by Incentive

	First Mailing and Post Card		Second Mailing		Reminder Call	
	Number Complete or Ineligible	Response Rate	Number Complete/ Ineligible	Response Rate	Number Complete or Ineligible	Response Rate
$20 Incentive	414	41.5	471	47.2	481	48.2
$25 Incentive	465	47.0	514	52	524	53

III. MODE DIFFERENCES BETWEEN THE CATI AND MAIL SURVEYS

The major goal of the study is to determine whether tracking questions included in prior rounds of the CATI survey can be compared with future mail versions of these questions, without having to adjust for large differences related to data collection mode. In the first section of this chapter, we compare unweighted and weighted mail and CATI survey sample characteristics. There were some differences in response patterns between the two surveys, with all but one eliminated after weighting. Next, we compare unadjusted weighted estimates of CATI and mail versions of tracking items since many tracking comparisons across surveys are weighted but not regression adjusted. Unimputed estimates were used for all comparisons of means, proportions and regressions. In the third section, we describe the model used in the regressions and show both unweighted and weighted mail survey dummy coefficients for ordinary least square and logistic regressions. Since item non-response affects the size of useable samples and, if large, may bias estimates, we also show item non-response rates for tracking items in the third section. In the final section, we summarize the impact of unadjusted, regression adjusted, and item non-response rates on tracking questions. (Consult Appendix B for more detailed tables of results described in this chapter.)

A. COMPARISON OF MAIL AND CATI SAMPLE CHARACTERISTICS

The characteristics of physicians choosing to respond to a mail and telephone survey may differ, which can exacerbate differences in responses to individual questions by mode of data collection. Table 7 compares mail and CATI practice and demographic characteristics for unweighted and weighted estimates. Most variables are based on the sample frame; a few were based on the survey (practice type and race) or frame variables updated on the survey (specialty and board certification). For the CATI survey, unweighted estimates are adjusted to correct for unequal probabilities of selection, but not for differences in response patterns. The mail survey was sampled with equal probability so unweighted estimates are unadjusted. Weighted estimates are based on the national weights described above for the two surveys. Sample characteristics shown in Table 7 were also used as control variables in regressions to test for mode effects.

There were slight differences in some sample characteristics for the comparison between CATI and mail survey unweighted samples. The mail survey included a slightly higher fraction of young (under 35 years of age) physicians, slightly more male physicians, slightly more physicians practicing in general or family medicine and internal medicine, slightly fewer specialists, and slightly more board certified physicians. The only large difference between the two surveys was in type of practice, where physicians to the mail survey were more likely to practice in partnerships or group practices and less likely to practice in community health centers, medical schools, and other settings. It is possible that the mail survey was more effective in getting past "gatekeepers" in group practice settings, but was less likely to get to its target in large institutions. Weighting the two surveys eliminated most of the differences in the unweighted samples, with only practice type remaining significant, but with the size of differences reduced.[3]

---

[3] As noted above, the large design effects observed for the mail survey sample were most likely due to the decision not to trim the weights. Weighting had a small impact on demographic and practice characteristics.

TABLE 7. Comparison of Mail and CATI Survey Sample Characteristics (Percentages)

Variable	Unweighted [a]		Weighted	
	Mail	CATI	Mail	CATI
AGE3 (Sample frame)				
<35	7.78	5.59	5.22	5.42
35-54	62.38	63.11	67.7	65.03
55-64	20.79	20.81	19.44	20.81
>=65	9.05	10.49	7.63	8.73
Chi-Square P-Value	0.08	0.08	NS	NS
GENDER2 (Sample frame)				
Male	76.67	74.5	74.82	74.83
Female	23.33	25.5	25.18	25.17
Chi-Square P-Value	<.01	<.01	NS	NS
RACE (Survey)				
White Non-Hispanic	79.65	78.66	79.51	76.41
Asian Non-Hispanic	12.99	12.21	14.28	13.75
Black Non-Hispanic, Hispanic, Other	7.36	9.13	6.21	9.84
Chi-Square P-Value	NS	NS	NS	NS
CENREG (Sample frame)				
Northeast	24.61	24.17	25.21	21.94
Midwest	20.79	19.21	21.63	19.36
South	33.52	35.25	35.62	35.83
West	21.07	21.37	17.64	22.87
Chi-Square P-Value	NS	NS	NS	NS
MSA2 (Sample frame)				
Metro	88.53	86.55	87.94	85.68
Non-metro	11.47	13.45	12.06	14.32
Chi-Square P-Value	NS	NS	NS	NS
MEDSCHUS2 (Sample frame)				
US	79.77	79.94	75.63	77.6
Foreign	20.23	20.06	24.37	22.4
Chi-Square P-Value	NS	NS	NS	NS
SPEC6CAT (Survey and sample frame)				
General and Family Medicine	19.35	17.01	15.24	16.64
Internal Medicine	15.54	12.19	15.46	11.8
Pediatrics	7.63	8.65	8.74	8.25
OBGYN	7.34	6.81	7.75	6.84
Psychiatry	4.38	7.44	5.08	6.71
Other Specialists	45.76	47.9	47.73	49.77
Chi-Square P-Value	<.01	<.01	NS	NS

22

TABLE 7. Comparison of Mail and CATI Survey Sample Characteristics (Percentages) (cont.)

Variable	Unweighted [a]		Weighted	
	Mail	CATI	Mail	CATI
BDCTPSP2 (Survey and sample frame)				
Board Certified	90.52	86.29	88.5	85.9
Not Board Certified	9.48	13.71	11.5	14.1
Chi-Square P-Value	<.01	<.01	NS	NS
APRCTYPE3 (Survey)				
Solo Practice	27.12	26.85	29.68	26.47
Partnership or Group Practice	43.64	36.43	41.04	37.58
Group or Staff Model HMO	4.94	4.1	4.13	4.52
Hospital	10.45	12.36	9.24	12.02
Community Health Center, Medical	13.84	20.26	15.81	19.42
School, and Other Settings				
Chi-Square P-Value	P<.01	P<.01	P<.01	P<.01

[a] The mail survey sample, which was sampled with equal probability, was unweighted and the CATI survey sample included only base weights, adjusting for unequal probabilities of selection.

B. UNADJUSTED WEIGHTED ESTIMATES

Table 8 shows unadjusted weighted estimates from the mail and CATI surveys for tracking questions. The significance of any mail-CATI differences is indicated by asterisks corresponding to the appropriate p-values. (Refer to Table A in Appendix B for detailed results of the unadjusted weighted estimates from the surveys.) The weight variables were based on the CATI and mail surveys described above. In the following sections, we highlight differences by data collection mode.

TABLE 8. Unadjusted Weighted Estimates and Differences from the Mail and CATI Surveys.

| | | WEIGHTED | | |
| | | Mail | CATI | Mean |
Variable	Label and location in the mail survey	Mean	Mean	Difference
CAREER SATISFACTION AND OWNERSHIP				
RCARSAT3	Q1(R):Dissatisfied in career in medicine	0.1472	0.1439	0.0033
RCARSAT2	Q1(R):Very satisfied in career	0.3413	0.4268	-0.0855***
XOWNPR2	Q6(R):Non-owner vs. owner	0.4207	0.4562	-0.0355
FINANCIAL INCENTIVES				
XINCENT3	Q8(R):Financial incentives - reducing services	0.1407	0.1229	0.0178
XINCENT2	Q8(R):Financial incentives -expanding services	0.2886	0.2369	0.0516*
UTILIZATON OF TIME AND PATIENT MIX				
XHRSPAT	Q10_1:Hours direct patient care	42.6192	45.5941	-2.9749***
XHRSMED	Q10_3:Hours all medical activity	48.5495	53.2135	-4.6640***
XHRFREE	Q11:Hours charity care	5.8762	7.2492	-1.3730*
XHRFREE2	Q11 Zero hours charity care	0.4133	0.3228	0.0905***
XHRFREE2	Q11_a:Location charity care in main practice	0.7447	0.7099	0.0347
XLOCFREE2	Q12_a: % Black	15.4122	18.69	-3.2777***
XBLCKPT	Q12_b: % Hispanic	12.722	14.7882	-2.0661*
XHISPPT	Q12_c: % Asian	5.5773	5.6837	-0.1064
XASIAPT	Q12_d: % Chronic	49.9418	56.9414	-6.9995***
CHRNPT	Q13: % Language problem	5.0416	5.1021	-0.0605
INFORMATION TECHNOLOGY, CARE MANAGEMENT				
IT_TRT	Q14a_cp:IT-Treatment options	0.7044	0.6483	0.0561*
ITRMNDR	Q14c_cp:IT-Reminders	0.2562	0.2929	-0.0367
ITNOTES	Q14a_pi:IT-Patient Notes	0.5597	0.504	0.0558*
ITCLIN	Q14b_pi:IT-Exchange Data Other Phys.	0.5293	0.5009	0.0284
ITHOSP	Q14c_pi:IT-Exchange Data Hosp. & Labs	0.6008	0.6639	-0.0631*
ITCOMM	Q14a_c:IT-Communicate with Patients	0.3045	0.2417	0.0627**
ITDRUG	Q14a_pd:IT-Drug Interactions	0.6319	0.5965	0.0354
IT_FORM	Q14b_pd:IT-Rx Formularies	0.4419	0.4525	-0.0106
ITPRESC	Q14d_pd:IT-Write Rx	0.3206	0.2193	0.1013***
EPRESC	Q15:% Electronic Prescriptions	14.5334	13.0795	1.4539
EPRESC2	Q15(R):No electronic Prescriptions	0.7706	0.8139	-0.0433
CPOEHSP	Q17:Computerized Test Ordering	0.7460	0.7647	-0.0187
ERRREPT	Q18:Medical Error Reporting	0.7702	0.8122	-0.0421
XHSPLST	Q19:% Pts using Hospitalist	27.0482	29.9613	-2.9131
PERCEPTIONS OF QUALITY				
RADQTIME3	Q20_a(R):Disagree-Adequate Time with patient	0.2784	0.2883	-0.0098
RADQTIME2	Q20_a(R):Strongly Agree -Adequate Time with patient	0.3085	0.2994	0.0091
RCLNFREE3	Q20_b(R):Disagree-Freedom for Clinical Decisions	0.1435	0.1073	0.0362*
RCLNFREE2	Q20_b(R):Strongly Agree-Clinical Decisions	0.4664	0.5572	-0.0909***

