NHSN Quarterly Validation Call
For State HAI Coordinators

Friday, September 28, 2018

2:00pm – 3:00pm EST
Today’s Agenda

- Introduction
- Update - 2018 HAI Validation Guidance and Toolkits
- Presentation – Facility Selection for External Validation of HAI Data Reported to NHSN: Alternative Approach
- Presentation – Data Validation in North Carolina 2018
- Question & Answer Session
- Wrap-up
NHSN HAI Validation Team

- Suparna Bagchi, MSPH, DrPH, HAI Validation Lead
  - iyj9@cdc.gov
- Bonnie Norrick, MT(ASCP), EdM, CIC, CPHQ
  - ojd8@cdc.gov
- Jennifer Watkins, RN, BSN, MPH
  - nub7@cdc.gov
2018 Validation Guidance and Toolkits

- 2018 External and Internal Validation Guidance and Toolkits are posted!  
  - https://www.cdc.gov/nhsn/validation/index.html
2018 External Validation Guidance and Toolkit

- 2018 External Validation Guidance and Toolkit Updates:
  - Two methods of facility selection
  - Updated instructions, including NHSN screenshots
  - MRATs updated and reformatted

- 2018 Internal Validation Guidance and Toolkit Updates:
  - Addition of Data Quality checklists
MRAT Updates 2018 - Location

NHSN Validation Guidance and Resources for 2018

For Reporting Facilities: 2018 Internal Validation Guidance and Toolkit

For Auditors: 2018 External Validation Guidance and Toolkit

- 2018 External Validation Guidance and Toolkit [PDF - 3 MB]

Medical Record Abstraction Tools (MRAT) and Instructions

- 2018 CLABSI Medical Record Abstraction Tool (MRAT) [PDF - 300 KB] (print-only)
  - 2018 Instructions for CLABSI MRAT [PDF - 300 KB] (print-only)
- 2018 CAUTI Medical Record Abstraction Tool (MRAT) [PDF - 300 KB]
  - 2018 Instructions for CAUTI MRAT [PDF - 300 KB] (print-only)
## MRAT Updates 2018 – New Field

### Case Determination

<table>
<thead>
<tr>
<th>(A) Correctly Classified</th>
<th>(B) Over-reported HAI</th>
<th>(C) Underreported HAI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If CLABSI was misclassified (over- or underreported) by facility, what was the reason?</strong></td>
<td><strong>(II) CLABSI criteria misapplied</strong></td>
<td></td>
</tr>
<tr>
<td>(I) General HAI definition misapplication</td>
<td>(I) Central Line not in &gt; 2 days in an inpatient location on date of event</td>
<td></td>
</tr>
<tr>
<td>(Ia) Incorrect location of attribution</td>
<td>(Iib) Missed CLABSI due to central line removed day of or day before the date of event</td>
<td></td>
</tr>
<tr>
<td>(Ib) Date of event incorrect</td>
<td>(Iic) Missed CLABSI due to location transfer/discharge day of or day before the date of event</td>
<td></td>
</tr>
<tr>
<td>(Ic) IWP set incorrectly</td>
<td>(IId) CLABSI incorrectly identified as secondary BSI</td>
<td></td>
</tr>
<tr>
<td>(Id) RIT applied incorrectly</td>
<td>(Ile) Secondary BSI incorrectly identified as a primary CLABSI</td>
<td></td>
</tr>
<tr>
<td>(Ie) Did not identify elements present in IWP</td>
<td>(II) Other __________________________</td>
<td></td>
</tr>
<tr>
<td>(If) POA/HAI applied incorrectly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ih) Other __________________________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Additional Reasons

| (IIIa) Missed case finding/failure to review positive specimen/culture | (IIId) CLABSI incorrectly identified as secondary BSI |
| (IIib) Clinical over-rule | (IIle) Secondary BSI incorrectly identified as a primary CLABSI |
| (IIic) Used outdated criteria | (IIif) Other __________________________ |
Appendix G: Data Quality Checklist - CLABSI/CAUTI Data

This checklist is intended to ensure completeness and accuracy of CLABSI and CAUTI data entered into NHSN and can be used at acute care hospitals, long term acute care facilities, critical access hospitals, and inpatient rehabilitation facilities.

