Appendix D: Methods Used in the OTA Clinical Scenario Surveys

This appendix summarizes the methods used to develop and analyze surveys of three physician professional societies. The Office of Technology Assessment (OTA) cooperated with three physician associations to conduct clinical scenario surveys of association members by mail from February through August of 1993.[1] The three physician associations, listed in the order in which they were surveyed, were:

- the American College of Cardiology (ACC),
- the American College of Surgeons (ACS), and
- the American College of Obstetricians and Gynecologists (ACOG).

The ACS component actually involved two separate surveys: one for general surgeons and the other for neurosurgeons. Thus, four distinct surveys were actually conducted.

The questionnaire for each survey was developed jointly between OTA and the respective association. ACC maintains an ongoing "practice panel" sample of its practicing members and conducted its own mailout, data entry, and initial data editing. For the other two surveys, these tasks were shared between OTA and the respective association. OTA performed all final data editing, processing, and analysis. Strict rules protecting respondent confidentiality were observed by all participating organizations.

SURVEY INSTRUMENT CONTENT AND FORMAT

The main goal of each survey was to ascertain, as unobtrusively as possible, the extent to which physicians would choose "malpractice concerns" from among several reasons for selecting or rejecting specific diagnostic or therapeutic procedures in treating specific hypothetical cases. Respondents were presented two or three specific clinical scenarios appropriate to their respective specialties. Introductory letters from both the physician association and OTA described the purpose of the survey in general terms, without mentioning malpractice or defensive medicine. Two separate instruction pages, including an example scenario, explained how the questionnaire should be

[1] Dr. Russell Localio of Pennsylvania State University and Dr. Jeremy Sugarman of Duke University were consultants to OTA on the design of the survey instruments and statistical analysis. Dr. Localio designed the sampling plan and data analysis components of the surveys and participated extensively in the analysis and interpretation of the survey results. Dr. Sugarman consulted on the development of the format and content of the clinical scenarios used in the surveys.

completed. Copies of a all survey instruments are presented in a technical appendix available from OTA upon request.

Clinical Scenarios

Scenario Format and Content

The clinical scenarios in each of the four surveys were developed by an expert panel containing from seven to 10 members of the relevant physician association (selected by association leadership in cooperation with OTA project staff and consultants). During a one-day meeting at the association headquarters, the panel members were asked to "brainstorm" at least 20 clinical scenarios in which concerns about liability would be expected to strongly influence clinical actions. Then the panel was asked to select from these candidates three or four scenarios that would be expected to elicit the strongest defensive medicine responses for inclusion in the survey.

Panel members were also asked to create a ● 'control" version of each selected case by adding or deleting one or more key clinical indicators (e.g., a result from a laboratory or radiologic test) that would, in the opinion of the panelists, greatly reduce the likelihood that malpractice concerns would be cited as the primary reason for choosing any action. OTA staff and consultants then selected and refined the final scenarios, with input from association leaders and panel members. Each questionnaire was pretested on a small sample of association members who were excluded from the final survey.

Each clinical scenario:

- described the patient's demographic characteristics, symptoms, vital signs, and initial diagnostic test results;
- presented between 3 and 13 diagnostic or therapeutic procedures, including the option of essentially doing nothing; and
- presented four reasons for choosing or rejecting each procedure:
  - —medical indications,
  - —concerns about costs versus benefits,
  - —malpractice concerns, and
  - —patient expectations.

"Other (specify)" was also a choice under both the procedures and the reasons for choosing them.[2]

The respondent was asked to:

- choose "yes" or "no" for each procedure,
- check one or more reasons for that choice, and
- double-check the most important reason for the choice.

Only one double-check was allowed for each procedure. These choices were presented in a grid format, with the procedures as rows and the reasons as columns. The first "procedure" listed was typically "do nothing," and the rest were diagnostic and therapeutic interventions with varying degrees of "invasiveness" or technological sophistication.