TABLE 8. Unadjusted Weighted Estimates and Differences from the Mail and CATI Surveys (cont.)

| | | WEIGHTED | | |
Variable	Label and location in the mail survey	Mail Mean	CATI Mean	Mean Difference
RHIGHCAR3	Q20_c(R):Disagree-Possible to Provide Quality Care	0.1984	0.1736	0.0248
RHIGHCAR2	Q20_c(R):Strongly Agree-Quality Care	0.4056	0.4505	-0.0448
COST SHARING				
GENERIC2	Q21_a(R):Prescribe Generic Usually or Always	0.8171	0.7817	0.0354
DIAGCST2	Q21_b(R):Consider OOP Cost-Usually or Always	0.5255	0.4017	0.1239***
IOPTCST2	Q21_c(R):Consider OOP Cost-InP. vs OutP Usually or Always	0.5680	0.5119	0.0561*
ABILITY TO OBTAIN SERVICES				
OBREFSR	Q22:Unable to Obtain Referrals	0.4153	0.339	0.0764**
REFPRVRVM	Q22_a(R): lack of quality providers important	0.5873	0.4733	0.114*
REFHPRVM	Q22_b(R):Adm. Barrier important	0.8412	0.8077	0.0335
REFINSRVM	Q22_c(R):Lack of coverage important	0.8402	0.8324	0.0079
OBHOSPR	Q23:Unable to obtain non-emergency hospital admissions	0.1292	0.2002	-0.0710***
HSPPRVRVM	Q23_a(R): lack of quality providers important	0.4965	0.3713	0.1252
HSPHPRVM	Q23_b(R): Adm. Barrier important	0.7674	0.7536	0.0138
HSPINSRVM	Q23_c(R):Inadequate insurance coverage important	0.7761	0.698	0.0780
OBOUTPTR	Q24:Unable to Obtain Mental Health	0.6119	0.6325	-0.0206
MHPROVRVM	Q24_a(R):Lack of quality providers Important	0.7327	0.798	-0.0654
MHHPRVM	Q24_b(R):Adm. Barrier Important	0.8922	0.7991	0.0931***
MHINSRVM	Q24_c(R):Inadequate insurance coverage important	0.9262	0.8416	0.0847***
PRACTICE ACCEPTANCE OF NEW PATIENTS				
XNWMCARE1	Q25:Accept Some or No Medicare Patients	0.2422	0.1647	0.0775***
XNWMCARE2	Q25:Accept All Medicare Patients	0.6066	0.7054	-0.0989***
MRBILL1	Q25_a(R):Medicare billing requirements important	0.6173	0.6083	0.0090
MRAUDIT1	Q25_b(R): Concern about Medicare audit important	0.3710	0.3040	0.0670
MRREIMB1	Q25_c(R):Medicare inadequate reimbursement Important	0.7612	0.7294	0.0318
MRNUFPT1	Q25_d(R):Practice already has enough patients Important	0.6065	0.5833	0.0232
MRPTBUR1	Q25_e(R):High Clinical Burden Important	0.4674	0.4379	0.0296
XNWMCAID1	Q26(R):Accept No or Some Medicaid Patients	0.4402	0.3851	0.0551
XNWMCAID2	Q26(R):Accept All Medicaid Patients	0.4284	0.5227	-0.0943***
MDBILL1	Q26_a(R):Medicaid billing requirements Important	0.7658	0.7059	0.06
MDDELAY1	Q26_b(R):Medicaid delayed reimbursement Important	0.8036	0.6591	0.1445***
MDREIMB1	Q26_c(R):Medicaid inadequate reimbursement important	0.9321	0.8642	0.0679***
MDREIMB2	Q26_c(R):Medicaid inadequate reimbursement very important	0.8278	0.6874	0.1404***

TABLE 8. Unadjusted Weighted Estimates and Differences from the Mail and CATI Surveys (cont.)

| | | WEIGHTED | | |
| | | Mail | CATI | Mean |
Variable	Label and location in the mail survey	Mean	Mean	Difference
MDNUFPT1	Q26_d(R):Important practice already has enough patients	0.5348	0.4798	0.0549
MDPTBUR1	Q26_e(R):Important New Medicaid Pts-Clinical Burden	0.4987	0.5296	-0.0309
XNWPRIV1	Q27(R):Accepting No or Some Private Insurance Pts	0.1702	0.1355	0.0347
XNWPRIV2	Q27(R):Accepting All Private Insurance Pts	0.5664	0.7162	-0.1498***
XNWNPAY1	Q28(R):Accepting No or Some Uninsured Pts	0.5641	0.4652	0.099***
XNWNPAY2	Q28(R):Accepting All Uninsured Pts	0.3507	0.4462	-0.0955***
FACTORS AFFECTING QUALITY OF CARE				
QNOTIME3	Q29_a(R):Major Problem-Quality Care -Inadequate Time	0.2726	0.1666	0.106***
QPRBPAY3	Q29_b(R):Major Problem-Quality Care-Inability to Pay	0.3189	0.2308	0.088***
QINSREJ3	Q29_c(R):Major Problem-Quality Care -Rejection by Insurance	0.3742	0.2476	0.1267***
QNOSPEC3	Q29_d(R):Major Problem-Quality Care-Lack of Specialists	0.1401	0.1083	0.0318
QNOREPT3	Q29_e(R):Problem-Quality Care -Reports from MDs	0.6159	0.7341	-0.1182***
QLANG3	Q29_f(R):Problem-Quality Care-Language Barrier	0.4185	0.5496	-0.1311***
QERRHSP3	Q29_g(R):Problem-Quality Care-Medical Errors	0.3660	0.5849	-0.2189***
PRACTICE REVENUE				
xpmcare	Q30_1:% Revenue from Medicare	29.7336	32.3568	-2.6232*
xpmcaid	Q30_2:% Revenue from Medicaid	17.3666	16.448	0.9186
xpcaprev	Q31: % Revenue Capitated	8.5727	12.8734	-4.3006***
COMPENSATION AND INCOME				
SALPAID	Q32:Salaried Physician	0.5382	0.6587	-0.1205***
SPROD	Q33:Practice Considers Own Productivity	0.6660	0.7044	-0.0384
IMPPROD2	Q33_a(R): Own Productivity Important	0.9337	0.9513	-0.0176
SSAT	Q34:Practice Considers Satisfaction Surveys	0.1642	0.2459	-0.0817***
IMPPSAT2	Q34_a(R):Satisfaction Surveys Important	0.6924	0.7678	-0.0755
SQUAL	Q35:Practice Considers Measure of Quality of Care	0.1138	0.2019	-0.0881***
IMPQUAL2	Q35_a(R):Quality of Care Measures Important	0.7607	0.8413	-0.0806
SPROF	Q36:Practice Considers Practice Profiling	0.0700	0.1388	-0.0688***
IMPPROF2	Q36_a(R):Practice Profiling Important	0.7360	0.7471	-0.011
SPERF	Q37:Practice Considers Overall Financial Performance	0.5769	0.6895	-0.1126***
IMPRPRF2	Q37_a(R):Practice Financial Performance Important	0.9092	0.8949	0.0143
INC4CAT2	Income 150K or more	0.6126	0.5806	0.032
INC4CAT3	Income 250K or more	0.2845	0.2852	-0.0007

[*] Statistically significant differences (at $0.5 \leq p < .10$)

1. Career Satisfaction

The tracking measure "dissatisfied in career" (which combined somewhat and very dissatisfied) did not differ by mode. However, physicians participating in the mail survey were significantly less likely to say they were *very* satisfied with their career in medicine (34 percent) than those participating in the CATI survey (42 percent). This difference may indicate a social desirability affect since physicians responding to an interviewer administered survey may be more likely to express enthusiasm about career decisions than those self responding.

2. Ownership

The difference in the ownership question by mode was not significant.

3. Financial Incentives

Although the question on financial incentives includes options of reducing services to patients, increasing services to patients, and favor neither, only the first two options are tracking questions. The difference by mode was not statistically significant for reducing services (mail survey 14 percent and CATI survey 12 percent), but physicians to the mail survey were more likely to select expanding services for their patients in response to financial incentives (29 percent) compared with the CATI survey (24 percent). Combining the responses of reducing and expanding services, mail survey respondents were more likely to choose expanding or contracting services (43 percent) than CATI respondents (36 percent). CATI respondents were more likely to choose the passive category (favor neither), which may be the more socially desirable response since it indicates that financial incentives do not affect behavior.