### Summary Denominator Data

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description/Action</th>
<th>Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) Missing summary data</td>
<td>Verify that summary data has been entered for the location and month/year. (Go to NHSN Application -&gt; Alerts -&gt; Missing Summary Data)</td>
<td></td>
</tr>
<tr>
<td>ii) Missing denominator variables (Incomplete summary data)</td>
<td>Verify that all mandatory/required fields are completed, and that &quot;Report No Events&quot; is checked, if appropriate. (Go to NHSN Application -&gt; Alerts -&gt; Incomplete Summary Data)</td>
<td></td>
</tr>
<tr>
<td>iii) Verify denominator data accuracy</td>
<td>Generate Rate Tables to display location and month in a table!</td>
<td></td>
</tr>
</tbody>
</table>

### Event Data Entry

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description/Action</th>
<th>Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) All CLABSI and CAUTI events reported</td>
<td>Verify that all CLABSI and CAUTI events have been reported. Go to NHSN Application -&gt; Analytic -&gt; Reports -&gt; Device-Associated (DA) Module -&gt; Central Line-Associated BSI -&gt; Line Listing - All CLAB Events OR NHSN Application -&gt; Analytic - Reports - Device-Associated (DA) Module - Urinary Catheter-Associated UTI - Line Listing - All CAU Events</td>
<td></td>
</tr>
<tr>
<td>ii) Missing numerator variables (Incomplete events)</td>
<td>Verify that all mandatory/required data fields (marked with an *, **, or &gt; on the event form) are completed. (Go to NHSN Application -&gt; Alerts -&gt; Incomplete Events, Event Type: BS/UTI)</td>
<td></td>
</tr>
<tr>
<td>iii) Confirm that date of event occurred on or after</td>
<td>If the event did not occur on or after the third</td>
<td></td>
</tr>
</tbody>
</table>
Today’s Speakers

- Suparna Bagchi, MSPH, DrPH
  - HAI Validation Lead
  - CDC NHSN Protocol and Validation Team
  - iyj9@cdc.gov

- Savannah Carrico, MPH
  - HAI Epidemiologist, SHARPPS Program
  - North Carolina Division of Public Health
  - savannah.carrico@dhhs.nc.gov
Facility Selection for External Validation of HAI Data Reported to NHSN: Alternative Approach

Suparna Bagchi, MSPH, DrPH
HAI Validation Lead
Protocol and Validation Team
September 28, 2018
Objectives

- Review the methods of facility selection in NHSN External Validation Guidance
- New method of facility selection in 2018 Guidance
- Comparison of facility selection methods
- Recommended data analysis and summarization
Facility Selection Method 1

- Targeted sampling: facility specific predicted events and SIR
- Facilities are sorted based on predicted number of events
- Top third of facilities (tertiles):
  - Targeting and prioritization
  - Facility specific SIR relative to median SIR for the top tertile of the facilities
- SIR does not estimate absolute burden of HAIs in a facility
- Ratio of observed/predicted events
- Focuses on larger (higher burden facilities), excludes smaller facilities where underreporting could be a potential problem
Method 2: Alternative Approach

- Underreporting of HAI remains primary concern
- Cumulative Attributable Difference (CAD) approach
  \[ \text{CAD} = \text{Observed HAIs} - (\text{Predicted HAIs} \times \text{SIR Goal}) \]
  \[ \text{CAD} = \text{Observed events} - \text{Predicted events} \]
- Facilities could have both positive and negative CAD values
- Facilities reporting zero or very few events: negative CAD value
- Prioritization based on highest negative CAD values can help assess the data accuracy among facilities with high predicted and very few or no reported events during a time frame
## Comparison of Facility Selection Methods

