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.

Allegheny Science & Technology Corporation (AST) will randomly sample 35 facilities (0.6% of all dialysis facilities), per contract and Quality Incentive Program (QIP) rule guidelines, for participation in the validation project. As a random sample, this should be a representative sample of all included facilities nationally. The sample pool will consist of Medicare-certified dialysis facilities that are required to submit administrative and clinical data into CROWNWeb in order to meet Section 494.108(h) of the 2008 updated Conditions for Coverage for ESRD Dialysis Facilities. The 35 facilities will be asked to submit records that will be validated for NHSN dialysis event elements. The patient sample size is limited to 10 patients per facility, as per contract and QIP rule guidelines. Team AST will sample 10 patients (or the maximum patients possible) from each selected facility for NHSN reviews. Historically, facility response rates have been solid. The response rate for the 2016 validation study was 100%; of the nine facilities selected for participation, all nine complied with our records request.

Sample Size Estimates

Using data from 2015, which was validated in our previous work, we estimate the numbers of facilities per three strata by the numbers of patients reported per year. Using the ESRD QIP rule guidelines of 10 records per facility and no more than 35 facilities sampled, Team AST calculated the estimated sample size from each of three strata of facilities. **Table 1** below provides the sample size estimates for the validation effort where 0.6% (n=35) of all dialysis facilities will be sampled. Projected confidence intervals for facility-specific error rates for the NHSN dialysis event validation can be found in **Table 2.**

Table 1: Sample Size Estimates

Number of Patients per Quarter	% Total Facilities	% Total Patients	Number of Facilities	Number of Records to Sample per Facility	Number of Records
0 to 25	14.51%	3.25%	5	10	50
26 to 85	54.48%	40.09%	19	10	190
>85	31.01%	56.66%	11	10	110

Number of Patients per Quarter	% Total Facilities	% Total Patients	Number of Facilities	Number of Records to Sample per Facility	Number of Records
Total	**100.00%**	**100.00%**	**35**		**350**

Table 2: Confidence Interval for NHSN Data

Agreement Rate	Sample Size	Confidence Interval
90%	0.6%	±9.9%
80%	0.6%	±13.3%
70%	0.6%	±15.2%
60%	0.6%	±16.2%

Some smaller facilities may have fewer than 10 patients treated for the period; in these cases, the facilities will submit records for all of the patients treated at the facility during the study period for validation. For NHSN reviewed facilities, the facilities will submit up to 10 patient records based upon an NHSN extract that will contain patients for which a dialysis event has been reported. If there are not 10 patients with dialysis events for a particular facility, Team AST will then pull a random sample from the general pool of patients at the facility. The random sample of patients will allow for an examination of possible underreporting of dialysis events in the NHSN system. **Table 3** depicts the methodology to be used when sampling for patients for NHSN reviews.

Table 3: Sampling Methodology for NHSN Reviews

Sampling Source	Sample to be Taken
NHSN Extract	Maximum number of patients available, up to 10
General Pool of Patients from Patients within CROWNWeb	Remainder of patients needed to equal 10 after NHSN patients are sampled

Sampling Time Frame

For this year's current NHSN validation, Team AST possesses CDC-approved NHSN system

access. The three dialysis event elements (dialysis event type, dialysis event date, and vascular access type) found in the medical records are manually compared with the dialysis event element data found in the NHSN system. The use of this manual comparison methodology was only possible due to the smaller facility sampling size (nine facilities) mandated by the PY 2018 QIP Rule. In accordance with the PY 2019 QIP Rule, the future NHSN validation will expand to 35 facilities sampled for validation. Further sampling size expansion can be anticipated in outlying years. This expansion in sampling size renders manual comparison of the two data sets as impractical due to the increased time, effort, and analysis costs required. Therefore, a different mechanism will be necessary for comparing the two data sets. One option will involve the CDC providing an NHSN system data extract to Team AST (spanning the two quarters worth of data specified by the PY 2019 QIP Rule) for use in comparing the two data sets. With this option, facilities will be sampled from the Facility/Patient data extract file from CROWNWeb (January —June 2018). Team AST will send the list of 35 sampled facilities to the CDC and the CDC will provide an extract of all patient dialysis events reported by the selected facilities during the validation time frame. A second option would be to provide the CDC with the data file Team AST compiles/develops from medical record abstractions. The CDC would conduct the comparison between the Team AST abstracted data and the NHSN data. The CDC would then return their findings (percent match, percent non-match, and percent missing) associated with the three data elements to Team AST for processing and incorporation into the final report.

