

SUPPORTING STATEMENT

Part B

VALIDATION PILOT FOR THE AHRQ PATIENT SAFETY INDICATORS PHASE II

Version June 19, 2008

Agency for Healthcare Research and Quality (AHRQ)

Table of contents

| | |
|--|---|
| B. Collections of Information Employing Statistical Methods..... | 3 |
| 1. Respondent universe and sampling methods..... | |
| 2. Information Collection Procedures..... | |
| 3. Methods to Maximize Response Rates..... | |
| 4. Tests of Procedures..... | |
| 5. Statistical Consultants..... | |

B. Collections of Information Employing Statistical Methods

The primary purpose of the validation pilot is to test the feasibility of the data collection itself through a collaborative of voluntary hospital participants. As part of that feasibility assessment, the validation pilot will test the deployment of the medical record abstraction tools and protocols for data collection as applied in various hospital settings. In addition, the validation pilot will test sampling methods intended to be sufficient to allow for the accomplishment of the validation analysis objectives under the constraints of 1) the limited burden per hospital needed to ensure voluntary participation and 2) the relative infrequency of the Patient Safety Indicator events.

1. Respondent universe and sampling methods

The sampling and analysis methods are intended to address two issues. First, to test the feasibility of estimating the *positive predictive value* (PPV) of the adverse events flagged in the administrative data using the medical record data as the gold standard. In Phase II, we will test the feasibility of estimating PPV for the five Patient Safety Indicators that are less common and more clinical complex than the indicators used in Phase I. Second, to test the feasibility of estimating the *sensitivity* of the adverse events flagged in the administrative data using the medical record data as the gold standard. In Phase II, we will test the feasibility of estimating sensitivity for all ten Patient Safety Indicators.

Participating hospitals will collect 30 cases to estimate PPV and an additional 30 cases to estimate sensitivity. We expect to recruit 40 non-randomly selected hospitals to participate in the pilot. Hospitals will be recruited through an email to hospitals that participated in Phase I and through an announcement on the AHRQ QI website and distributed through the AHRQ QI listserv. Based on our experience in Phase I, we anticipate that the hospitals volunteering to participate in the pilot will include a higher percentage of large, not-for-profit hospitals than community hospitals nationally (Table 1), but that hospitals of various types will be represented.

Table 1. Characteristics of Phase I Validation Pilot and Community Hospitals Nationally

| Ownership | Phase I Pilot Hospitals | Community Hospitals, National | Bed Size | Phase I Pilot Hospitals | Community Hospitals, National |
|------------------------------|-------------------------|-------------------------------|--------------|-------------------------|-------------------------------|
| State | 7.1% | 1.4% | <50 beds | 0.0% | 7.9% |
| District | 7.1% | 11.4% | 50-99 beds | 3.5% | 20.0% |
| Other public | 0.0% | 11.4% | 100-199 beds | 17.8% | 19.1% |
| Not-for-profit, religious | 3.5% | 11.3% | 200-299 beds | 25.0% | 23.2% |
| Not-for-profit, nonreligious | 78.5% | 48.9% | 300-399 beds | 17.8% | 12.8% |
| For-profit | 3.5% | 15.6% | 400-499 beds | 7.1% | 7.6% |
| | | | 500+ beds | 28.5% | 3.6% |

Source: American Hospital Association Survey (N=4,648)

a. Estimating PPV

The validation pilot will test the feasibility of estimating the positive predictive value of the adverse events flagged in the administrative data using the medical record data as the gold standard. PPV is defined as the crude percentage of PSI-flagged cases that were confirmed by detailed medical record review. In general, we estimate that an actual sample size of approximately 225 cases can be expected to yield 95% confidence intervals for the PPV parameter of width less than 12 percentage points when the population PPV is 70% (the average sample PPV from Phase I)¹. However, for less common indicators our eventual sample sizes will be much smaller and the resulting estimates less precise. For the more clinically complex indicators, our eventual sample sizes will depend on the degree of clustering within hospitals and the resulting variance inflation factor (see below) and may need to be larger to achieve similar precision. In addition, we will want to test the association with specific processes of care and potential confounding factors, and to determine the required sample sizes to conduct these analyses.