<table>
<thead>
<tr>
<th>Facility Selection criteria</th>
<th>Method 1 - Prioritizing Facilities with Highest Likelihood of Event Occurrence</th>
<th>Method 2 - Cumulative Attributable Difference (CAD) Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on highest likelihood of event occurrence.</td>
<td>Based on difference of predicted and observed number of events.</td>
<td>Prioritization focuses on facilities with negative values of difference, primarily under-reporters</td>
</tr>
<tr>
<td>Larger facilities with higher predicted/expected number of events are more likely to be selected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Which type of facilities are selected?**
  - Larger facilities with higher predicted/expected number of events are more likely to be selected
  - Prioritization focuses on facilities with negative values of difference, primarily under-reporters
# Comparison of Facility Selection Methods

<table>
<thead>
<tr>
<th>Method 1 - Prioritizing Facilities with Highest Likelihood of Event Occurrence</th>
<th>Method 2 - Cumulative Attributable Difference (CAD) Approach</th>
</tr>
</thead>
</table>
| **Ranking algorithm** | • SIR metric is a ratio of and is subject to variability  
• A small facility with low predicted volume of events with even one observed event could lead to a high SIR value.  
• Cumulative attributable difference (CAD)  
• CAD metric is robust, stable and reflects the true facility HAI burden |
| **Which method should my state use?** | • No prior validation, use Method 1 to determine errors in HAI misclassification  
• If already aware of underreporting concerns - select Method 2 |
| | • Previous validation history that have identified underreporting as a potential concern would benefit additionally with this method |
CAD Method of Facility Selection

- Generate new datasets in NHSN
- After successful dataset generation, navigate to Analysis
- Navigate to the SIR report of interest
- Export Analysis Data Set screen - export to an Excel spreadsheet
- Exported SIR report file will display multiple levels of aggregation
- In Excel, select the aggregation level that provides a facility-specific SIR for all validation locations
<table>
<thead>
<tr>
<th></th>
<th>1.71106</th>
<th>1030</th>
<th>0.1252</th>
<th>2.338</th>
<th>0.743</th>
<th>5.639</th>
<th>IN:ACUTE:CC:M_PED</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.18614</td>
<td>2824</td>
<td>0.5558</td>
<td>0.628</td>
<td>0.105</td>
<td>2.074</td>
<td>IN:ACUTE:CC:NS</td>
</tr>
<tr>
<td>7</td>
<td>9.78433</td>
<td>7188</td>
<td>0.3846</td>
<td>0.715</td>
<td>0.313</td>
<td>1.415</td>
<td>IN:ACUTE:CC:NURS</td>
</tr>
<tr>
<td>9</td>
<td>17.3151</td>
<td>15646</td>
<td>0.0327</td>
<td>0.52</td>
<td>0.253</td>
<td>0.954</td>
<td>IN:ACUTE:CC:S</td>
</tr>
<tr>
<td>2</td>
<td>3.26009</td>
<td>2134</td>
<td>0.531</td>
<td>0.613</td>
<td>0.103</td>
<td>2.027</td>
<td>IN:ACUTE:CC:T</td>
</tr>
</tbody>
</table>