Case and Control Scenarios

ACC and ACS respondents each received two scenarios, while ACOG respondents received three (see below). In each survey, the "case" version of one scenario was given to a randomly chosen subgroup of respondents, and the "control" version of that same scenario was given to the remaining respondents. One or two additional scenarios in each survey, referred to here as "common" scenarios, were sent to all respondents. Thus, the first randomly selected subgroup of surveyed physicians received one or two scenarios (all of which were selected because concern about liability was

[2] In place of "other," the ACC survey used "institutional protocols/professional guidelines" as the fifth reason. Although "other" was listed as a procedure on the ACC questionnaire, the association did not code the presence or absence of a written response in that box. Consequently, OTA was unable to Include "other procedure" in its analysis of the ACC data.

expected to be frequent); the other received the control scenario and one or two common scenarios, The specific combination of scenarios presented to each group of respondents is summarized in table D- 1. Special analytical problems posed by this case-control design are discussed later in this appendix.

Open-Ended Version of the ACS General Surgeon Survey

A supplemental sample of general surgeons was sent an "open-ended" version of each ACS clinical scenario used in the main survey of general surgeons (case versions only—see previous section). The open-ended questionnaire offered no specific "reasons" for choosing procedures. Instead, a blank space was provided beside each procedure, in which respondents could fill in their own reasons, in their own words, for choosing the procedure. A senior OTA staff member coded the responses on these open-ended questionnaires into the categories of "reasons" given in the main questionnaire. Responses were coded as citing "malpractice concerns" if they contained any suggestion at all of defensive practice (e.g., ". . . to cover myself").

Attitudinal and Demographic Items

Each survey instrument contained items on two or three professional or demographic characteristics (e.g., practice setting) that were particularly relevant to malpractice issues within that specialty.[3] The instrument also contained a set of attitudinal items provided to OTA by Dr. Susan Goold of the University of Michigan, who had developed and tested three composite scales based on those items (77). For this report those attitude scales were labeled as follows:

TABLE D-1: Combinations of Clinical Scenarios in OTA Surveys of Defensive Medicine

Association	Group	Scenario 1 (case/control)	Scenario 2 (common)
American College of Cardiology	Group 1 (case)	Chest pain case	Syncope
	Group 2 (control)	Chest pain control	Syncope
American College of Surgeons			
General surgeons	Group 1 (case)	Rectal bleeding case	Breast pain
	Group 2 (control)	Rectal bleeding control	Breast pain
Neurosurgeons	Group 1 (case)	Back pain case	Head injury
	Group 2 (control)	Back pain control	Head injury
American College of Obstetricians and Gynecologists	Group 1 (case)	Perimenopausal bleeding case	Breast lump Complicated delivery
	Group 2 (control)	Perimenopausal bleeding control	Breast lump Complicated delivery

SOURCE Off Ice of Technology Assessment, 1994

[3] These characteristics were jointly selected by staff members of OTA and the relevant physician association, considering not only differences among the specialties, but also the unavailability of some characteristics in each association's membership database (also see the section on sampling, below). Most importantly, the following measures were not available: in the ACC survey, the number of years in practice; in the ACS survey, geographic region; and in the ACOG survey, whether the respondent held an academic appointment. Also, the categories of the respondent's usual practice setting differed slightly from survey to survey, reflecting the different categories used by the associations themselves. Finally, as measures of the number of years in practice, ACS used years since board certification, whereas A COG used years of membership m the association. These unavoidable variations in measurement reduced the comparability of results from the four surveys.

- Malpractice Concern,
- Cost Consciousness, and
- Discomfort with Clinical Uncertainty.

Additional items regarding satisfaction with medical practice were developed by OTA and Dr. Goold to serve as decoy items in the surveys.