4. Utilization of Time and Patient Mix

Physicians participating in the CATI survey reported more hours in direct patient care, in all medical activities, and in charity care than those completing mail questionnaires. Mail survey physicians were more likely to report zero hours in charity care (41 percent) than CATI physicians (32 percent). Mode had no impact on whether or not the location of charity care was the physician's main practice.[4]

CATI physicians also reported that a higher percentage of their patients were black non-Hispanic and Hispanic and had chronic conditions compared with the mail survey sample. There were no differences by mode for percentage of patients who are Asian non-Hispanic and percentage of patients with language problems. Differences in instrumentation could have been

---

[4] We reviewed distributions to assess the impact of extreme values for the hours worked variables and also constructed categorical versions of the variables to determine whether respondents to the CATI survey still reported higher hours (CATI respondents were more likely to report more than 50 hours per week in direct patient care and in medically related activities).

27

affected by responses to hours delivering charity care. The CATI survey first asks for hours spent providing charity care and then defines charity care.

5. Information Technology and Care Management

Questions on the use of information technology could have been affected by the timing of the surveys, as well as mode, since the mail survey took place several months after the CATI survey and use of technology is likely to be increasing. The mail survey value was higher than CATI or differences between the two surveys were non-significant for all items except one – exchanging date with hospitals and labs, where 66 percent of the CATI physicians compared with 60 percent of the mail survey physicians use the technology in their practices.

The mail survey value was larger than CATI and statistic ally significant for the following variables:

- Obtaining information about treatment alternatives and guidelines (70 versus 65 percent)
- Access to patient notes, medication lists or problem lists (56 versus 50 percent)
- Communications with patients by e-mail (30 versus 24 percent)
- Writing prescriptions (32 versus 22 percent)

However, these differences may have reflected real increases rather than mode effects, since the items are factual and are less likely than attitudinal items to be affected by interviewer mediation.

6. Perceptions of Quality

Only one tracking item ("I have the freedom to make clinical decisions that meet my patients' needs.") differed between the two surveys; 14 percent of the mail but only 10 percent of the CATI physicians disagreed. In addition, only 47 percent of the mail versus 56 percent of the CATI respondents agreed strongly that they have the freedom to make clinical decisions that meet their patients' needs (not a tracking measure). Respondents to the telephone interview were more likely to agree strongly with that statement compared with a self administered questionnaire, suggesting that physicians were more likely to give the socially desirable response when responding to an interviewer.

7. Cost Sharing

Of the three cost sharing tracking items (always and usually combined), two differed by mode:

- How often consider an insured patients out-of-pocket costs in deciding recommended tests (53 percent mail versus 40 percent CATI)
- How often consider patients out-of-pocket costs in deciding between outpatient and inpatient care (57 percent mail versus 51 percent CATI)

Since the questions were worded and scaled identically, these differences should be related to interviewer mediation. However, one would have expected that the socially desirable response would be to consider out-of-pocket costs, which was reported more frequently for the mail rather than the CATI survey. On the other hand, physicians responding to an interview may be less likely to acknowledge cost considerations even if they reduce the patient's expenditure.

8. Ability to Obtain Services

The CATI and mail survey questions were structured differently because not all of the CATI logical controls could be applied to the mail survey. The CATI survey first rotated the three services (unable to obtain referrals to high quality specialists, unable to obtain non-emergency hospital admissions, and unable to obtain high quality outpatient mental heath services). If the answer to any item was yes, CATI rotated the reasons why the respondent was unable to obtained needed care, asking the physician to rate each one in a four point Likert scale. The mail survey first asked if the respondent was unable to obtain referrals to high quality specialists; if yes, the respondent was directed to answer the three rating questions for that service. Then the respondent was asked whether he was unable to obtain non-emergency hospital admissions and, if yes, to answer the rating questions. Finally, the respondent was asked whether he was unable to obtain outpatient mental health services and, if yes, to answer the rating questions for that service. Question wording was the same for the two modes.

Mail survey physicians were more likely than CATI physicians to indicate they were unable to obtain referrals to high quality specialists (42 versus 34 percent). However, CATI physicians were more likely to say they were unable to obtain non-emergency hospital admissions (20 versus 13 percent). Mode was not related to ability to obtain high quality outpatient mental health services.

Small sample sizes, coupled with high item nonresponse (discussed below) for the mail survey limits interpretation of mode differences for the rating questions, since even fairly large differences were not statistically significant at the .10 level for many items. Among physicians who said they were unable to obtain quality outpatient mental health services, CATI physicians were more likely than mail survey physicians to say that administrative barriers were important (89 versus 80 percent) and that inadequate insurance coverage was important (93 versus 84 percent). Nevertheless, the percentage rating these items as important was very high on both surveys.

9. Practice Acceptance of New Patients

Physicians were asked whether they accept all, most, some or no new patients insured through Medicare, Medicaid, and private plans, respectively, and for the uninsured. Those who are limiting access (accepting some or no new patients) were asked to rate the importance of several reasons why their practices limited access to Medicare and Medicaid, but not for private plans or the uninsured. Questions were worded identically, although the logic differed slightly by mode. The four access questions were rotated for CATI respondents and logical tests were used to identify physicians who should be asked the Medicare and/or Medicaid rating questions. For the mail survey, the rating questions followed the Medicare and Medicaid access questions, respectively, based on arrows and written instructions. Pediatricians, who could respond to the mail survey Medicare questions, were excluded from mode comparisons to maintain consistency.

Mode was related to the Medicare, Medicare, and uninsured access tracking questions, with mail survey respondents more likely to say they limit access (accepting some or no new patients combined) than CATI respondents. (Mode was not significant for accepting some or no new private patients.)

- Medicare : 24 percent mail versus 16 percent CATI
- Medicaid : 44 percent mail versus 39 percent CATI
- Uninsured : 56 percent mail versus 46 percent CATI

In addition, physicians self-responding to the mail survey were less likely to say that they accept all new patients for all three plan types (Medicare, Medicaid, and private) and the uninsured.

- Medicare : 61 percent mail versus 71 percent CATI
- Medicaid : 42 percent mail versus 52 percent CATI
- Private plans : 57 percent mail versus 72 percent CATI
- Uninsured : 35 percent mail versus 45 percent CATI

Among physicians limiting access to Medicare, none of the rating items differed by mode. However, two rating items differed by mode for physicians limiting Medicaid access. Delayed reimbursement was rated as important by 80 percent of mail and 66 percent of CATI physicians. Inadequate reimbursement was rated as important by 93 percent of mail and 86 percent of mail survey respondents. Moreover, mail survey respondents were much more likely to rate inadequate reimbursement as very important (83 percent) versus 69 percent for CATI respondents.

Since question wording was identical and formatting similar, it is likely that interviewer mediation affected responses to access questions, with CATI respondents less likely to say they limit patient access and more likely to say they accept all new patients compared with those completing mail questionnaires.

10. Factors Affecting Quality of Care

The only difference between the survey questions on factors affecting quality of care was that the CATI items were rotated while the mail survey items were not. Question wording and scales (major problem, minor problem, not a problem) were identical. The distributions of all items were related to mode and all but one tracking measure (lack of qualified specialists) differed between the mail and CATI surveys. (Note that major and minor problem were combined for the last three items cited below because relatively few physicians selected major problem.)

- Inadequate time with patients a major problem: 27 percent mail versus 17 percent CATI
- Patients inability to pay for needed care a major problem: 32 percent mail versus 23 percent CATI
- Rejections of care decisions by insurers a major problem: 37 percent mail versus 25 percent CATI
- Not getting timely reports a problem (major and minor combined): 62 percent mail versus 73 percent CATI
- Language barriers a problem (major and minor combined): 42 percent mail versus 55 percent CATI
- Medical errors in hospitals a problem (major and minor combined): 37 percent mail versus 58 percent CATI

Respondents to the mail survey were more likely to cite practice barriers (inadequate time with patients), financial limitations (patients inability to pay), and interference by insurers (rejection of care decisions) as major problems than CATI respondents. Telephone survey respondents were more likely to mention physician or hospital related issues (timely reports and medical errors in hospitals) or language barriers as problems.

11. Practice Revenue

There were differences in question design between the two surveys. Although the wording of the Medicare and Medicaid revenue questions were the same, probes instructing respondents to answer for their main practice and to provide best estimates were used at the interviewer's discretion in the CATI survey but displayed below the question for the mail survey. The mail survey mean for percent revenue from Medicare (29.7 percent) was slightly smaller than the CATI mean (32.4 percent). Mode was unrelated to percent revenue from Medicaid.

The effect of mode was greater on the mean percentage of patient care revenue paid on a capitated basis (8.6 percent mail versus 12.9 percent CATI). The CATI survey first asks the question and then provides the definition of capitation whereas the mail survey first defines capitation and then asks the question. If some CATI respondents answered before hearing the definition, they may have misunderstood the question, which could account for a mode difference. Practice revenue questions are unlikely to be affected by social desirability; on the other hand, mail survey respondents have more time to think about financial questions and have access to practice information if they choose to access it.

12. Compensation and Income

The first question in the compensation section ascertains whether the physician is salaried; then physicians are asked to identify factors affecting their compensation and the importance of each factor; finally, they are asked to estimate their income. Owners of solo practices are skipped out of the salary and compensation questions in CATI; they were excluded from the mail survey by editing to maintain comparability. All physicians were asked the income question. Although question wording was consistent between the two modes, the compensation questions were structured slightly differently between the two modes. In CATI, the survey first rotates the factors used in determining compensation; for each "yes" response, the CATI program then brings up the rating question on the importance of the factor. After asking whether a factor affects compensation, the mail survey uses an arrow and instruction to direct the respondent to the rating question, and then goes on to the next factor and rating until the question sequence is completed.