SIRs for each location types

<table>
<thead>
<tr>
<th></th>
<th>5.18005</th>
<th>4183</th>
<th>0.3493</th>
<th>0.578</th>
<th>0.147</th>
<th>1.574</th>
<th>HOSP-GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.06437</td>
<td>7682</td>
<td>0.0013</td>
<td>0.11</td>
<td>0.006</td>
<td>0.544</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>7.57817</td>
<td>6272</td>
<td>0.0235</td>
<td>0.264</td>
<td>0.044</td>
<td>0.872</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>0.42346</td>
<td>562</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>0.7199</td>
<td>823</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>1.0873</td>
<td>1253</td>
<td>0.3934</td>
<td>1.839</td>
<td>0.308</td>
<td>6.077</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>0.44531</td>
<td>591</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>1</td>
<td>1.05264</td>
<td>933</td>
<td>1</td>
<td>0.95</td>
<td>0.048</td>
<td>4.685</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>13</td>
<td>18.5196</td>
<td>15267</td>
<td>0.1921</td>
<td>0.702</td>
<td>0.390</td>
<td>1.170</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>0.15574</td>
<td>232</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>1</td>
<td>1.52725</td>
<td>1760</td>
<td>0.7659</td>
<td>0.655</td>
<td>0.033</td>
<td>3.229</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>3</td>
<td>1.57237</td>
<td>1812</td>
<td>0.2846</td>
<td>1.908</td>
<td>0.485</td>
<td>5.193</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>3.62945</td>
<td>2146</td>
<td>0.4283</td>
<td>0.555</td>
<td>0.093</td>
<td>1.8 CC</td>
<td>HOSP-GEN</td>
</tr>
</tbody>
</table>

Facility-specific SIRs combining all location types

<table>
<thead>
<tr>
<th></th>
<th>1.5876</th>
<th>1034</th>
<th>0.7334</th>
<th>0.63</th>
<th>0.032</th>
<th>3.1 CC_N</th>
<th>HOSP-GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.26831</td>
<td>7012</td>
<td>0.0026</td>
<td>0.121</td>
<td>0.006</td>
<td>0.5 CC</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>0.79606</td>
<td>670</td>
<td></td>
<td></td>
<td></td>
<td>CC_N</td>
<td>HOSP-GEN</td>
</tr>
</tbody>
</table>

THIS IS THE LEVEL TO EVALUATE

Facility and location types
Calculate the 75th Percentile Value of numPred

Sort the facilities in the descending order of number of predicted infections (numPred) and compute the 75th percentile value of the variable numPred.
Selection of Facility Sampling Frame

<table>
<thead>
<tr>
<th>infCount</th>
<th>numPred</th>
<th>numclday</th>
<th>SIR_pval</th>
<th>SIR</th>
<th>sir95ci</th>
<th>locationTy</th>
<th>locCDC</th>
<th>orgID</th>
<th>facType</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>18.51959</td>
<td>15267</td>
<td>0.1921</td>
<td>0.702</td>
<td>0.390, 1.170</td>
<td></td>
<td></td>
<td>100008</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>22</td>
<td>15.32671</td>
<td>9910</td>
<td>0.1034</td>
<td>1.435</td>
<td>0.922, 2.138</td>
<td></td>
<td></td>
<td>100030</td>
<td>HOSP-CHL</td>
</tr>
<tr>
<td>10</td>
<td>9.736101</td>
<td>8387</td>
<td>0.8926</td>
<td>1.027</td>
<td>0.522, 1.831</td>
<td></td>
<td></td>
<td>100014</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>8</td>
<td>9.542312</td>
<td>7595</td>
<td>0.6509</td>
<td>0.838</td>
<td>0.389, 1.592</td>
<td></td>
<td></td>
<td>100046</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>1</td>
<td>9.0643/3</td>
<td>7/682</td>
<td>0.0013</td>
<td>0.11</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100001</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>7.578169</td>
<td>6272</td>
<td>0.0235</td>
<td>0.264</td>
<td>0.044, 0.872</td>
<td></td>
<td></td>
<td>100002</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>7</td>
<td>5.689505</td>
<td>4581</td>
<td>0.5585</td>
<td>1.23</td>
<td>0.538, 2.434</td>
<td></td>
<td></td>
<td>100022</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>4</td>
<td>5.504663</td>
<td>4879</td>
<td>0.558</td>
<td>0.727</td>
<td>0.231, 1.753</td>
<td></td>
<td></td>
<td>100027</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>3.159258</td>
<td>2784</td>
<td>0.5651</td>
<td>0.76</td>
<td>0.389, 1.592</td>
<td></td>
<td></td>
<td>100015</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>2.437844</td>
<td>2304</td>
<td>0.8601</td>
<td>0.143</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100001</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>1.945079</td>
<td>1724</td>
<td>0.2846</td>
<td>0.12</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100002</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>3</td>
<td>1.572374</td>
<td>1812</td>
<td>0.2846</td>
<td>0.12</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100003</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>1</td>
<td>1.527251</td>
<td>1760</td>
<td>0.2846</td>
<td>0.12</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100004</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>1.329405</td>
<td>1357</td>
<td>0.5333</td>
<td>0.727</td>
<td>0.231, 1.753</td>
<td></td>
<td></td>
<td>100027</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>0</td>
<td>1.242188</td>
<td>1101</td>
<td>0.2846</td>
<td>0.12</td>
<td>0.006, 0.544</td>
<td></td>
<td></td>
<td>100004</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>2</td>
<td>1.087298</td>
<td>1253</td>
<td>0.3934</td>
<td>1.839</td>
<td>0.308, 6.077</td>
<td></td>
<td></td>
<td>100005</td>
<td>HOSP-GEN</td>
</tr>
<tr>
<td>4</td>
<td>1.087298</td>
<td>1253</td>
<td>0.3934</td>
<td>1.839</td>
<td>0.308, 6.077</td>
<td></td>
<td></td>
<td>100005</td>
<td>HOSP-GEN</td>
</tr>
</tbody>
</table>