Each attitude item offered five response categories, scored as 1 through 5 (respectively): strongly agree, agree, unsure, disagree, and strongly disagree. The Malpractice Concern scale contained five items, the Cost Consciousness scale contained six items, and the Discomfort with Clinical Uncertainty scale originally contained three items. However, OTA did not use the entire Uncertainty scale for the ACOG survey (only one Uncertainty item was included in that survey), after receiving written comments from ACS respondents regarding how similarly worded the items were.

Each respondent's scores (1 through 5) on all the items in a given scale were summed to obtain a total scale score.[4] To make a "5" represent agreement rather than disagreement (so that the summed scores would measure agreement), the item scores were reversed by subtracting them from 6, except where an item was worded negatively (e.g., where agreement represented low malpractice concern). The scores for the five-item Malpractice Concern scale thus ranged from 5 (minimal malpractice concern) to 25 (maximal malpractice concern), whereas the six-item Cost Consciousness scale ranged from 6 (minimal cost consciousness) to 30 (maximal cost consciousness). The three-item Uncertainty scale, which ranged from 3 (minimal discomfort with clinical uncertainty) to 15 (maximal discomfort with clinical uncertain y), was computed on] y for ACC and ACS respondents because the ACOG survey contained only one Uncertainty item (see above).

SAMPLING

OTA and its consultant, Russell Localio, developed a sampling plan for each survey, with input from association staff. Sampling fractions were based on statistical power calculations for two-sample comparisons, with rough assumptions about the survey response rate and the number of respondents who would choose clinical procedures primarily because of malpractice concerns. Sampling fractions varied across sampling strata to ensure adequate numbers of respondents in each subclass of physicians. Each physician association then drew a sample from its membership database according to detailed instructions provided by OTA. Population sizes, sample sizes, numbers of respondents, and response rates for each survey are displayed in table D-2. All four surveys targeted only association members who, according to the membership database:

- had earned the degree of either Medical Doctor (MD) or Doctor of Osteopathy (DO).
- were not in residency training,
- were not retired,
- were board certified in the relevant specialty, and
- were currently practicing in the United States.

All four samples were drawn from the association's membership database through systematic stratified random sampling. However, due to 1imitations of the membership databases and special association concerns, the stratification factors differed somewhat from survey to survey. These and other features of the four samples are summarized in table D-3. Other differences also existed among the four samples:

- ACC used its existing "Professional Practice Panel," a standing sample of about 1,500 practicing members who are occasional] y surveyed

[4] Dr. Goold reported that this simple additive approach was most appropriate, given that factor analysis had failed to create satisfactory composite scales with weighted individual items (76).

TABLE D-2: Samples for OTA Clinical Scenario Surveys of Defensive Medicine

Survey	Group	Population	Sample	Respondents[a]	Response rate
American College of Cardiology[b]	Total	11,541	622	352	566
	Case		311	184	591
	Control		311	168	540
American College of Surgeons					
General surgeons	Total	12,972	3,004	1,793	597
	Closed-ended		2,401	1,412	588
	Case		1,196	739	618
	Control		1,205	673	559
	Open-ended		603	381	63.2
Neurosurgeons	Total	1,384	859	503	586
	Case		427	252	59.0
	Control		432	251	581
American College of Obstetricians and Gynecologists[c]	Total	20,832	1,983	1,230	623
	Case		1,002	634	633
	Control		981	596	608

a The numbers of respondents shown In this table may differ silghtly from the scenario-specific numbers of respondents shown in text tables 3-2 through 3-5 in chapter 3 because a few respondents completed one scenarlo but not the other
b The American College of Cardiology sample included only adult cardiologists
c The American College of Obstetrlclans and Gynecologists sample excluded gynecological oncologists and reproductive endocrinologists

SOURCE Office of Technology Assessment, 1994

on various issues regarding the practice of cardiology. This sample is drawn using similar methods to those used in the ACS and ACOG surveys (see table D-3). For this survey, only adult cardiologists on the panel as of February 1993 were included. As with the ACS and ACOG samples, questionnaires were sent to all 622 adult cardiologists on the ACC panel. Their overall response rate was slightly lower than the response rates in the ACS and ACOG surveys (see table D-2). ACC panel members may have been more sensitized to practice issues raised by previous surveys.