The “respondent universe” will include a census of all of the cases flagged for a Patient Safety Indicator adverse event over a three year period. Patient Safety Indicator adverse events are relatively rare, occurring in less than 1% of all hospitalizations. Therefore, the

¹ This is based on the classical variance estimator $p*q/n$ for the sample proportion p where n is the number of independent observations from a Bernoulli process.

intent is to sample 100% or close to 100% of the medical record data flagged in the administrative data. In the 40 participating hospitals, we estimate that there would be approximately 1,200 flagged PSI adverse events for the five Patient Safety Indicators included in this Phase II pilot (Table 2). The estimated average number of flagged cases per hospital over three years is calculated from the AHRQ Healthcare Cost and Utilization Program (HCUP) State Inpatient Data (SID) for 2002-2004.

Table 2. Estimated Number of Flagged Cases in Phase II Validation Pilot Hospitals

| PSI | Label | Estimated Average Number Of Flagged Cases Per Hospital Over Three Years | Estimated Total Number Of Flagged Cases In 40 Participating Hospitals Over Three Years |
|-----|--|---|--|
| 05 | Foreign Body left in During Procedure | 1.1 | 44 |
| 09 | Postoperative Hemorrhage or Hematoma | 8.7 | 348 |
| 10 | Postoperative Physiologic or Metabolic Derangement | 2.6 | 106 |
| 11 | Postoperative Respiratory Failure | 17.1 | 684 |
| 14 | Postoperative Wound Dehiscence | 2.2 | 86 |
| | Total | 31.7 | 1268 |

Source: HCUP State Inpatient Data (SID), 2002-2004

However, under the constraint that each hospital will have no more than 30 cases and as close to 30 cases as possible selected into the sample, and given that the actual number of cases per PSI will vary by hospital, the sampling fractions must be adjusted for each hospital, with priority given to those cases with adverse events from the lower-yield PSI: Foreign Body left in During Procedure, Postoperative Physiologic or Metabolic Derangement and Postoperative Wound Dehiscence. The higher-yield PSI are Postoperative Hemorrhage or Hematoma and Postoperative Respiratory Failure. Although we expect that virtually all participating hospitals will have fewer than 30 total cases pertaining to the lower-yield PSI included in the pilot, our sampling procedure will allow for hospital-specific adjustments to the within-hospital indicator-specific medical record sampling fractions so that the burden on each hospital is limited to no more than 30 total sampled cases. The sampling procedures for each hospital will be:

- Apply the AHRQ QI Windows Software to three years of hospital administrative data
- Identify the number of cases flagged for each of the five PSI included in Phase II of the validation pilot
- Apply the initial sampling fraction (100%) to randomly select the cases for medical record abstraction using a random number generator
- If the number of cases with the 100% sampling fraction is less than 30, then select all of these cases.
- If the number of selected cases is greater than 30, then reduce the sampling fraction for the higher-yield PSI by $((30 - \text{lower yield PSI}) / \text{higher yield PSI})^*$
- Re-apply the adjusted sampling fraction to select the cases
- If the number of selected cases is still greater than 30, then reduce the sampling fraction for all PSI by $(30 / \text{selected cases})^*$
- Re-apply the adjusted sampling fraction to select the 30 cases
- If the number of selected cases is less than 30, then increase the sampling fraction for the higher-yield PSI by $((30 - \text{lower yield PSI}) / \text{higher year PSI})^*$

* For each hospital and PSI, the minimum sampling fraction is always zero and the maximum is always one.

