Supporting Statement - Part B

Collections of Information Employing Statistical Methods

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

RESPONSE:

The Physician Quality Reporting System (PQRS) and Electronic Prescribing Incentive (eRx) Program Data Assessment, Accuracy and Incorrect Payments Identification Support contract was created to identify and address problems with data handling, data accuracy, and incorrect payments for the PQRS and eRx Programs.

Because the data submitted by, or on behalf of, eligible professionals (EPs) to the PQRS and eRx Programs is used to calculate incentive payments and payment adjustments, it is critical that this data is accurate. Additionally, the data is used to generate Feedback Reports for EPs and, in some cases, is posted publicly on the CMS website, further supporting the need for accurate and complete data.

The ultimate use of the clinical quality reporting data is to improve the quality of care for Medicare beneficiaries. This aligns with the CMS mission and helps to make healthcare more cost-effective and efficient.

To determine if data quality issues exist and if the incentive payments are correct, additional information is required. Surveys are one tool that will be used to collect this data, and they will be sent to the following reporting entities: Group Practices using the Group Practice Reporting Option (GPROs), Registries, Qualified Clinical Data Registries (QCDRs), Electronic Health Record (EHR) Data Submission Vendors (DSVs), and Eligible Professionals (EPs) submitting via the Electronic Health Record (EHR) Direct and Claims reporting options.

The survey is completely automated and was designed with simplicity as a core requirement – it does not require a login and can be accessed via a link provided in a survey invitation email. There is no Protected Health Information (PHI) or Personally Identifiable Information (PII) submitted in the survey. In order to minimize the burden on the participant community, the number of questions in a survey will not exceed 30. The majority of the questions in the survey are "point and click," allowing the participant to complete the survey quickly. There is a Feedback section included in the survey, which allows for free-form text entry and document upload; however, document uploads are not required.

Sampling, as it relates to this effort, will limit the number of entities that receive the survey and, consequently, the data examined to identify errors and incorrect payments made from the Physicians Quality Reporting System (PQRS) data. The projected samples by contract option year are provided in Table 1: Sampling Size Distribution, below. The samples are generated following the analytical process described in the next sections. For Option Years 2 and 3, the total number of participants is 115; however, the exact number by reporting option will be determined after the analytical process is completed. Table 1: Sampling Size Distribution, below, shows how the sampling is distributed among reporting entities.

Table 1: Sampling Size Distribution

Reporting Entity	Base Year	Option Year 1	Option Year 2	Option Year 3
GPRO	NA	20	TBD	TBD
Registries	9	30	TBD	TBD
EHR Direct	NA	30	TBD	TBD
EHR DSV	NA	26	TBD	TBD
EPs submitting via Claims	NA	9	TBD	TBD
Total	55	115	115	115

2. Describe the procedures for the collection of information, including:

- Statistical methodology for stratification and sample selection,

- Estimation procedure,

- Degree of accuracy needed for the purpose described in the justification,

- Unusual problems requiring specialized sampling procedures, and

- Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

RESPONSE:

The methodology begins with the Measures Analytics Phase. During this phase, the Team conducts detailed literature reviews, environmental scans, and analyses, such as comparison of results against national benchmarks and year-to-year comparisons, to assess the validity of submitted PQRS results at various levels of granularity. Where feasible, the analyses performed during the Measure Indicator Analysis phase are completed for each of the applicable reporting options. The outcomes of these analyses define the business rules for measure submission and validation that are incorporated into rule sets during the Data Analytics phase. Where feasible, the Team re-uses the rules developed in previous option years by applying required updates or changes.

During the Data Analytics Phase, the Team applies the rules to the data, via the Cross Industry Standard Process for Data Mining (CRISP-DM) process to derive the samples for the Electronic Survey (Survey). The CRISP-DM process, which is the most widely adopted data mining life cycle model, uses discrete phases to increase data understanding, and in turn, increase the precision of the analytics that are applied during each subsequent phase. By following the CRISP-DM life cycle model, the Team ensures that a robust process is used to look at the data from many angles to detect and report data inaccuracies. During this process, the entities and EPs are "scored" based on how many rules they did or did not violate.