The income question in the CATI survey first asked respondents to estimate their annual (2003) net income from the practice of medicine to the nearest $1,000. If they are unable to provide an estimate, the CATI program brings up a probe in which they are asked to choose a larger interval containing their net income. The mail survey was worded identically, but had to display the income interval probe for respondents who checked a box indicating they cannot estimate their net income to the nearest $1,000. Consequently, mail survey respondents may have been induced to skip the more burdensome open-ended question and answer the income interval question. In addition, we had to ask for 2004 net income because of the difference in timing of the two surveys.

a. *Salaried Physicians*. CATI physicians were much more likely to be salaried (66 percent) than those completing the mail survey (54 percent). Since the question is brief, worded identically in both modes, and unlikely to be affected by social desirability, the difference is likely to be related to differences in response patterns. Compared with mail survey physicians, CATI respondents were more likely to practice in hospitals, community health centers, medical schools, and other institutional settings where physicians are more likely to be salaried (see Table 7). Differences in practice type presumably accounted for at least some of the difference in the percentage salaried between the two surveys.

b. *Compensation*. For each factor, other than own productivity where mode had no effect, CATI physicians were more likely than mail survey respondents to say that their practices consider it in determining compensation:

- Satisfaction surveys: 16 percent mail survey versus 25 percent CATI
- Quality of care measures: 11 percent mail survey versus 20 percent CATI
- Practice profiling: 7 percent mail survey versus 14 percent CATI
- Practice financial performance: 58 versus mail survey versus 69 percent CATI

Mode was not related (at the .10 level) to any of the tracking measures for the rating questions (very and moderately important combined). However, mail survey sample sizes were very small for satisfaction surveys, quality of care measures, and practice profiling, so our assessment of mode impacts for those factors is uncertain. Mode was unrelated to the ratings of

own productivity and practice financial performance, factors that over half of the physicians said were considered in determining their compensation.

c. *Income.* As expected, item non-response for the mail version of the open-ended income question was too high (35 percent missing) to compare that question to the CATI version. Instead, we compared annual income in broad intervals, which also mitigated the impact of the difference in timing (2003 for CATI versus 2004 for the mail survey). There was no difference between the two distributions (see Table A in Appendix B). Nor were there any differences by mode between two tracking measures, income of $150,000 per year or more and income of $250,000 per year or more.

C. REGRESSION ADJUSTED ESTIMATES OF DATA COLLECTION MODE

Both unweighted and weighted regressions were run to show the effect of data collection mode on tracking variables, after controlling for practice and demographic characteristics. As described above, design effects for the weighted regressions were much larger than for the unweighted regressions and there were differences in the weighting procedures used for the two surveys, particularly the decision to trim the CATI weights but not the mail survey weights. Comparing the unweighted and weighted mode survey coefficients indicates where weighting affected mode comparisons, after controlling for demographic and practice characteristics.

The dependent variables included in the regressions are the tracking variables shown above in Table 8 for the unadjusted weighted comparisons. Control variables were identical for both weighted and unweighted regressions and included age, gender, race, census region, whether or not a metropolitan statistical area, U.S versus foreign medical school, specialty, board certification, and type of practice (shown in Table 7), with a mail survey dummy to estimate mode impacts after controlling for those factors. Reference values for the control variables included age 35-54, male, white non-Hispanic, practicing in the northeast census region, practicing in an metropolitan statistical area, attended U.S. medical school, general and family practice, board certified, partnership or group practice. All regressions were run in SUDAAN; imputation was not used for missing items. The unweighted regressions included first stage adjustments for the CATI sample to correct for oversampling.

We ran ordinary least square (OLS) regressions for continuous variables, as well as for binary tracking measures because the coefficients are more intuitive than odds ratios and can be compared to unadjusted differences between CATI and mail survey means. [5] Table 9 summarizes results for OLS mail survey dummy coefficients; detailed results can be found in Appendix B, Table B. [6] Due to the large design effects for the weighted coefficients, we use p-value<.10 as the significance level for describing differences in mail survey coefficients.

---

[5] We also ran logistic regressions for binary tracking measures (see Appendix B, Table C). The results of logistic regressions are similar to the OLS regressions for binary tracking variables. As noted above, we are discussing OLS coefficients because they are more intuitive and can be compared with unadjusted mode differences. However, the logit regressions may be more appropriate if we decide to publish findings.

[6] Table 9 also includes item non-response rates which are discussed in the following section.

In the following sections, we summarize the effect of the added controls provided by the regressions on weighted estimates of tracking measures. The unadjusted mode differences in means and proportions (shown above in Table 8) are compared to both weighted and unweighted OLS mail coefficients. After controlling for demographic and practice characteristics, the weighted mail survey coefficients for most variables were similar in size, direction, and significance level to the unadjusted weighted differences in means and proportions between the mail and CATI surveys, presumably because many of the nonresponse and post-stratification adjustments included in the weights were also used as control variables in the regressions.

The sign and size of significant coefficients for weighted and unweighted mail survey coefficients also were similar for most variables. Of the 92 variables for which unweighted and weighted mail survey coefficients were compared, only 14 were significant for the unweighted but not for the weighted mail survey coefficients and 8 were significant for the weighted but not the unweighted coefficients. Generally, coefficients that were significant (p-value<.10) for the unweighted but not the weighted regressions were similar in size and direction to the weighted coefficients and were significant due to the smaller design effects. However, weighting the regression coefficients increased mode effects for key tracking measures of access to care (accepting some or no new Medicare patients and accepting some or no new privately insured patients) and for percent Medicare and percent Medicaid revenue (discussed below).

TABLE 9. Unweighted and Weighted OLS Mail Survey Regression Coefficients and Item Nonresponse

| | | WEIGHTED | UNWEIGHTED | | |
| | | Mail coefficient | Mail coefficient | Item Nonresponse | |
Variable	Label and location in the mail survey	OLS Beta	OLS Beta	CATI	MAIL
CAREER SATISFACTION AND OWNERSHIP					
RCARSAT3	Q1(R):Dissatisfied in career in medicine	-0.0048	-0.0017	0.21%	0.71%
RCARSAT2	Q1(R): Very satisfied in career	-0.0739**	-.0539***	0.21%	0.71%
XOWNPR2	Q6(R):Non-owner vs. owner	0.0225	0.039**	0.06%	0.14%
FINANCIAL INCENTIVES					
XINCENT3	Q8(R):Financial incentives- reducing services	0.0143	0.0434***	4.39%	1.83%
XINCENT2	Q8(R):Financial incentives -expanding services	0.052*	0.0443**	4.39%	1.83%
UTILIZATON OF TIME AND PATIENT MIX					
XHRSPAT	Q10_1:Hours direct patient care	-3.7485***	-3.5695***	0.36%	1.55%
XHRSMED	Q10_3:Hours all medical activity	-5.8295***	-3.9561***	0.24%	9.73%
XHRFREE	Q11:Hours charity care	-1.487**	-2.0246***	8.74%	2.13%
XHRFREE2	Q11: Zero hours charity care	0.1162***	0.1290***	8.74%	2.13%
XHRFREE2	Q11_a:Location charity care main practice	0.0232	0.0274	2.24%	3.19%
XLOCFREE2	Q12_a: % Black	-3.3496***	-3.3255***	0.98%	2.26%
XBLCKPT	Q12_b: % Hispanic	-0.8372	0.4412	1.06%	3.67%

34

TABLE 9. Unweighted and Weighted OLS Mail Survey Regression Coefficients and Item Nonresponse (cont.)

		WEIGHTED	UNWEIGHTED		
		Mail coefficient	Mail coefficient	Item Nonresponse	
Variable	Label and location in the mail survey	OLS Beta	OLS Beta	CATI	MAIL
XHISPPT	Q12_c: % Asian	-0.1153	-0.71**	1.34%	7.62%
XASIAPT	Q12_d: % Chronic	-7.6786***	-6.6775***	1.00%	3.67%
CHRNPT	Q13: % Language problem	0.3345	0.9687***	0.48%	1.55%
INFORMATION TECHNOLOGY, CARE MANAGEMENT					
IT_TRT	Q14a_cp:IT-Treatment options	0.0629**	0.0512***	0.17%	0.42%
ITRMNDR	Q14c_cp:IT-Reminders	-0.0339	-0.0171	0.57%	0.71%
ITNOTES	Q14a_pi:IT-Patient Notes	0.0763***	0.0367*	0.15%	0.42%
ITCLIN	Q14b_pi:IT-Exchange Data Other Phys.	0.035	-0.0066	0.26%	0.85%
ITHOSP	Q14c_pi:IT-Exchange Data Hosp. & Labs	-0.0577**	-0.1031***	0.21%	0.56%
ITCOMM	Q14a_c:IT-Communicate with Patients	0.0605**	0.0636***	0.15%	0.42%
ITDRUG	Q14a_pd:IT-Drug Interactions	0.0486	0.0311	0.26%	0.71%
IT_FORM	Q14b_pd:IT-Rx Formularies	0.0032	-0.0056	0.27%	1.97%
ITPRESC	Q14d_pd:IT-Write Rx	0.1102***	0.0823***	0.06%	0.56%
EPRESC	Q15:% Electronic Prescriptions	2.1733	0.8195	0%	1%
EPRESC2	Q15(R):No electronic Prescriptions	-0.0545**	-0.0252	0%	1%
CPOEHSP	Q17:Computerized Test Ordering	-0.0208	-0.0474**	1.61%	4.55%
ERRREPT	Q18:Medical Error Reporting	-0.0313	-0.0552***	27.71%	6.27%
XHSPLST	Q19:% Pts using Hospitalist	-1.1559	-2.8671	2.55%	6.04%
PERCEPTIONS OF QUALITY					
RADQTIME3	Q20_a(R):Disagree-Adequate Time with patient	-0.0142	0.0034	0.48%	0.28%
RADQTIME2	Q20_a(R): Strongly Agree- Adequate Time with patient	0.0097	-0.0485***	0.48%	0.28%
RCLNFREE3	Q20_b(R):Disagree-Freedom for Clinical Decisions	0.0234	0.0468***	0.17%	0.42%
RCLNFREE2	Q20_b(R):Strongly agree -Freedom for Clinical Decisions	-0.0834**	-0.0769***	0.17%	0.42%
RHIGHCAR3	Q20_c(R):Disagree-Possible to Provide Quality Care	0.0184	0.0185	0.15%	0.56%
RHIGHCAR2	Q20_c(R):Strongly agree-Possible to Provide Quality Care	-0.0448	-0.0479**	0.15%	0.56%
COST SHARING					
GENERIC2	Q21_a(R):Prescribe Generic Usually or Always	0.0451*	0.034**	1.27%	0.28%
DIAGCST2	Q21_b(R):Consider OOP Cost-Usually or Always	0.1155***	0.097***	2.63%	0.42%
IOPTCST2	Q21_c(R):Consider OOP Cost-InP. vs OutP Usually or Always	0.0601**	0.0774***	6.89%	1.00%