Our measure of PPV will be a point estimate of the proportion of flagged cases for which the medical record review validates the flagging. Although the cases within each hospital are a random sample (using the hospital-specific sampling probabilities described above) the confidence interval for the PPV point estimate from our convenience sample of hospitals should be adjusted to account for the clustering of cases within hospitals.. In other words, to the extent that there are between-hospital differences in true PPV probabilities, the effective sample size of a cluster sample is reduced (i.e., will achieve the same precision as a simple random sample with a smaller sample of cases). The effective sample size in a cluster sample depends on the ICC (intra-class correlation), which is ratio of the between-hospital variance component for the outcome and the sum of the between-hospital and within-hospital variance components. Based on our Phase I data, we assume an ICC of 1%, which results in a Variance Inflation Factor (VIF) of 1.11. A sample size of $n \geq 225 \times 1.11 = 250$ would be required for the width of the 95% asymptotic confidence intervals of 12 percentage points on an estimated PPV of 70 percent. Therefore, we expect that a cluster sample of 250 will provide us the reasonably narrow confidence intervals that we desire, at least for the higher-yield PSIs.

b. Estimating Sensitivity

The validation pilot will also test the feasibility of estimating the sensitivity of the PSI. Sensitivity is defined as the proportion of adverse events documented in the medical record data (considered as the gold standard) that are identified by the PSI flagged cases from the administrative data. Higher sensitivity results in a relatively fewer number of these “false negative” cases.

The challenge in estimating sensitivity is to sample medical records efficiently. The least efficient approach would be to randomly select 30 cases from each hospital because the likelihood that a randomly selected medical record would have an adverse event is very low. Therefore, we will use an unequal probability stratified sample design to select hospitalizations whose medical records will be reviewed, where sampling fractions will be higher in strata judged to be more likely to have hospitalizations with adverse events.

Strata will be defined based on auxiliary information (from within the administrative record for the hospitalization) and will be in three tiers:

- The first tier will be based on indicator specific “markers” judged to be associated with an increased likelihood of the medical record showing an adverse event. For example, each indicator definition excludes cases from the denominator when the adverse event is more likely to be present-on-admission or less likely to be preventable. However, these excluded cases are also more likely to have a true un-flagged adverse event. We will use subject matter expertise to formulate classification rules to identify which administrative records should be placed in strata in this top tier, where the sampling fractions will be highest.
- The second tier of strata will be based on propensity score estimates from a multifactorial regression model that relates administrative data fields (e.g. gender, age, Diagnosis Related Group and coded co-morbidities) to the probability that the record would have at least one of the Patient Safety Indicators. Noting that the PSI definitions are themselves based on the administrative data, we will take due

care to avoid a tautological specification for this multifactorial model. Based on the regression model estimates, each record will be placed in a propensity class stratum. Sampling fractions will be highest in the highest propensity classes, although all records will have a positive probability of being selected into the sample.

- The third tier of strata will be a random sample of medical records. A weighted average of the three tiers will yield the overall sensitivity estimates for each indicator.

We intend to select (approximately) 30 hospitalizations from each hospital for purposes of assessing sensitivity. The medical record data for each sampled hospitalization will be assessed for adverse events pertaining to each PSI. With 40 hospitals participating, we estimate there will be approximately 1,200 selected medical records assessed for “denominator” cases for purposes of estimating the sensitivity of each indicator. It should be carefully noted that although our sampling strategy will make judicious use of the administrative data, it can not ensure that enough denominator cases will be yielded in our sample for any particular indicator to guarantee that the confidence intervals for sensitivity will be comparable in precision to those for PPV.