Team AST intends on validating records for the first and second quarter of 2018. This time frame was selected after considering several factors. To ensure that the validation can be completed during the period of performance, Team AST considered the data reporting periods allowed to facilities to submit clinical data into CROWNWeb and NHSN. Facilities are given 60 days to enter data from the previous month for CROWNWeb clinical data and are given three months after quarter close to submit dialysis event data in NHSN (for Q2 the data submission deadline is September 30). These mandated reporting periods limit the time frame we can validate expeditiously, as we will not be able to obtain an extract until after the close of the data-reporting period. We also took into account the time it would take to request and receive the CROWNWeb and NHSN data extracts; we anticipate that it may take the CDC up to a month to provide the data extract needed for sampling. Another important consideration is the ESRD QIP rule that makes it mandatory for us to give facilities up to 60 days to submit records. Taking into consideration these factors as well as the need to ensure that our reviewers have adequate time to complete NHSN reviews and that there is adequate time to perform analysis and prepare reports, we decided on using the first two quarters of 2018 as the validation time frame. A breakdown of the mandated reported deadlines that were taken into consideration is displayed in **Table 4**.

Table 4: Mandated Reporting Deadlines

Submission Type	Mandated Reporting Deadlines
CROWNWeb Data Submission	60 days after month close (Q2 – August 31, 2018)
NHSN Data Submission	90 days after quarter close (Q2 – September 30, 2018)
Facility Record Submission Deadline - NHSN	60 days after request receipt per QIP rule

Assuming the CROWNWeb data team will need at least 10 days and the CDC data team will need at least one month to export and send the data, Team AST has estimated preliminary dates for availability of January, February, March, April, May, and June 2018 data. It is anticipated that the NHSN data will be received during the month of November. Due to the tight time frame for data abstraction, effective coordination and management, as well as adherence to established schedules, will be crucial to the success of the project. As noted below in response to question 3, our response rate for the NHSN Feasibility Study was high; of the nine facilities selected for participation, nine responded (100%).

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.**

Please see response to question 1 for statistical methodology for stratification and sample selection, including estimation procedure and degree of accuracy needed for the purpose of this

work. As noted below in response to question 4, there are no unusual problems requiring specialized sampling procedures as our previous experience on past CMS NHSN validation efforts have shown near universal compliance by the hospitals with medical record requests. The period for data collection cycles is expected to be no more frequently than annually.

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.

As part of our previous work to collect medical records for the 2016 NHSN Feasibility Study, upon which the NHSN project is based, facilities were contacted via certified letter in January 2017 and were asked to participate in the validation effort. The letter provided instructions on the types of records to be submitted, methods to submit records to Team AST, and identified patients selected for validation. Facilities that did not respond to the initial request for records were contacted via phone by Team AST and received a final request letter in March 2017.

To aid in maximizing facility response rates, our team presented our project details during multiple ESRD community forums (Town Halls, NICE Calls, etc.) in order to increase facility exposure to our validation study. Facilities that did not respond to the request for records were subject to a 10-point reduction to their Total Performance Score (TPS). The response rate for the feasibility study in 2016 was high; of the nine facilities selected for participation, nine responded (100%). For future validations, we plan to follow the same records request methodology, follow-up, and ESRD community outreach approach we've used in the past since it has been effective in producing desired response rates.

Data Validation

The main objective of this analysis is to perform a single comparison of the NHSN system against NHSN "candidate event" data obtained from the facilities' records, leading to an evaluation of the reliability (i.e., the data are reasonably complete and accurate) and validity (i.e., the data actually represent what is being measured) of NHSN data. Candidate events include positive blood cultures, intravenous antimicrobials, or vascular site infection (e.g. pus, redness, or increased swelling).

Reliability: Reliability means data are reasonably complete and accurate, meet intended purposes, and are not subject to inappropriate alteration, where:

- Completeness refers to the extent that relevant records are present and the fields in each record are populated appropriately, and,
- Accuracy refers to the extent recorded data reflect the actual underlying information. In this instance, that the data in the NHSN system accurately reflects the data contained within the source documents, i.e. the facilities' medical records.

Validity: (as used here) refers to whether the data actually represent what one believes is being measured. A number of measures are commonly used to assess validity of any measure.

In our interpretation of these measures, we identify the key sources of overall disagreement between the NHSN data and the patients' medical records, which would serve as the "gold standard." Typical sources of disagreement include missing information about events, inaccurate dates and inaccurate dialysis catheter information.