TABLE 9. Unweighted and Weighted OLS Mail Survey Regression Coefficients and Item Nonresponse (cont.)

		WEIGHTED	UNWEIGHTED		
		Mail Coeffic ient OLS Beta	Mail Coefficient OLS Beta	Item Nonresponse CATI	MAIL
Variable	Label and location in the mail survey				
ABILITY TO OBTAIN SERVICES					
OBREFSR	Q22:Unable to Obtain Referrals	0.0663**	0.0399**	0.95%	0.59%
REFPRVRVM	Q22_a(R): lack of quality providers important	0.1027**	0.1046***	0.78%	10.42%
REFHPRVM	Q22_b(R):Adm. Barrier important	0.0239	-0.0338	0.74%	8.88%
REFINSRVM	Q22_c(R):Lack of coverage important	0.0302	0.0272	0.83%	8.88%
OBHOSPR	Q23:Unable to obtain non-emergency hospital admissions	-0.0637***	-0.0579***	5.97%	1.62%
HSPPRVRVM	Q23_a(R): lack of quality providers important	0.1383*	0.0571	1.45%	16.67%
HSPHPRVM	Q23_b(R): Adm. Barrier important	0.0369	0.034	0.92%	15.48%
HSPINSRVM	Q23_c(R):Inadequate insurance coverage important	0.0807	0.0853*	0.76%	15.48%
OBOUTPTR	Q24:Unable to Obtain Mental Health	-0.0463	-0.071***	3.07%	1.12%
MHPROVRVM	Q24_a(R):Lack of quality providers Important	-0.0347	0.0405	0.63%	3.43%
MHHPRVM	Q24_b(R):Adm. Barrier Important	0.0604*	0.047*	0.79%	2.45%
MHINSRVM	Q24_c(R):Inadequate insurance coverage important	0.0768***	0.0475**	0.52%	3.43%
PRACTICE ACCEPTANCE OF NEW PATIENTS					
XNWMCARE1	Q25:Accept Some or No Medicare Patients	0.0624**	0.0159	4.16%	2.06%
XNWMCARE2	Q25:Accept All Medicare Patients	-0.0744***	-0.0679***	4.16%	2.06%
MRBILL1	Q25_a(R):Medicare billing requirements important	-0.0127	0.0366	4.45%	5.36%
MRAUDIT1	Q25_b(R): Concern about Medicare audit important	0.0677	-0.0087	5.17%	7.14%
MRREIMB1	Q25_c(R):Medicare inadequate reimbursement Important	0.0212	0.0191	4.65%	7.14%
MRNUFPT1	Q25_d(R):Practice already has enough patients Important	-0.0676	0.0195	4.76%	3.57%
MRPTBUR1	Q25_e(R):High Clinical Burden Important	0.0256	0.0011	4.96%	6.25%
XNWMCAID1	Q26(R):Accept No or Some Medicaid Patients	0.0101	-0.0124	3.44%	1.42%
XNWMCAID2	Q26(R):Accept All Medicaid Patients	-0.0543**	-0.0365*	3.44%	1.42%
MDBILL1	Q26_a(R):Medicaid billing requirements Important	0.0226	0.0248	4.74%	6.92%
MDDELAY1	Q26_b(R):Medicaid delayed reimbursement Important	0.1153***	0.098***	5.76%	7.96%
MDREIMB1	Q26_c(R):Medicaid inadequate reimbursement important	0.0336	0.0262	3.88%	5.54%

TABLE 9. Unweighted and Weighted OLS Mail Survey Regression Coefficients and Item Nonresponse (cont.)

| | | WEIGHTED | UNWEIGHTED | | |
| | | Mail Coefficient | Mail Coefficient | Item Nonresponse | |
Variable	Label and location in the mail survey	OLS Beta	OLS Beta	CATI	MAIL
MDREIMB2	Q26_c(R):Medicaid inadequate reimbursement very important	0.0926***	0.0786***	3.88%	5.54%
MDNUFPT1	Q26_d(R):Important practice already has enough patients	0.0223	0.0218	4.03%	9.34%
MDPTBUR1	Q26_e(R):Important New Medicaid Pts-Clinical Burden	-0.0708*	-0.0102	4.29%	9.34%
XNWPRIV1	Q27(R):Accepting No or Some Private Insurance Pts	0.0406*	-0.0057	2.94%	1.42%
XNWPRIV2	Q27(R):Accepting All Private Insurance Pts	-0.1412***	-0.1275***	2.94%	1.42%
XNWNPAY1	Q28(R):Accepting No or Some Uninsured Pts	0.067**	0.0328*	4.69%	1.42%
XNWNPAY2	Q28(R):Accepting All Uninsured Pts	-0.0605**	-0.0587***	4.69%	1.42%
FACTORS AFFECTING QUALITY OF CARE					
QNOTIME3	Q29_a(R):Major Problem-Quality Care-Inadequate Time	0.0887***	0.1132***	1.57%	0.71%
QPRBPAY3	Q29_b(R):Major Problem-Quality Care-Inability to Pay	0.0921***	0.1144***	1.03%	0.85%
QINSREJ3	Q29_c(R):Major Problem-Quality Care-Rejection by Insurance	0.1092***	0.169***	1.13%	0.85%
QNOSPEC3	Q29_d(R):Major Problem-Quality Care-Lack of Specialists	0.0337	0.0428***	0.33%	0.56%
QNOREPT3	Q29_e(R):Problem-Quality Care-Reports from MDs	-0.1195***	-0.1007***	0.47%	0.99%
QLANG3	Q29_f(R):Problem-Quality Care-Language Barrier	-0.1254***	-0.0829***	0.23%	0.42%
QERRHSP3	Q29_g(R):Problem-Quality Care-Medical Errors	-0.2169***	-0.1837***	2.14%	1.99%
PRACTICE REVENUE					
xpmcare	Q30_1:% Revenue from Medicare	-3.4622***	-1.2625	17%	6%
xpmcaid	Q30_2:% Revenue from Medicaid	2.6608*	1.188	15%	7%
xpcaprev	Q31: % Revenue Capitated	-3.4453***	-1.733 **	17%	9%
COMPENSATION AND INCOME					
SALPAID	Q32:Salaried Physician	-0.076**	-0.0847***	0.29%	0.57%
SPROD	Q33:Practice Considers Own Productivity	-0.0554*	-0.0349	1.35%	0.57%
IMPPROD2	Q33_a(R): Own Productivity Important	-0.0489**	-0.0185	0.56%	1.33%

TABLE 9. Unweighted and Weighted OLS Mail Survey Regression Coefficients and Item Nonresponse (cont.)

		WEIGHTED	UNWEIGHTED		
		Mail Coefficient	Mail Coefficient	Item Nonresponse	
Variable	Label and location in the mail survey	OLS Beta	OLS Beta	CATI	MAIL
SSAT	Q34:Practice Considers Satisfaction Surveys	-0.0826***	-0.0642***	1.66%	0.38%
IMPPSAT2	Q34_a(R):Satisfaction Surveys Important	-0.0933*	-0.0021	0.99%	3.54%
SQUAL	Q35:Practice Considers Measure of Quality of Care	-0.0826***	-0.0752***	1.58%	0.57%
IMPQUAL2	Q35_a(R):Quality of Care Measures Important	-0.0546	-0.104	0.83%	1.11%
SPROF	Q36:Practice Considers Practice Profiling	-0.0578***	-0.0361**	2.07%	0.57%
IMPPROF2	Q36_a(R):Practice Profiling Important	-0.0214	0.0448	1.73%	2.90%
SPERF	Q37:Practice Considers Overall Financial Performance	-0.1107***	-0.0972***	1.66%	0.75%
IMPRPRF2	Q37_a(R):Practice Financial Performance Important	-0.0151	0.0111	0.87%	4.93%
INC4CAT2	Income 150K or more	0.0564	0.0025	4.86%	12.27%
INC4CAT3	Income 250K or more	0.0059	-0.0285	4.86%	12.27%

* Statistically significant differences (at $0.5 \leq p < .10$)

** Statistically significant differences (at $0.01 \leq p < .05$)

*** Statistically significant differences (at $p < .01$)

1. Career Satisfaction

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. The weighted regression results are similar to unadjusted weighted comparisons. Mode was not significant for tracking measures (very and somewhat dissatisfied with career or practice ownership), but was for "very satisfied in career." Results indicate mail survey physicians were less likely than CATI physicians to be very satisfied with their career choice.