- The ACOG survey excluded gynecological oncologists and reproductive endocrinologists. The sample size was limited to 2,000 to meet administrative and budgetary constraints at both OTA and the association.
- In both the ACC and ACOG surveys, a second mailing of the questionnaire was sent to members who had not responded to the first mailing. In the ACS survey, one mailout was used because the association preferred not to track individual respondents. The method of identifying each respondent's sampling stratum is described in the next section.
- The ACS survey included physicians practicing in U.S. territories (Puerto Rico, Guam, etc.), whereas the ACC and ACOG surveys did not.
- The ACC and ACS surveys contained government-employed physicians, including military doctors (except those practicing overseas,), whereas the ACOG sample excluded military physicians.

In the ACS and ACOG surveys, the numbers of case and control respondents were not equal, for two reasons. First, for ease of data processing, random assignment of respondents to the case or control group (every other respondent) was performed within each sampling stratum rather than throughout the entire sample. In the ACC survey, the overall numbers of case and control respondents were equal; however, the case respondents were selected by taking a simple random subsample of the overall sample, without regard to the stratification variable of geographic region. Second. response rates differed slightly between the

TABLE D-3: Features of Sampling Plan for OTA Clinical Scenario Surveys of Defensive Medicine

Feature	American College of Cardiology[a]	American College *of* Surgeons	American College of Obstetricians and Gynecologists[b]
Stratification factors	Census region	Academic appointment yes, no Year of first board certification post-1981, 1972-81, pre-1972 Practice setting solo, group, medical school, hospital, other	Geographic region (4 regions) Years in ACOG < 6, 6-10, 11-20, >20 Gender
Number of strata	9	30, plus two additional, one for some missing data, the other for all missing data	32
Special exclusions[c]	U S trust territories	None	U S trust territories, military, Public Health Service
First mailing	Feb. 4, 1993	March 4, 1993	May 27,1993
Second mailing	Feb. 23, 1993	None	June 30, 1993

a The ACC survey ıncluded only adult cardiologists
b The ACOG survey excluded gynecological oncologists and reproductive endocrinologists
c For general exclusion criteria see text

SOURCE Office of Technology Assessment 1994

case and control groups. The numbers of case and control respondents therefore differed within each region by as much as 11 percent. Differences in response rates were corrected by reweighting the respondents according to case/control group and sampling stratification factors (e.g., region).

DATA PROCESSING

ACC conducted its own mailouts, data entry, and initial data editing. Individual respondents were tracked, and initial nonrespondents were sent another copy of the questionnaire. In the ACS and A COG surveys, the general procedure was as follows:

- The association providcd OTA with mailing labels for sampled members.
- OTA produced the questionnaires and mailed them with a prepaid return envelope addressed to the association's Washington. DC. office.
- Upon receiving the responses, the association photocopied them and shipped the originals to OTA for processing.

There were several variations on this basic process between the ACS and ACOG surveys. The identity of individual ACOG respondents was tracked by ACOG personnel by means of a relatively unobtrusive identification number printed on the first page of the questionnaire as well as on the mailout label and the postage-paid return envelope. As noted earlier, a second mailing of the ACOG questionnaire was sent to initial nonrespondents. Five such respondents apparently returned both questionnaires, for they had duplicate ID numbers. We allowed one of each pair of data records for these duplicate respondents to be randomly discarded through a computer sorting and matching routine (see the next section).

ACS, on the other hand, preferred not to track individual respondents; thus, no followup mailing of the questionnaire to initial nonrespondents was possible. To track the sampling stratum to which the respondent belonged, OTA devised a method of unobtrusively tracking the respondent's sampling stratum by varying the features of the return mailing label.