Table 3. Sampling Strata for Sensitivity Analysis

| Type of Chart | Sampling Probability | Analysis |
|----------------|----------------------|-------------|
| Clinical Event | Census | PPV |
| Marker Event | Probability | Sensitivity |
| Model | Propensity | Sensitivity |
| Other | Random | Sensitivity |

c. Estimating Reliability

The final statistical analysis will assess whether the medical record abstraction, sampling and data collection tools are reliable, adequately documented, compliant with applicable standards (i.e. Section 508) and otherwise suitable for public release. Reliability is the

extent to which the tools provide consistent results across institutions and abstractors with equivalent skills and methods. Reliability testing will require that the hospitals re-abstract approximately 10% of records for selected items for constructing metrics of inter-rater reliability. It will also require the team to modify the training process to ensure uniformity, for example by including mock chart training using one or two sample charts. In the validation pilot, the focus will be the consistency of the abstraction among hospital staff from the same hospital who are similarly trained and equally familiar with the medical record documentation practices in that hospital. The Pearson correlation coefficient will be used as our measure of reliability based on a few selected items from the medical record data abstraction instruments. We will also discuss any disagreements with hospital staff to determine whether the items or documentation might be improved to increase uniformity. Finally, we will evaluate the intra-class correlations discussed above to determine whether there are any systematic differences among hospitals, and to identify hospitals for further discussions in order to identify potential sources for such differences. We will use this information to assess whether the full OMB clearance package will include a mechanism for external review and the nature of the review (e.g. whether the support team would re-abstract selected medical records as a “gold standard” and to compute “error” rates for selected items)².

2. Information Collection Procedures

In order to identify the cases for medical record data collection, participating hospitals will use the existing AHRQ QI Windows software to apply the sampling approach outlined above. The Windows software and documentation are available for download from the AHRQ QI website at <http://qualityindicators.ahrq.gov>. The software was originally released in September, 2005 for use by hospitals and hospital systems to apply the AHRQ QI specifications to hospital administrative data. Use of the existing Windows software will significantly reduce the burden on participating hospitals as most have already used the software and have already imported batch extracts of the

² We considered such a mechanism of external review to be too burdensome on our volunteer hospitals in the validation pilot, both because of the time and resources required to make a paper record from electronic medical records and to work through any potential modifications to the Institutional Review Board or Privacy Board process established in Phase I.

administrative data into the database created by the software. The only additional step is to apply the sampling approach to that existing data, which will generate a list of the 60 cases selected for data collection for each hospital.

Once that list is generated, the hospital will either pull the paper medical record for each case or access the medical record via the Electronic Medical Record EMR. Although the AHRQ QI support team will use a web-based version of the data collection tool for data entry and data extraction, the Phase I hospitals indicated a preference for filling out a paper version of the data collection tool because it facilitated access to the EMR. The hospitals will have three months to complete the data collection.

3. Methods to Maximize Response Rates

The web-based application includes the capability of generating meta-data reports that will indicate the number of cases entered for each participating hospital. As participating hospitals submit the paper tools to the AHRQ QI support team for data entry, we will track this data entry on a weekly basis to determine whether there are any hospitals that have not yet completed a reasonable portion of the assigned cases. Hospitals that fall behind will be contacted to determine the reason. If the reason is lack of available staff to conduct the medical record abstraction, we may be able to provide limited staff support to assist in the review. If the reason is lack of data entry personnel, we may be able to provide limited staff support to assist in the data entry. Both of these methods will depend on available resources.

4. Tests of Procedures

We will be testing all of our tools and protocols using members of the support team at the University of California-Davis Medical Center. Staff at UC-Davis will apply the software to identify the sampled cases for data collection, work through the abstraction instruments for these cases in a medical record review, and enter the data into the web-based data collection application. Staff conducting this testing will be generally

unfamiliar with the details of AHRQ PSI project, in order to simulate the level of expertise of a typical user of the PSI.

In addition, we will conduct webinars to train hospital staff on the data collection instruments and on the administrative data. These trainings will review the rationale behind the instruments and the indicators, and provide guidance to hospital staff on how to locate the required information from the medical chart, and how to apply the AHRQ QI Windows software to the administrative data and how to interpret the results.