Weighted versus unweighted regression adjusted mode coefficients. The direction, size, and significance levels of the mail survey coefficients for career satisfaction also are similar for unweighted and weighted regressions.

2. Ownership

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. The weighted unadjusted and regression results are similar.

38

Weighted versus unweighted regression adjusted mode coefficients. Although the size and sign of the coefficient were similar, the unweighted regression (smaller design effect) was significant but the weighted regression was not.

3. Financial Incentives

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. The results of the weighted regressions were similar to unadjusted comparisons, with mode not significant for reducing services but significant and positive for expanding services. After controlling for personal and practice characteristics, the OLS mail survey coefficient was virtually the same as the unadjusted difference (5 percent).

Weighted versus unweighted regression adjusted mode coefficients. The mail survey coefficient for expanding services was virtually the same for unweighted and weighted regressions, but the coefficient for reducing services was significant and larger (.043 versus .014) for the unweighted compared with the weighted coefficient.

4. Utilization of Time and Patient Mix

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Regression adjusted and unadjusted mode differences were consistent, except for percent Hispanic, where the mode coefficient was not significant.

Weighted versus unweighted regression adjusted mode coefficients. Weighting had little impact on the sign, size, and significance level for most mail survey coefficients. Howver, the percentage of patients who are Asian and the percentage of patients with language problems were significant and the coefficients were larger for the unweighted compared with the weighted regressions.

5. Information Technology and Care Management

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Regression adjusted and unadjusted mode differences were consistent, except for the percentage of physicians who write no prescriptions electronically (EPRESC2). The mode coefficient was significant, whereas the unadjusted difference was not. However, the OLS mail survey coefficient (-.055) was only slightly larger than the unadjusted difference in means (-.043). The negative coefficient indicated that physicians responding to the self-administered survey were less likely to indicate that they write no electronic prescriptions.

Weighted versus unweighted regression adjusted mode coefficients. Unweighted mail survey coefficients were significant for three coefficients for which weighted coefficients were not: whether the physician's hospital has computerized systems to order tests and medications, whether the physician's hospital has a system for reporting medical errors, and the percentage of hospitalized patients treated by hospitalists. However, the signs (negative) of the coefficients were consistent and they differed in size by only one to two percentage points. In addition, the coefficient for EPRESC2 was significant for the weighted (-.055) but not the unweighted (-.025) coefficient. The combined impact of weighting and the control variables in the regression

resulted in a mode effect for this variable. Other coefficients were similar in size, direction, and significance (p<.10), regardless of whether the regressions were weighted.

6. Perceptions of Quality

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Weighted regression adjusted mode differences were generally consistent with weighted unadjusted differences for mo st tracking measures (percentage disagreeing). However, the unadjusted difference (.036) for the tracking measure (percentage disagreeing that they had the freedom to make clinical decisions that meet their patients' needs) was marginally significant (p=.10), but did not approach significance forin the weighted regression mode coefficient (.023), after controlling for practice and demographic characteristics.

Weighted versus unweighted regression adjusted mode coefficients. The unweighted coefficient (.047) for the percentage disagreeing that they have the freedom to make clinical decisions was significant, indicating that weighting reduced the size of the mode coefficient, as well as increasing the variance due to the larger design effects.

We also compared coefficients for *agree strongly* indicators (which are not tracking measures) since we observed mode effects for some extreme values, even when tracking measures that combined weak and strong agreement or disagreement were not significant. All three of the unweighted mode coefficients are negative and significant, whereas only one of the weighted mode coefficients (strongly agree freedom to make clinical decisions) is significant (and also negative).

7. Cost Sharing

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Mail survey coefficients were statistically significant, large, and positive, and were consistent with unadjusted comparisons of differences.

Weighted versus unweighted regression adjusted mode coefficients. The direction and size of the three weighted and unweighted coefficients also are similar and all are significant, although the p-values of the unweighted coefficients were smaller.

8. Ability to Obtain Services

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Regression adjusted and unadjusted mode differences were consistent in significance level and sign for tracking questions on inability to obtain referrals to high quality specialists (positive and significant), non-emergency hospital admissions (negative and significant), and high quality outpatient mental health services (not significant). Although the signs and size of OLS dummy coefficients were generally consistent with unadjusted differences for rating questions, there were a few differences in significance levels.

- Lack of quality providers important for inability to obtain non-emergency hospital admissions: not significant for unadjusted difference but significant for mail survey coefficient

- Administrative barrier important for inability to obtain high quality outpatient mental health services: significant for the unadjusted difference but not significant for mail survey coefficient

Weighted versus unweighted regression adjusted mode coefficients. Weighting did not affect significance (p<.10) and the size and direction of significant coefficients for most regression coefficients. However, the coefficients for inability to obtain mental health services and the importance of inadequate health insurance coverage for the inability to obtain non-emergency hospital admissions were significant for the unweighted but not for the weighted regressions. Nevertheless, the weighted and unweighted mode coefficients were virtually equivalent for the importance of inadequate health insurance coverage and differed by only three percentage points for the inability to obtain mental health services, indicating that while weighting the regression increased the design effect and reduced the likelihood of detecting a significant coefficient, it did not have very much of an impact on the sign and size of the coefficients themselves.

The importance of lack of quality providers for the inability to obtain non-emergency hospital admissions was significant for the weighted but not for the unweighted coefficient. Here, the weighted mode coefficient (.14) also was considerably larger than the unweighted coefficient (.06). Since the difference in unadjusted weighted means also was not significant for this tracking measure, it appears that the combination of weighting and control variables provided by the regression model was required for mode to be significant.

9. Practice Acceptance of New Patients

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Regression adjusted and unadjusted mode differences were consistent in significance level and sign for all but two items:

- Accept no or some Medicaid patients is marginally significant for the unadjusted difference but does not approach significance for the mail survey OLS coefficient,

- Inadequate reimbursement is an important reason for the practice limiting or not accepting new Medicaid patients is significant for the unadjusted difference but is not significant for the mail survey OLS coefficient.

Weighted versus unweighted regression adjusted mode coefficients. Weighting resulted in the coefficients for two key tracking measures (accepting some or no new Medicare patients and accepting some or no new privately insured patients) being significant, whereas these coefficients were not significant for the unweighted regressions, despite much lower design effects. Moreover, the mail survey coefficients for the weighted measures were considerably larger than the unweighted-- .062 versus .016 for accepting some or no new Medicare patients and .041 versus -.006 for accepting no or some privately insured patients. In addition, while both coefficients were significant (p<.10), the weighted mail survey coefficient for accepting no or some uninsured patients (.067) was larger than the unweighted coefficient (.033). Also, the

41

weighted (but not the unweighted) coefficient for high clinical burden as an important reason for limiting Medicaid access was significant.

Weighting increased the percentage of respondents to the mail compared with the CATI survey who said they would restrict access to patients (by accepting no new patients or only some new patients). Since both the unweighted and weighted regressions included the same demographic and practice control variables, it is possible that this result was related to differences in mail and CATI weighting procedures, with the main difference being that the CATI weights were trimmed and the mail weights were not.

10. Factors Affecting Quality of Care

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. The sign and size of OLS mail survey coefficients were consistent with unadjusted differences.

Weighted versus unweighted regression adjusted mode coefficients. Weighting had little impact on coefficients. Six of the seven coefficients were significant, had the same signs, and were about the same size for both the weighted and unweighted regressions. One coefficient (lack of qualified specialists is a major problem) was significant (p<.10) for the unweighted but not for the weighted regression. However, the signs were in the same direction and there was less than one percentage point difference between the two coefficients.

11. Practice Revenue

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient. Mail survey coefficients were significant for all three tracking variables: percentage of practice revenue from Medicare and Medicaid and percentage of patient care revenue from capitation. The significance levels, signs, and size of OLS coefficients were consistent with unadjusted differences for Medicare and capitation revenue. However, the percentage of practice revenue from Medicaid was not significant for the unadjusted difference, but was significant and positive for the mail survey coefficient (2.7 percent).

Weighted versus unweighted regression adjusted mode coefficients. Weighting adjustments increased mode effects. Only one coefficient (percent capitated revenue) was significant for the unweighted regressions, but all three coefficients (percent capitated, percent Medicare, and percent Medicaid revenue) were significant for the weighted regressions. Also, the size of each of the three coefficients increased as a result of weighting.

12. Compensation and Income

Weighted unadjusted mode difference versus weighted regression adjusted mode coefficient:

- *Whether or not salaried.* The OLS mail survey coefficient, while significant, and smaller (-.076) than the unadjusted mode difference (-.121). While adjusting for practice and demographic characteristics reduced the mode effect, mail survey physicians were still less likely to be salaried than those participating in the CATI survey.

- *Compensation Factors.* Mail survey coefficients were significant and negative for all five compensation factors. For unadjusted mean differences by mode, consideration of the physician's own productivity was not significant, although the other four factors were significant and negative (mail survey value was smaller). The signs and sizes of OLS mail survey coefficients for rating measures (percentage important) were generally consistent with unadjusted differences. Exceptions included the importance of own productivity and satisfaction surveys, where the regression coefficients were significant but the unadjusted differences were not. Since salaried physicians were underrepresented in the mail survey and they may have had different views toward factors affecting their compensation than non-salaried physicians, we added whether salaried to the model. However, adding salary to the model for compensation factors did not affect mail survey coefficients.