Eighty-nine respondents did not use the return envelope provided but instead sent the questionnaire back in an "irregular" envelope (i e., without the tailored mailing label). For 61 of these respondents (68.5 percent), ACS was able to use the return address or postmark on that envelope to identify the sampling stratum to which the respondent belonged. ACS kept the individual identity of these 89 respondents confidential.

OTA made no attempt to identify any individual respondents and analyzed all data separately from any identifying materials.

DATA EDITING AND ENTRY

The major rules used to edit the data in all four surveys are summarized in a technical appendix available from OTA upon request. OTA and the associations made concerted efforts to refine the questionnaire instructions based on responses to the three pretests. Despite these precautions, respondents in all four surveys sometimes provided answers that were inconsistent with the instructions; these responses required editing.

The most frequent ● *error" was failure to circle "no" for unselected clinical options or failure to check the reasons for circling "no" for such options. That is, many respondents circled '*yes" only for selected options and checked reasons for choosing only those options. Fortunately, this kind of "error" did not substantially affect the analysis, which focused on respondents who chose "yes" for a given option (see the next section).

Another very infrequent "error" (on the order of 0.1 to 0.6 percent of all responses) that would affect the analysis was failure to check reasons for c1inical options where "yes" was circled. These respondents (who circled "yes" for an option but failed to check any reasons for doing so) were included in the denominator when the percentage of "choosers" (see below) was calculated—implying that, if the respondent had cited a reason, it would *not* have been "malpractice concerns." The alternative approach—to exclude such respondents from the denominator of that percentage—would have further reduced the size of that denominator, which might have slightly weakened the reliability of the analysis.

All edits of the ACS and ACOG data were performed by OTA. ACC performed similar edits on its own data. After receiving the data from ACC (see below), OTA then made further edits that had not been performed by ACC.

Data for all four surveys were key-entered by the same contractor (Office Remedies, Inc., of Vienna, Virginia) with double-entry verification. Keyed data were returned to OTA in database files on floppy diskettes. (ACC contracted directly with Office Remedies, Inc.) OTA converted these files into SAS (203) format for analysis on a microcomputer using both SAS-PC and SUDAAN (193), a program that computes variance estimates properly weighted for disproportionate stratified sampling and nonresponse. We also used StatXact-Turbo (49) for analyses involving small numbers of respondents, for which large-sample statistical methods might be inappropriate. The use of these programs is discussed in further detail below.

DATA ANALYSIS

■ General Approach

The focus for the analysis of all four surveys was the percentage of respondents who cited "malpractice concerns" as a reason for choosing a diagnostic or therapeutic procedure in a given scenario—i.e., positive defensive medicine (see chapter 2). Analysis of "malpractice concerns" as a reason for choosing *not* to perform a procedure (a form of negative defensive medicine—again see chapter 2) was deemed to be outside the scope of the study.[5] The analysis thus focused on respondents who chose "yes" for one or more procedures (and

[5] A possible exception here is the clinical option of "refer to surgeon," which appeared in the ACOG breast lump scenario. Physicians who chose this option had possibly decided not to intervene themselves (depending on whether they chose to perform other procedures I isted in the scenario), and thus may have been engaging in negative defensive medicine. On the other hand, referral to a surgeon can imply an expectation that relatively aggressive and potentially costly intervention will he undertaken, and may thus reflect positive defensive medicine.

hence chose "no" for the "do nothing"' option). Thus, for each procedure. the denominator was the group of respondents who chose "yes" for that procedure. Excluded from this denominator were not only respondents who explicitly chose "no," but also those who chose neither "yes" nor "no" (i.e., those who had left that entire row of the questionnaire blank). Respondents who did not respond at all to a given scenario, but who responded to other parts of the questionnaire, were excluded only from the analysis of that particular scenario.

