SUPPORTING STATEMENT

Part B

National Study of the Hospital Adverse Event Reporting System

Version: July 8, 2008

Agency of Healthcare Research and Quality (AHRQ)

Table of contents

B. Collections of Information Employing Statistical Methods

1. Respondent universe and sampling methods

A random sample, using random selection within strata, of 1,200 will be selected from the 1,652 hospital Risk Managers who completed the survey in our baseline study completed in 2005. We expect a response rate of 85%, based on an 81% response rate for the baseline survey, as well as an expectation that a high percentage of those who already completed the survey will respond to the follow-up survey. This results in an estimated sample size of 1020 for the follow-up survey.

The Sample Frame and Methodology

Unit of analysis = the hospital

Statistical power to obtain results by:

- hospital bed size
- teaching versus non-teaching
- ownership (for-profit, non-profit, public)
- urban versus rural (including frontier counties)
- in a multi-hospital system or not

The sample design for the baseline survey was a stratified random sample of hospitals using a combination of the strata dimensions listed above. To achieve a representative sample in this survey across strata used in the baseline survey, different proportions of responders to the baseline survey will be selected. Specifically, we will select higher proportions of respondents in baseline survey low response rate strata (e.g. 50% rate) and smaller proportions from baseline survey high response rate strata (e.g., 100% rate).

If the follow-up survey non-response rates vary considerably by stratum, two additional sampling steps may be taken to obtain proportionate samples within each stratum:

- In follow-up survey low response strata, additional randomly selected hospitals that responded to the first survey may be selected for this survey.

- If this is insufficient to obtain enough responses in a follow-up survey low responding stratum, additional hospitals from that stratum that either were not selected to be in the first survey or who were selected and did not respond to the first survey may be randomly sampled for this survey. This would serve to 'refresh' the sample to obtain a cross sectionally representative sample at the second time point. Because these hospitals would be new to the study, they could not contribute to any analyses of change over time.

Assumptions and Goals

1. The sample design for the initial survey was a stratified random sample of hospitals using a combination of the strata dimensions listed above.
2. The Risk Manager in each hospital in the sample is the survey respondent. This approach is taken to capture information on hospital-wide, comprehensive reporting systems.
3. The followup national survey will be designed to have the statistical power to:
   - Estimate differences in the use and comprehensiveness of adverse event reporting systems by hospitals of differing characteristics and locations, and
   - Estimate changes over time in the comprehensiveness of adverse event reporting systems by hospitals of differing characteristics and locations.
4. The national survey sample will consist of hospitals drawn from those that responded to the first survey in 2005, which were drawn from the sampling frame of the 2003 AHA hospital survey database.

Data on hospital distributions. Using the 2003 AHA hospital survey data, we obtained distributions of hospitals by each of the characteristics of interest. In the 2000 database (used to calculate power), about 5,795 non-federal hospitals in the AHA guide were in the sampling frame (including not-for-profit hospitals, for-profit hospitals, and hospitals operated by cities, counties, or states). The smallest cell size among the five dimensions of hospital ownership is for-profit hospitals, which are 19% of the total number of hospitals. There were 2,323 rural hospitals (defined as those located outside of Metropolitan Statistical Areas), which represent 40% of the total 5,795 hospitals in the

sampling frame. An estimated 26 percent of the hospitals had teaching programs, and 48% had fewer than 100 beds. We ran correlations among the hospital characteristics, finding high correlations between rural and teaching (-0.34); rural and bed size (-0.42) and therefore necessarily teaching and bed size (0.41). Rural hospitals tend to be relatively more government owned and non-rural ones tend to be relatively more for-profit ownership. Government hospitals were less likely than non-profit and for-profit hospitals to be teaching hospitals. The distribution of governmental hospitals by bed size differs from distributions for non-profit or for-profit hospitals. Government hospitals are least likely to be in multi-hospital system than the non-profit or for-profit hospitals.

Power calculations.

Our sample will include large, medium and small size, non-federal hospitals (including not-for-profit hospitals, for-profit hospitals, and hospitals operated by cities, counties, or states). It will include hospitals in states where adverse event reporting is "mandatory" as well as where it is not.

With this relatively balanced distribution of hospitals, we needed a total of about 1,200 randomly sampled hospitals to have sufficient power for reasonable hypotheses, with the goal of achieving 1,020 completed surveys (85% response rate). This sample size would allow us sufficient power to detect the following differences between subgroups for cross-sectional analysis (within the second survey) and for before-after analysis (between the 2005 survey and this survey):

1. Power for cross-sectional analysis – within the Second AERS Only
    - Sufficient power (80%) to detect a difference of 50% versus 65% (or 50% versus 35%) in a dichotomous outcome (e.g., whether or not a hospital has a comprehensive adverse event reporting system) between two sub-samples that are each 20% of the full sample. For a dichotomous outcome, the most difficult differences to detect are in the vicinity of 50%; therefore this 15% difference is a worst-case scenario.
    - Margin of Error (1/2 confidence interval) for a single full sample proportion = 0.03 with $\alpha = 0.05$

- Margin of Error (1/2 confidence interval) for a single full sample mean of a continuous measure = 0.056 standard errors with $\alpha = 0.05$
- Can distinguish an effect size of 0.28 (a difference in means of 0.28 standard deviations) between two 20% sub-samples for a continuous outcome with $\alpha = 0.05$ and 80% power.