- *Income.* Neither unadjusted differences in tracking measures by mode nor mail survey coefficients were significant.

Weighted versus unweighted regression adjusted mode coefficients:

- *Whether or not salaried.* Mail survey coefficients were not affected by weighting the regressions.

- *Compensation Factors.* Coefficients for key tracking measures (whether practice considers productivity, satisfaction surveys, quality of care measures, practice profiling, and financial performance in determining compensation) were unaffected by weighting. However, there were differences in significance levels for coefficients measuring the importance of each factor. The importance of productivity and satisfaction surveys were significant for the weighted but not the unweighted coefficients, and the importance of quality of care measures was significant for the unweighted but not the weighted coefficients. Adding salary to the model for the compensation questions resulted in the coefficients for both the weighted and unweighted measures of the importance of productivity being significant. Mail survey sample sizes were small for the follow-up questions on the importance of each factor, so these results should be viewed cautiously.

Income. Mail survey coefficients did not approach significance for weighted or unweighted regressions.

D. ITEM NONRESPONSE

For most questions, control over questionnaire logic and interviewer probing should reduce item nonresponse to a CATI compared with a self-administered mail survey. The final column of Table 9 (above)[7] compares unweighted item nonresponse rates between the CATI

---

[7] Table 9 only shows dependent variables included in difference of means and regressions; in this section, we also discuss item nonresponse rates for selected independent variables included in the survey. Item non-response rates for all variables are shown in Appendic C, Table D.

and mail surveys of physicians. We used unweighted rather than weighted measures because item nonresponse is largely a function of instrumentation and we wanted to assess the likely impact of mode change on future item nonresponse rates regardless of weighting methods used in subsequent surveys. Item non-response rates are defined as the sum of don't know, refused, and missing responses divided by the eligible sample (items that were legitimately skipped or coded as not applicable were excluded from the denominator).

1. Career Satisfaction

Item nonresponse rates were low for both modes.

2. Financial Incentives

Item nonresponse was low in both surveys, but was lower on the mail (1.8 percent) than CATI survey (4.4 percent).

3. Utilization of Time and Patient Mix

Item nonresponse rates were generally consistent. However, item nonresponse was higher on the mail survey for total hours in medically related activities (9.7 percent), compared with 0.2 percent on the CATI survey. Mail survey respondents sometimes omitted responses to hours in administrative tasks and professional activities and left the summation question blank. This type of error did not occur in the CATI instrument. Design of this item requires revision and testing for use in a future mail survey.

On the other hand, mail survey respondents were less likely to omit hours spent providing charity care (2.1 percent) than those completing CATI interviews (8.7 percent). It is possible that some physicians were reluctant to acknowledge to an interviewer that they didn't provide charity care and decided not to answer the question.

4. Information Technology and Care Management

Item nonresponse rates were comparable for the information technology questions but differed for the medical error and hospitalist questions. For the CATI survey, over one quarter of the eligible responses (27.7 percent) were omitted, compared with only 6.3 percent for the mail survey. Physicians who said that did not admit patients were excluded from the base for both modes. Review of the raw frequencies showed that nearly all of the CATI item nonresponse was due to values of "don't know." Although the questions were worded identically, physicians responding to the telephone interview may have been confused by the length of the question and complexity of the definition. Respondents to a mail survey have more time to review complex definitions before responding.

For the hospitalist question, item nonresponse was higher on the mail (6.0 percent) than the CATI (2.6 percent) survey. It is possible that the definition used to exclude mail survey physicians, which included two criteria, may have confused some physicians. (The questionnaire included an instruction that read, "If you did not admit any patients to a hospital in the last year or you are a practicing hospitalist, check the box.")

5. Perceptions of Quality

Item nonresponse rates were less than one percent for all three items on both surveys.

6. Cost Sharing

Item nonresponse rates were consistent between modes for all three items, but fairly high for the rating item concerning the choice between outpatient and inpatient care (6.9 percent CATI and 5.9 percent mail). Item nonresponse rates for the generic and test recommendations rating items were much lower. Perhaps physicians who rarely had to recommend inpatient versus outpatient care were less likely to answer the question on both modes.

7. Ability to Obtain Services

Item nonresponse rates were consistent between surveys for the questions concerning inability to obtain specialist referrals and outpatient mental health services, but were higher on the CATI (6.0 percent) than mail (1.6 percent) survey for non-emergency hospital admissions. However, mail survey respondents were less able to respond to the rating questions than were CATI respondents. For the specialist referral ratings, item nonresponse rates were under one percent for CATI and eight to nine percent for the mail survey. For the non-emergency hospital admission ratings, CATI item nonresponse rates were under two percent, but were 16-17 percent for the mail survey. (The number of physicians responding to the non-emergency admissions ratings also was very low—84 respondents, since relatively few physicians indicated they were unable to obtain this service.) Item nonresponse rates for the mental health services ratings were under one percent for CATI and slightly higher (two to three percent) for the mail survey.

Since mail survey item nonresponse rates were low for the rating questions on outpatient mental health services (the last service in the sequence), it is unlikely that questionnaire formatting was a key factor. It is more likely that the higher mail survey item nonresponse rates to the specialist referral and non-emergency hospital admissions rating questions reflected inability to answer the questions. Interviewer assistance apparently kept item non-response lower on the CATI rating questions.

8. Practice Acceptance of New Patients

Mail survey item non-response rates were slightly higher (by one to two percentage points) for most items. However, the gap was wider for the ratings concerning the practice's decision to limit new Medicaid patients because it already has enough patients (4.3 percent CATI versus 9.3 percent mail) and because Medicaid patients have high clinical burden (4.3 percent CATI versus 9.3 percent mail). On the other hand, respondents to the mail survey had slightly lower item nonresponse rates to the question on accepting all uninsured patients (1.4 percent) compared with CATI (4.7 percent).

9. Factors Affecting Quality of Care

Item nonresponse rates were consistent between the two modes and less than or equal to two percent for all items.

10. Practice Revenue

Item nonresponse rates were moderate for the mail survey (8 to11 percent) and high for the CATI survey (15 to 17 percent). Self-administration sometimes yields higher item response rates to quantitative questions, assuming the respondent has the time and ability to answer. Time pressure during a telephone interview may force some respondents who do not have an answer readily available to say they are unable to respond.

11. Compensation and Income

Item nonresponse rates were low for the compensation items. For the income question based on intervals, they were higher on the mail (12.3 percent) compared with the CATI survey (4.7 percent). As noted above, the structure of the mail survey income question may have confused some respondents. To maintain comparability with the CATI survey, we first asked mail survey respondents to estimate their net income from medical practice to the nearest $1,000. Respondents were given the option of checking a box indicating that they cannot estimate their income to the nearest $1,000 and were then asked to estimate their income in broad intervals. As noted above, 35 percent of the sample did not answer the open ended question, so that question was not useable. For a future mail survey, we recommend only asking the interval question. Overall item non-response might be lower if we avoid the confusion of a two-part income question on a self-administered survey.

12. **Demographic Characteristics (see Appendix C, Table D)**

Item nonresponse rates were low for both modes, except for number of physicians in the practice, where the item nonresponse rate for the mail survey (5.1 percent) was much higher than for the CATI survey (1.0 percent). Mail survey formatting (Q.7) may have confused some doctors because the instruction is different for practices with fewer than 50 physicians (arrow directing the respondent to an item that asks "How many?") versus 50 or more physicians (a check box). An alternative that may be simpler would be a check list with intervals, e.g., one, two, three, four five, 6-10, 11-25, 26-50, more than 50. Item nonresponse for the ethnicity question also was low for both modes, but wasslightly lower for the mail (2.3 percent) than CATI survey (3.7 percent).

E. SUMMARY

To facilitate decisions regarding retention of tracking questions if HSC shifts from a CATI to mail survey, below we have summarized our findings for unadjusted differences, regression adjusted diffrences, and item non-response rates by topic area.

1. Career Satisfaction. Although physicians responding to the CATI survey were more likely to say they were very satisfied with their careers than were those completing mail questionnaires (the socially desirable response), there was no difference by mode for dissatisfied in career, which is the tracking measure. This pattern indicates that while mode effects may occur for extreme values (very satisfied), they may not affect measures that combine moderate (somewhat dissatisfied) and extreme (very dissatisfied) measures.

2. *Ownership.* Practice ownership is unaffected by data collection mode for weighted unadjusted and regression adjusted estimates. However, mail survey respondents were more likely than CATI respondents to own their practices in the unweighted regressions, indicating that weights help to correct for differences in ownership response patterns.

3. Financial Incentives. Physicians responding to the CATI survey were more likely than mail survey respondents to say that financial incentives do not affect the provision of services, which may be the more socially desirable response since it indicates that financial incentives do not affect behavior. The effect of mode on the impact of financial incentives in reducing services is more ambiguous, with mail survey respondents more likely to say that financial incentives reduce services for the unweighted regression, but not for the weighted regression.

4. Utilization of Time and Patient Mix. Most items have mode effects, with possible social desirability bias. CATI respondents report more hours worked, more hours in charity care, and say they treat more patients from minority groups and with chronic conditions. Given the importance of these measures, some or all items may be retained; however, it is likely that self-administered measures will be lower than CATI.

5. Information Technology and Care Management. Most items either had no mode effects or the mail survey measures were larger, which is likely given the temporal gap between the two surveys and the increasing use of information technology by medical practices. In addition, item nonresponse was much lower for the mail survey medical error item, which also had no mode effect.