We propose to use Cohen's Kappa (κ) because it is an overall measure of agreement between the test and reference databases. The kappa coefficient, κ, is calculated as $\kappa = (p_o - p_e) * (1 - p_e)$, where p_o is the observed agreement between two classifications and p_e is the expected agreement between two classifications based on the marginal distributions.

Cohen's Kappa is usually presented as a proportion; we will convert key measures of agreement from proportions to percentages for easier interpretation. All measures will range from -100 % (perfect disagreement) to +100%(perfect agreement). Landis and Koch (1977) have suggested overall agreement between two classifications using Cohen's Kappa be interpreted as shown in **Table 5.**

Table 5: Interpretation of Cohen's Kappa

Kappa (κ)	**Interpretation**
< 0%	Poor
0% to 20%	Slight
20% to 40%	Fair
40% to 60%	Moderate
60% to 80%	Substantial
80% to 100%	Almost Perfect

Although the Landis and Koch interpretation is an old one, it is still widely referenced and is still the dominant one used to indicate the strength of agreement (Cunningham, 2009).

Other coefficients of agreement have been suggested but as Fleiss pointed out (1981), κ has a number of qualities that has made it an attractive option for the measurement of agreement:

- If there is complete agreement, $\kappa = +1$,
- If observed agreement is greater than or equal to chance agreement, $\kappa \geq 0$, and
- If the observed agreement is less than or equal to chance agreement, $\kappa < 0$.

However, over the years a number of difficulties in the interpretation of Cohen's Kappa have been pointed out and several statistical fixes have been proposed. Kappa not only measures agreement, but it is affected in complex ways by the distribution of data across categories that are used ("prevalence") and by bias that may be inherent in the measures used. These are the

problems associated with Kappa (Feinstein and Cicchetti, 1990):

1. If the expected agreement (p_e) is large, the correction process can convert a relatively high value of the observed agreement (p_o) into a low value of Kappa (κ).
2. Unbalanced marginal totals produce higher values of κ than balanced totals.

Kappa is also affected both by any bias between the two measures of gender and by the overall prevalence (the relative probability of the responses – the "Yes" and "No" responses). Byrt, Bishot, and Carlin (1993) introduced measures of prevalence (prevalence index – p_{index}) and bias (bias index – b_{index}) that can be used to compensate for the biases and suggest that these measures are reported together with κ. The p_{index} and b_{index} can then be used to produce a *prevalence-adjusted, bias-adjusted kappa* (*PABAK*) that takes on the values of -100% when there is no agreement and +100% when there is perfect agreement and 0 when the agreement is equal to 50%. Additionally, PABAK is linearly related to p_o.

Given the following standard table with two measures and a dichotomous response, we can therefore add the following measures of agreement shown in **Table 6** below (Byrt, Bishot and Carlin, 1993; Cunningham, 2009).

Table 6: Standard Fourfold Table

Measurement A	**Measurement B, Yes**	**Measurement B, No**	**Total**
Yes	*a*	*b*	*a + b*
No	*c*	*d*	*c + d*
Total	***a +c***	***b +d***	***N***

1. The observed proportion of agreement, $po = (a+d) \div N$
2. Expected proportion of agreement, $p_e = ((a + c)(a + b)+(b + d)(c + d)) \div N^2$
3. Proportion of positive agreement, $p_{pos} = (2a) \div (N + a – d)$
4. Proportion of negative agreement, $p_{neg} = (2d) \div (N – a + d)$
5. Prevalence Index, $p_{index} = (a – d) \div N$
6. Bias Index, $b_{index} = (b – c) \div N$
7. Prevalence-adjusted, bias-adjusted Kappa, $PABAK = 2p_o – 1$

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.

As noted above, the sample pool will consist of Medicare-certified dialysis facilities that are required to submit administrative and clinical data into CROWNWeb in order to meet Section 494.108(h) of the 2008 updated Conditions for Coverage for ESRD Dialysis Facilities. Our previous experience on past CMS NHSN validation efforts have shown near universal compliance with medical record requests. No additional tests of procedures or methods to be undertaken are expected.

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.

Jon Mark Hirshon, University of Maryland-Baltimore, 1 (410) 328-7474
Joseph Kufera, University of Maryland-Baltimore, 1 (410) 328-4161
Darren Childers, Allegheny Science & Technology, 1 (304) 612-1571
Heather Duvall, Healthcare Management Solutions, 1 (304) 816-5262