2. Power for before-after analysis – using linked First and Second Surveys
   - For the full sample, the average change over time (within hospital) in a continuous outcome that can be detected with $\alpha = 0.05$ and 80% power is an effect size of 0.09 (0.09 standard deviation change where this is the standard deviation of the changes, not of the original factor – if the pre-post correlation = 0.30 this is an effect size of 0.11, corr = .5 → ES= 0.09, corr=0.70 → ES= 0.07 where the effect sizes are in terms of number of standard deviations of the outcome instead of change in the outcome).
   - For a 30% sub-sample, the average change over time (within hospital) in a continuous outcome that can be detected with $\alpha = 0.05$ and 80% power is an effect size of 0.16 (0.16 standard deviation change where this is the standard deviation of the changes, not of the original factor – if the pre-post correlation = 0.30 this is an effect size of 0.19, corr = .5 → ES= 0.16, corr=0.70 → ES= 0.12 where the effect sizes are in terms of number of standard deviations of the outcome instead of change in the outcome).

Note: These power calculations are conservative if:

- the analysis accounts for the stratified design and the variation is lower within strata than between strata, or
- other covariates that are correlated with the outcomes are controlled for in testing for differences over time or between groups.

2. Information Collection Procedures

The survey mode for the follow-up survey will be a mail survey with two waves of mail follow-ups, and finishing with a CATI telephone survey follow-up for the remaining non-responders. The CATI survey will be tested to ensure that the questionnaire items appear

as designed, that the logical flow is correct, that there are appropriate range checks and that the data are being recorded correctly. The survey questions were revised based on the pilot survey, and takes approximately 25 minutes to complete. The 1,020 follow-up surveys will be completed with the Risk Manager at each hospital (one per hospital).

The steps in the process are as follows:

1. A cover letter and copy of the follow-up survey will be mailed to the Risk Manager.

2. A reminder post card will be sent to the Risk Managers who have not returned the follow-up survey within 2 weeks of the initial mailing, and a re-mail of the follow-up survey will be sent 2 weeks after the reminder post card is sent.

3. If a follow-up survey has not been returned 2 weeks after the second re-mail, a telephone interviewer will attempt to complete the follow-up survey with the Risk Manager over the telephone.

The survey and survey procedures have been cognitively tested and piloted. They also have been used successfully in the baseline survey, yielding an 81 percent response rate.

3. Methods to Maximize Response Rates

The survey will first be mailed twice, with a reminder letter between mailings, and if there is no response, a telephone interview will be conducted using Computer-Assisted Telephone Interviewing. Respondents will be called at different times of days and different days of the week, and messages will be left on voice mail or with a gatekeeper. The survey methodology includes telephone follow-up with the mail non-respondents to maximize response rates. The methods proposed for data collection should yield fairly high response rates.

For hospital nonresponse, we will construct nonresponse weights. We are fortunate that we have information on both responding and nonresponding hospitals via the AHA database. Initially we will assess how different responding and nonresponding hospitals

are in terms of AHA dimensions, an analysis which will be informative to AHRQ in its own right. We will then fit a multivariate logistic regression model using AHA variables as covariates and respond or did not respond as the outcome. Based on this model, we will form nonresponse classes that consist of responding and nonresponding hospitals who are similar in terms of predicted nonresponse. Responding hospitals in a particular nonresponse class will all receive the same nonresponse weight, which will be calculated based on the number of nonresponding hospitals that need to belong to that class, and thus need to be represented. All analyses of the survey will utilize these nonresponse weights to ensure that our results are generalizable to the target population of hospitals.

For item nonresponse on the surveys, we will apply imputation methods as appropriate. For items that fall above? a certain level of response, generally about 60-70% depending upon the item, we will report the observed (weighted for unit-level nonresponse as described above) results and make no attempt at imputation. Indeed nonresponse is itself an informative category for some survey questions. For items that have lower item nonresponse and for which imputation is appropriate, we will impute using methods such as hot-deck imputation that incorporate uncertainty into the imputed responses. We will consider multiple imputation but do not anticipate the need for this advanced technique at this time. We note that for both unit and item level nonresponse, we will balance analytic sophistication with the interpretability and acceptability of the approach. Transparency in our methodology will be a primary goal of the analytic phase of the project.

4. Tests of Procedures

The majority of the questions and procedures in the baseline and follow-up survey have been cognitive tested during the pilot study and have been revised based on those recommendations. We will be employing essentially the same procedures and survey during the follow-up study. Further, the study materials (in attachments) and procedures have been reviewed and approved by RAND's Human Subjects Protection Committee.

5. Statistical Consultants

Dr. Amelia Haviland at RAND (412) 683-2300 was consulted on the statistical aspects of this survey design.