75th percentile value of numPred = 5.5. Select facilities with numPred > 5.5. Only facilities in red box (numPred > 5.5) are included in the sampling frame for targeted validation.
Compute the CAD Values for Sampling Frame

- **Variable infCount**
  - Pooled total observed events from all validation locations, for the timeframe of validation for each facility selected in sampling frame

- Insert a column (CAD) next to the numPred

- Compute CAD as difference: \( \text{infCount} - \text{numPred} \)

- Could generate – all negative, positive and negative, all positive
Sort the Facilities by CAD Values

Compute the CAD values for facilities in the sampling frame.

Sort the CAD values in descending order (highest negative on the top). If the sampling frame has greater than 15 facilities, select the top 15 facilities.
Facility Selection: If Sampling Frame > 30 Facilities

- Divide the total facilities in the sampling frame into two strata:
  - Stratum 1: Includes all facilities in the sampling frame that had zero reported pooled observed events for the validation time frame
  - Stratum 1: will generate all negative CAD values
  - Stratum 2: includes all facilities in the sampling frame with non-zero reported pooled observed events for the validation time frame
  - Stratum 2: could generate positive and negative CAD values
Stratum 1: Facilities with Zero Reported Events

- All CAD values will be negative.
- Highest negative values: facilities with greater predicted and zero events reported
- Sort them in descending order of negative values of CAD
- Facilities with the highest negative CAD value should be at the top
- Select the first 15 facilities from Stratum A.
Stratum 2: Facilities with Non-zero Reported Events

- CAD values could be positive or negative
- Highest negative values: facilities with greater predicted and zero events reported
- Sort them in descending order of negative values of CAD
- Facilities with the highest negative CAD value should be at the top
- Select the first 15 facilities from Stratum B
Facility Sampling Using CAD Approach

- Distribution of predicted number of events, use the 75th percentile value as threshold
- If value > 1, then use the value corresponding to 75th percentile, otherwise value = 1
- Create a subset of facilities in state with predicted events greater than the threshold

If subset is ≤ 30 facilities – validate all
If subset > 30 facilities, facility selection

Calculate the pooled total of observed events among the facilities in sampling frame

**Stratum 1: Zero events reported**
- All values negative CAD
- Highest negative CAD: High predicted/zero events
- Sort – descending order absolute CAD values
- Select top 15 facilities