6. Perceptions of Quality. Weighted differences for the tracking measures (adequate time, freedom to make clinical decisions, and quality care) were not significant. However, the unweighted coefficient for the percentage disagreeing that they have the freedom to make clinical decisions was positive and significant.

7. Cost Sharing. Physicians responding to the mail survey were more likely to usually or always consider out-of-pocket costs. The results are difficult to interpret. One would have expected that the socially desirable response would be to consider out-of-pocket costs, which was reported more frequently for the mail than CATI survey. On the other hand, physicians responding to an interview may be less likely to acknowledge cost considerations even if they reduced the patient's out-of-pocket costs.

8. Ability to Obtain Services. Mail survey respondents were more likely to say they were unable to obtain specialist referrals and less likely to be unable to obtain non-emergency hospital admissions than CATI respondents. Item nonresponse rates also were high for the mail survey

rating questions. Given the complexity of the CATI wording and logic, it will be difficult to track these questions in a mail survey.

9. Practice Acceptance of New Patients. The key tracking measures are acceptance of some or all new patients for each major payor and the uninsured. Weighting the regressions increased the percentage of mail survey physicians, compared with CATI physicians, who said they would restrict access to new patients by accepting no new patients or only some new patients. For the weighted regressions, mail survey respondents were more likely than CATI respondents to limit access to Medicare, privately insured, and uninsured patients. Respondents to the CATI survey also were more likely to say they accept all new patients for all payors and uninsured patients. These results indicate that respondents to the CATI survey were more likely than respondents to the self administered survey to provide the socially desirable response (acceptance of all new patients) and less likely to acknowledge that their practices limit access to new patients. Weighting increased this pattern. Differences for most items rating reasons why the practice limits access to Medicare and Medicaid were not significant. Item nonresponse rates for rating items were slightly higher for the mail survey, indicating that respondents to the self-administered survey had slightly more difficulty following the questionnaire logic.

10. Factors Affecting Quality of Care. These questions had large mode effects and will be difficult to track if HSC shifts to a mail survey. Respondents to the mail survey were more likely to cite practice barriers (inadequate time with patients), financial limitations (patients inability to pay), and interference by insurers (rejection of care decisions) as major problems than CATI respondents. Telephone survey respondents were more likely to mention physician or hospital related issues (timely reports and medical errors in hospitals) or language barriers as problems. Relatively few physicians in either survey cited lack of qualified specialists as a major problem and there was no difference by mode.

11. Practice Revenue. Mode for all three practice revenue questions (percent capitated, percent Medicare, and percent Medicaid revenue) altered weighted estimates by several percentage points. (However, only percent capitated revenue was significant for the unweighted regressions.) Item non-response rates were lower for the mail survey for all items. Given the importance of the Medicare and Medicaid revenue items, inconsistency between the weighed and unweighted regressions, and the absence of a theoretical reason for social desirability biases, these items could be retained with the recognition that mode could confound tracking in round five.

12. Factors Determining Compensation. CATI survey respondents were more likely to say their practices consider each factor in determining compensation. Since CATI respondents also were more likely to be salaried, and salaried physicians may be more likely to cite these factors, we included salary as a control variable; however, the sign and size of the mail survey dummies were unaffected. Since these factors are factual items, it is not clear that social desirability was behind the mode difference. On the other hand, the regression model controlled for demographic and practice characteristics (including practice type and whether salaried), and the questions were worded and scaled the same. A possible explanation is that some CATI respondents may have been fatigued at the end of the survey and acquiesced when the interviewer asked the questions. Some studies have shown greater acquiescence on personal or telephone interviews

than on self-administered instruments (Dillman, 2000; 2004). In any case, tracking compensation factors between a CATI and self-administered survey will be difficult.

13. Income.   Estimates for the income question were not affected by mode, although item nonresponse was higher for the mail survey. However, shifting to a single interval-based income question may reduce item non-response in a self-administered survey.

Continuing the CTS physician survey by mail would dramatically reduce costs, sustain or increase response rates compared with a telephone survey, but would result in mode effects that are pervasive and often sizable. A key goal for the next round of the physician survey is evaluating statistical methods to determine whether mode effects can be controlled to permit continued tracking of key CTS measures.

REFERENCES

Armstrong, J. Scott . "Monetary Incentives in Mail Surveys." *Public Opinion Quarterly,* vol. 39, 1975, pp. 111-116.

Audet, AM, Doty, M, Peugh, J, et al. "Information Technologies: When Will They Make It Into Physicians' Black Bags?," *Medscape General Medicine,* Dec. 7, 2004.

Ballou, J, Milliner-Waddell, J and B Roff. *National Survey on Medical Professionalism*, report submitted to the Institute for Health Policy, Harvard University, September 2004.

Church, AH. "Incentives in Mail Surveys: A Meta-Analysis." *Public Opinion Quarterly,* vol. 57 1993, pp. 62-79.

de Leeuw, E and J. Van Der Zowen. "Data Quality in Telephone and Face-to Face Surveys: A Comparative Meta-Analysis." In *Telephone Survey Methodology,* edited by Groves, R, Biemer, P, Lyberg, L, Massey, J, Nicholls II, W, and J. Waksberg, J. New York: John Wiley & sons, 1988, pp. 283-299.

Dillman, DA and LM Christian. "*Survey Mode as a Source of Instability in Responses across Surveys,."* (Forthcoming in the journal, *Field Methods*). Paper presented at the Workshop on Stability of Methods for Collecting, Analyzing and Managing Panel Data, American Academy of Arts and Sciences, Cambridge, MA March 27, 2003.

Dillman, D. "*Mail and Internet Surveys, The Tailored Design Method"* 2nd Edition, Wiley, 2000.

Dillman DR, Sanster, R, Tarnia, J, and T. Rockwood. "Understanding Differences in People's Answers to Telephone and Mail Surveys," In *Current Issues in Survey Research, New Directions for Program Evaluation Series,* edited by ed. by Braverman and Slater San Francisco: Jossey-Bass, 1996, pp. 45-62.

Fowler FJ, Roman AM, and ZX Di. "Mode Effects in a Survey of Medicare Prostate Surgery Patients." *Public Opinion Quarterly,* vol. 62, 1988, pp. 29-46. *In the text, this is referred to as Fowler, Roman, and Dillman.*

Fox, RJ, Crask, MR and K Jonghoooo. "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response." *Public Opinion Quarterly* vol. 52 1988, pp. 467-491.

Heberlein, TA, and R Baumgartner. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature," *American Sociological Review* vol 43 (4) 1978 pp. 447- 462.

James, JM, and R Bolstein. "The Effect of Material Incentives and Follow-up Mailings on the Response Rate and Response Quality in Mail Surveys." *Public Opinion Quarterly,* vol. 54 1990, pp. 346-361.

James, JM, and R Bolstein "Large Monetary Incentives and their Effect on Mail Response Surveys Rates." *Public Opinion Quarterly* vol 56, 1992, pp. 442-453.

Kasprzyk, D, Montaño, DE, St.Lawrence, JS, and WR Phillips. "The Effects of Variations in Mode of Delivery and Monetary Incentive on Physicians: Responses to a Mailed Survey assessing STD Practice Patterns," *Evaluation and the Health Professions,* vol 24, 2001, pp. 3-17.

Martinez-Ebers, V. "Using Monetary Incentives with Hard-to-Reach Population in Panel Surveys,". *International Journal of Public Opinion* vol9 1997, pp. 77-86.

Montano, D, Kasprzyk, D, Carlin L, Freeman, C, and J. Christian. *HPV Clinician Survey: Knowledge, Attitudes, and Practices about Genital HPV Infection and Related Conditions*, final report from Battelle's Center for Public Health Research and Evaluation to The Centers for Disease Control and Prevention, July 2005

Pugh N, Iannacchione, V, Trang, L, and L Dimitropoulos. "Evaluating Mode Effects in the Medicare CAHPS Fee-For-Service Survey," RTI International, Research Triangle Park, NC 27709, Joint Statistical Meetings, Section on Survey Research Methods, 2002.

Stringfellow, V, Fowler, F, and B Clarridge. "Evaluating Mode Effects on A Survey of Behavioral Health Care Users," Proceedings of the Annual Meeting of the American Statistical Association, August 2001.

Singer, E, Van Hoewyk, J. and MP Maher. "Does the Payment of Incentives Create Expectation Effects?" *Public Opinion Quarterly* vol. 62 1998, pp.152-164.

Strouse, Richard, *Findings from Surveys on Public Attitudes Toward Tobacco Policy Issues for Smokeless States Grantees*, Final Report submitted by Mathematica Policy Research Inc. to The Robert Wood Johnson Foundation, November 2004.

Tambor, ES, Chase, GA, Faden, RR, Geller, G, Hofman, KJ and NA Holtzman. "Improving Response Rates Through Incentives and Follow-Up: The Effect of a Survey of Physicians' Knowledge of Genetics." *American Journal of Public Health* vol 83 1993, pp. 1599-1603.

Warriner, K, Goyder, J, Gjertsen, H, Hohner, P and K McSpurren. "Charities, No; Lotteries, No; Cash, Yes: Main Effects and Interactions in a Canadian Incentives Experiment." *Public Opinion Quarterly* vol 60, 1996, pp. 542-562.

Yammarino, F, Steven J, Skinner, J, and TL Childers. "Understanding Mail Survey Response Behavior: A Meta- Analysis." *Public Opinion Quarterly* vol. 55, 1991, pp. 613-639.

Yu, J, and H Cooper. "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires." *Journal of Marketing Research* vol. 20 1983, pp. 36-44.