Of this denominator (respondents who chose "yes" for a given procedure), the numerator of greatest interest was the group of respondents who checked "malpractice concerns" as a reason for choosing that procedure (with either a single- or double-check). However, the "malpractice" responses could not be analyzed in isolation, because another reason (usually "medical indications") was often cited along with "malpractice concerns" by the same respondents. This meant that these respondents were selecting procedures not only on the basis of malpractice concerns, but also in part because they felt that the procedures were at least somewhat medically indicated. These combinations of responses suggested that differing degrees or levels of defensive motivation were being expressed in these surveys. each of which required a separate measure. Tables showing the distribution of responses by clinical procedure and reason for procedure choice are presented in a technical appendix available from OTA upon request.

Specific Measures of Defensive Medicine

To gauge the extent of "defensive medicine" expressed in these surveys, we constructed six measures of defensive medicine based on specific patterns of reasons given for choosing a given diagnostic or therapeutic procedure. These response patterns involved particular combinations of check marks for "malpractice concern s," "medical indication s," and other reasons. The six measures are 1isted in order below from the most restrictive definition of defensive medicine to the least restrictive definition. The measures are cumulative, i.e., the least restrictive measure (measure 6) includes respondents meeting measures 1 through 5.

Measure 1:
DOUBLE check for "malpractice concerns"
AND
NO check at all for ANY other reason.

Measue 2:
Measure 1 PLUS
a DOUBLE check for "malpractice concerns"
AND
NO check for "medical indications"
(single checks for other reasons are allowed).

Measure 3:
Measure 2 PLUS
a DOUBLE check for "malpractice concerns"
AND
a SINGLE check for "medical indications"
(single checks for other reasons are allowed).

Meusure 4:
Measure 3 PLUS
a SINGLE check for "malpractice concerns"
AND
NO check for "medical indications"
(single or *double* checks for other reasons are allowed).

Measure 5:
Measure 4 PLUS
a SINGLE check for "malpractice concerns"
AND
a SINGLE check for "medical indications"
(single or double checks for other reasons are allowed).

Measure 6:
Measure 5 PLUS
a SINGLE check for "'malpractice concerns"
AND
a DOUBLE check for "medical indications"
(single checks for other reasons are allowed}.

The rationale underlying these measures is as follows. Defensive medicine is most strongly indicated when the respondent cites only "malpractice

concerns" and no other reason (measure 1). Even though there are no medical indications or patient expectations for performing the procedure, the physician would perform it anyway, solely out of fear of malpractice litigation. This response should be infrequent, since it is arguably a violation of medical ethics. Citing other reasons, particularly "medical indications," "dilutes" the degree of defensive medicine indicated. Moreover, a single check for '*malpractice concerns" represents a weaker level of defensive medicine than does a double check.

These six measures of defensive medicine were computed on the basis of two different denominators, thereby creating two separate measures that provide two different interpretations of the results for a given procedure in a given scenario:

Percentage of "choosers": Here the denominator was the number of respondents who would choose the procedure (i.e., circled *'yes"). The measure of defensive medicine was thus the percentage of respondents choosing the procedure who cited "malpractice concerns" as a reason for doing so.

Percentage of scenario respondents: Here the denominator was the total number of respondents to the overall scenario. The measure of defensive medicine was thus the percentage of all respondents who, when presented with the scenario, would choose the procedure for defensive reasons. This percentage was much smaller than the percentage of choosers and represents the frequency with which concerns about malpractice would be expected to enter clinical decisions in situations of this type.

With six separate measures of defensive medicine, the number of comparisons between the percentages for various groups of respondents (case versus control, academic versus nonacademic, etc.) would have been unmanageable. Consequently, for such comparisons we used only measure 3 (double-check for "malpractice concerns," with single checks allowed for any other reasons, including ● 'medical indications"). This measure most closely approximated OTA's working definition of positive defensive medicine: physicians performing procedures *primarily,* but *not necessarily solely,* out of fear of malpractice litigation (see chapter 2). Tables showing the distribution of responses on all six measures of defensive medicine are presented in appendix E.