The phases in the CRISP-DM process are shown in Table 2: CRISP-DM Process, below.

Table 2: CRISP-DM Process

#	CRISP-DM Phases	Description
1	Business Understanding	This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.
2	Data Understanding	The data understanding phase starts with an initial data collection and proceeds with activities to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.
3	Data Preparation	The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tools) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include selecting tables, records, and attributes, as well as transforming and cleaning data for modeling tools.
4	Modeling	In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements based on the form of the data. Therefore, stepping back to the data preparation phase is often needed.
5	Evaluation	At this stage in the project, models are built that appear to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to evaluate the model more thoroughly and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.
6	Deployment	Creating the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained must be organized and presented in a way that the customer can use. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. However, even if the analyst will not carry out the deployment effort, it is important for the customer to understand up front the actions which will need to be carried out to make use of the created models.

During the modeling phase, the business rules are translated into SAS code and the code is applied to the data mart. At the highest level, the data modeling process is iterative in nature, as shown in Figure 1: CRISP-DM Process, below.
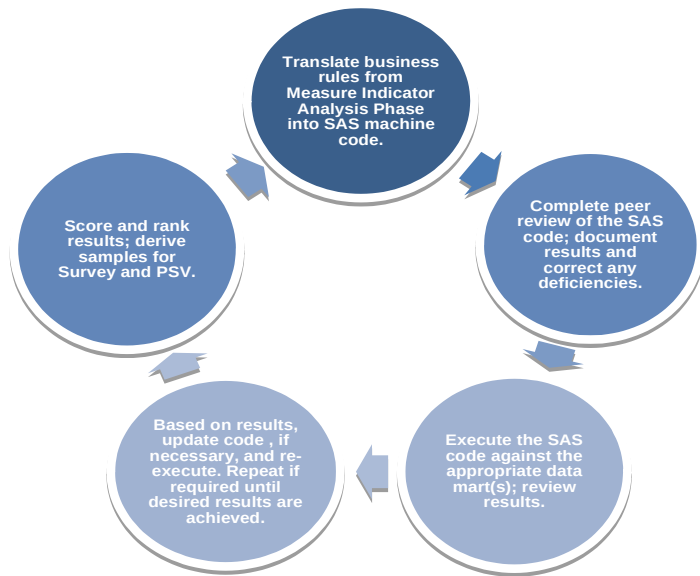
Figure 1: CRISP-DM Process

As stated above, when the rules are executed against the data mart, the entities are scored based on how many rules they did or did not violate. The results are then ranked and the samples derived for Survey. The intent of this process is two-fold: to identify sources of reporting error and develop recommendations for corrective action(s) and to identify candidates for payment recoupment.

The final step in the data modeling process is to use the experience gained to further refine the rules with the expectation that it will help the Team better identify reporting error and incorrect payments in future option years.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield 'reliable' data that can be generalized to the universe studied.

RESPONSE:

Knowing that timely and appropriate communication encourages participation by the survey participants, the Team has developed meaningful communications that include the initial invitation email and frequent reminder notices. In addition, prior to implementing a new survey, the Team conducts live demonstrations of the survey functionality with the participant community to increase their comfort level with the survey and encourage successful participation.

Regarding non-response, the Team follows up with participants regularly to remind them that survey participation is required and to ensure that they are on track to complete the survey within the prescribed timeframes.

The PRA package contains the communications, referenced above (PQRS Data Validation Electronic Survey - Att- C-Recruitment-comms.doc) that are shared with the participant community.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the

main collection of information.

RESPONSE:

During the Base Year of the contract, the survey was administered to a small group of Registry participants. Experience gained during the administration of the survey and subsequent analysis of responses was used to refine and improve the survey for subsequent option years. Prior to the implementation of new surveys, testing is completed to ensure that the survey is functioning as designed.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

The names and telephone numbers of individuals consulted on the statistical aspects of the design are shown in Table 3: Contact Information, below.

Table 3: Contact Information

Contact Name	Telephone Number	Agency Name	Email Address
Mary Braman	202-955-3583	NCQA	braman@ncaa.org
Andrew Weller	703-626-6106	IBM	andrew.j.weller@us.ibm.com