5. Statistical Consultants

Battelle Memorial Institute, in Arlington, Virginia has been contracted to conduct this pilot data collection. Battelle has subcontracted to the University of California-Davis and Stanford University to assist in this effort. The individuals assigned to the project and their title, roles and contact information are included below.

| Battelle Memorial Institute | |
|--|---|
| Jeffrey Geppert Senior Research Scientist Centers for Public Health Research and Evaluation (CPHRE) Battelle Memorial Institute 2101 Wilson Boulevard, Suite 800 Arlington, VA 22201-3008 Phone: 916-682-9965 Fax: 614-458-6698 geppertj@battelle.org | Theresa Schaaf Project Manager Battelle Arlington Operations 2101 Wilson Boulevard, Suite 800 Arlington, VA 22201-3008 Phone: 703-875-2990 Fax: 703-527-5640 SchaafT@battelle.org |
| Jennifer Cohen Principal Research Scientist Battelle Memorial Institute (CPHRE) 1100 Dexter Ave North, Suite 400 Seattle, WA 98109-3598 Phone: 206-528-3116 Fax: 614-458-6743 CohenJ@battelle.org | Sharon Xiong Research Associate Battelle Memorial Institute (CPHRE) 2101 Wilson Boulevard, Suite 800 Arlington, VA 22201-3008 Phone: 703-875-2987 Fax: 703-527-5640 XiongX@battelle.org |
| UC-Davis | |
| Patrick Romano, M.D. Professor of Medicine and Pediatrics | Ruth Baron, R.N. Nurse Researcher |

| | |
|--|--|
| <p>UC Davis Division of General Medicine 4150 V Street; PSSB Suite 2400 Sacramento, CA 95817 Phone: 916-734-7237 Fax: 916-734-2732 psromano@ucdavis.edu</p> | <p>UC Davis Center for Health Services Research in Primary Care 2103 Stockton Blvd. Ste. 2224 Sacramento, Ca 95817 Phone: 916-734-7878 Fax: 916-734-2349 ruth.baron@ucdmc.ucdavis.edu</p> |
| <p>Patricia Zrelak, R.N. Nurse Researcher UC Davis Center for Health Services Research in Primary Care 2103 Stockton Blvd. Ste. 2224 Sacramento, Ca 95817 Phone: 916-734-7878 Fax: 916-734-2349 Patricia.Zrelak@ucdmc.ucdavis.edu</p> | <p>Garth Utter, M.D. Assistant Professor of Surgery Division of Trauma & Emergency Surgery 2315 Stockton Blvd., Rm. 4206 Sacramento, CA 95817 Phone: 916-734-1768 Fax: 916-734-7755 garth.utter@ucdmc.ucdavis.edu</p> |
| <p>Banafsheh Sadeghi, M.D. Candidate to PhD in Epidemiology UC Davis Division of General Medicine 4150 V Street; PSSB Suite 2400 Sacramento, CA 95817 Phone: 510-918-7669 Fax: 916-734-2732 bsadeghi@ucdavis.edu</p> | <p>Daniel Tancredi Senior Statistician UC Davis Center for Health Services Research in Primary Care 2103 Stockton Blvd, Suite 2224 Sacramento, CA 95817 Phone: 916-734-3293 Fax: 916-734-2349 djtancredi@ucdavis.edu</p> |
| <p>Stanford University</p> | |
| <p>Sheryl Davies Project Manager Centers for Health Policy and Primary Care and Outcomes Research (PCOR) 117 Encina Commons Stanford University Stanford, CA 94305-6019 Phone: 650-498-9023 Fax: 650-723-1919 smdavies@stanford.edu</p> | <p>Kathryn McDonald Senior Scholar & Executive Director Centers for Health Policy and Primary Care and Outcomes Research 117 Encina Commons Stanford University Stanford, CA 94305-6019 Phone: 650-723-0559 Fax: 650-723-1919 kathy.mcdonald@stanford.edu</p> |
| <p>Olga Saynina Research Associate PCOR/NBER 30 Alta Road Stanford, CA 94305-6019 Phone: 650- 326-1958 Fax: 650-328-4163 olga@newage3.stanford.edu</p> | |

No other persons will consult on the statistical aspects of the pilot nor will any other persons either collect or analyze the data.