Statistical Analysis

All data were treated as coming from a sample survey with unequal probability of selection in a stratified (cross-classified) population (114,117, 124). Compared with simple random sampling, the effect of weighting the data to compensate for unequal probability of selection is generally to increase the variance of estimators, while the effect of stratification is generally to reduce that variance. Data from the surveys supported our reliance on this general experience. Test analyses using methods for 1) unweighed simple random samples, 2) weighted simple random samples, 3) unweighed stratified samples, and 4) weighted stratified samples demonstrated that the effects of stratification and weighting in fact did offset each other to a considerable degree. Variances were not increased markedly owing to the use of unequal weights in this sampling design.

Rates (or proportions) of respondents who would choose a clinical procedure, and of those who did so primarily because of malpractice concerns (see above), were calculated using sampling weights that compensated for nonresponse as well as unequal probability of selection across the sampling strata. Wherever possible, variance estimates and confidence intervals for these point estimates used methods that are common in survey analysis and assumed both stratification and sampling without replacement (i.e., use of the finite population correction).

Where possible, comparisons among subclasses of respondents were made by differences in rates (or proportions), and calculations of the variance of those differences took into consideration the sampling design. In several instances we departed from the use of rate differences in

comparing populations. In those cases, we used a sample-weighted logistic regression model (15,16) to compute odds ratios that tested for differences among groups of respondents, while controlling for a third factor.

Assumptions of simple random sampling were used only when data were too sparse to use survey sampling methods, owing to the small numbers of respondents (fewer than 40) who would choose procedures primarily because of malpractice concerns in some of the clinical scenarios. As a fallback method, in these cases we used StatXact-Turbo (49), a software package with advanced numerical algorithms that are especially appropriate for sparse data, i.e., where the numbers of respondents and the rates of citing malpractice concerns are small. The advantage of this additional analysis tool is the ability to produce confidence intervals and p-values that do not overstate the significance of results. The disadvantage is the risk of bias from the use of unweighted data: StatXact-Turbo software (49) assumes simple random sampling (unstratified) and cannot handle weighted data. Use of unweighted data had little effect on the point estimates, however, except when only one or two respondents cited malpractice concerns and their individual sampling weights were large. In those cases both the weighted and unweighted rates were close to zero. For these very small frequencies in this survey, therefore, reliance on StatXact as an alternative tool was acceptable. In addition, we used simple categorical analysis methods to compute chi-square tests for possible differences among groups of respondents.

Sampling Weights: Nonresponse

Prior to analysis, each respondent was assigned a weight that reflected the number of physicians in the population whom he or she represented. First, sampling weights were computed as:

$$swt = 1/p$$

where *swt* is the sampling weight and p is the respondent's probability of selection. Next, the sampling weights were adjusted for nonresponse using the method of sample weight adjustment classes (107,177). In each class of respondents (as determined by the sampling criteria, described earlier), we reweighted each respondent to represent the number of physicians sampled in that class. Thus, the adjusted sampling weight became:

$$adjswt = swt * (1/p_r)$$

where p_r is the probability of response. The weighting classes were created to lump similar groups of physicians together and to ensure that the adjustment factor ($1/p_r$) was not unstable owing to small class size. Finally, we adjusted all weights so that the sum of the weights across respondents exactly equaled the number of physicians in the population. This adjustment represented a change of no more than about 0.5 percent.

Point Estimates and Confidence Intervals

Point estimates and confidence intervals were computed using the PROC DESCRIPT procedure in SUDAAN (193) where, as was commonly the case, the numbers of respondents in most sampling strata were large enough to take advantage of the stratified sampling design. Where the number of respondents in either the numerator or denominator of a rate calculation was small (fewer than 10 in the numerator or fewer than 40 in the denominator), we calculated exact binomial confidence intervals according to the method of Daly (50). This method avoided the well-known problem of having confidence intervals that are both too narrow and too symmetric.

Group Comparisons

For comparisons between groups we used the DIFFVAR option in the PROC DESCRIPT procedure in SUDAAN (193) to compute differences in rates (or proportions) and the variances of those differences. For small-sample comparisons (fewer than 10 respondents in a category), where stratified sampling adjustments were inappropriate, we used exact methods as implemented in StatXact-Turbo (49) and computed odds ratios rather than

rate differences.6 This approach allowed us to take advantage of the stratified sampling design, where the numbers of respondents were sufficient, and alternative methods where the numbers of respondents were too small to justify large-sample techniques. Tests for rate differences and odds ratios are comparable for these data.

Case-Control Comparisons

Comparisons of responses to the case and control scenarios presented special problems. First, the design of the surveys did not permit "within-physician" comparison of case and control responses, because the same respondents could not be given both the case and control scenarios without possibly revealing our purpose. The case and control responses were thus independent, thereby reducing the efficiency of the case-control comparisons (greater variances for the same sample size). Second, although the case and control groups were each stratified random samples, they could differ in systematic ways—most importantly, in their propensity to cite "malpractice concerns. " As a proxy for this control variable, we examined whether or not the respondent double-checked "malpractice concerns" for one or more procedures in the common scenario for each survey (the scenario received by every respondent in a given survey—see table D-l). This adjustment was computed as follows.

Where the numbers of respondents were adequate (again, at least 10 in each category), we used sample-weighted logistic regression, as implemented in the PROC LOGISTIC procedure in SUDAAN (193), to perform the equivalent of stratified 2-by-2 contingency table analysis in which:

- the dependent variable was whether or not the respondent double-checked '*malpractice concerns" in the case-control scenario (labeled *response* in the model shown below);
- the independent variable was the respondent's group (case or control, labeled *group* in the model); and
- the control variable was whether or not the respondent double-checked "malpractice concerns" in the common scenario (labeled *common* in the model).

The saturated model for this analysis then became:

$$\text{response} = \beta_0 + \beta_g * \text{group} + \beta_c * \text{common} + \beta_{int} * (\text{group} * \text{common})$$

where *response* is the log odds of double-checking "malpractice concerns," and the ß's represent regression coefficients.

Using an interaction term representing the joint effects of *group* and *common* permitted us to test whether the impact of the respondent's group (case or control) on his or her defensive-medicine response in the case-control scenario differed according to his or her defensive-medicine response in the common scenario. If the interaction term was not statistically significant, then the model simplified to the two main effects (group and common), and the odds ratio of the case and control responses became $\exp(\beta_g)$.

Where the numbers of respondents were small (again, usually fewer than 10), we used exact analysis of these stratified 2-by-2 contingency tables, as implemented in StatXact-Turbo (49). Here we computed exact common odds ratios (case versus control) and their 95-percent confidence intervals and p values, as well as the exact test for the homogeneity of odds ratios across the categories of the control variable (*common*).

Global Differences

Global tests for the significance of difference across the categories of the demographic variables (e.g., practice setting) in the rate of double-checking of "malpractice concerns" in the common scenario for each survey were initially assessed using

---

6 Except where noted, the calculations are evict odds ratios and their accompanying exact 95-percent confidence intervals and p-values, computed according to the methods of Mehta, Gray, and Patel (156).

the PROC FREQ procedure and Cochran-Mantel - Haenszel statistics on the normalized weighted data in SAS (203) (see table D- 1).[7] The DIFFVAR option in PROC DESCRIPT in SUDAAN (193) was used to test the significance of difference in mean attitude scale scores between respondents who double-checked "malpractice concerns" in the common scenario for each survey (see table D-1) and those who did not.

[7] The common scenarios were used for this analysis so that it would be based on all respondents in a given survey.