**Stratum 2: Non-zero events reported**
- CAD values: negative and positive
- Sort – descending order absolute CAD values
- Select top 15 facilities
Medical Record Selection: CAD Approach

- Before requesting medical records: download ("freeze") data
- Request facilities to send line lists of candidate HAI events
- For facilities with reported events in validation locations:
  - Events reported to NHSN in the validation time frame (select all)
  - Randomly select additional medical records for a total of 40 medical records for candidate cases.
- For facilities with no reported event in validation locations:
  - Randomly select 40 medical records for review for each HAI candidate event.
### Recommended Data Summary

<table>
<thead>
<tr>
<th>Facility Events</th>
<th>Auditor Determination</th>
<th>Not Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events reported</td>
<td>True Positive (a)</td>
<td>False Positive (b) Over reports</td>
</tr>
<tr>
<td>Events not reported</td>
<td>False Negative (c) Missed events</td>
<td>True Negative (d)</td>
</tr>
<tr>
<td>(a+c)</td>
<td>(b+d)</td>
<td>Total</td>
</tr>
</tbody>
</table>

- **True positive (a):** facility identified and reported the events and auditor agreed
- **True negative (d):** facility did not identify/report event and auditor agreed
- **False negative (c):** facility did not identify/report event and auditor disagreed *(MISSED)*
- **False positive (b):** facility identified and reported the events and auditor disagreed *(OVER REPORT)*
### Recommended Data Analysis

<table>
<thead>
<tr>
<th></th>
<th>Auditor Determination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Facility Events</td>
<td>Not Events</td>
</tr>
<tr>
<td>Events reported</td>
<td>True Positive (a)</td>
<td>False Positive (b)</td>
</tr>
<tr>
<td>Events not reported</td>
<td>False Negative (c)</td>
<td>True Negative (d)</td>
</tr>
<tr>
<td></td>
<td>(a+c)</td>
<td>(b+d)</td>
</tr>
</tbody>
</table>

- **Sensitivity**: Ability of a test to correctly identify those with the disease (true positive rate) = \( \frac{a}{a+c} \)
- **Specificity**: Ability of the test to correctly identify those without the disease (true negative rate) = \( \frac{d}{b+d} \)
- **Positive Predictive Value**: Proportion of individuals who test positively (a+b) AND truly have the disease (a) = \( \frac{a}{a+b} \)
- **Negative Predictive Value**: Proportion of individuals who test negatively (c+d) AND truly do not have the disease (d) = \( \frac{d}{c+d} \)
Reasons for Misclassification

- For each misclassified case, list the reasons for errors in reports
- Compute proportion of each error type – identify gaps, training opportunities

**Reasons for under-reported CDI events**
- Incorrect understanding of protocol definition (n1)
- Laboratory records missed (n2)
- Reason ....

Total Under-reported events

**Reasons for over-reported CDI events**
- Incorrect specimen (n1)
- Duplicate record (n2)
- Reason ....

Total Over-reported events
Summary and Recommendations

- Both facility selection methods use a targeted approach
- Generalizability is still limited
- Select the method as deemed appropriate
- Compare same HAI validated previously validated using alternative method
- Feedback on implementation: challenges and successes
Questions!

Suparna Bagchi – iyj9@cdc.gov
Jennifer Watkins – nub7@cdc.gov
Bonnie Norrick – ojd8@cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.
Data Validation in North Carolina 2018

Savannah Carrico, MPH
HAI Epidemiologist
September 28, 2018
Outline

I. Importance of Data Validation

II. Hospital Selection Method: SIR and CAD

III. Results of North Carolina’s CDI and CLABSI validations
Importance of Data Validation

• Non punitive validation
Importance of Data Validation

- Non punitive validation
- Engages health care facilities in accurate data collection methods
Importance of Data Validation

• Non punitive validation
• Engages health care facilities in accurate data collection methods
• The goal identify the true burden of HAIs
Importance of Data Validation

- Non punitive validation
- Engages health care facilities in accurate data collection methods
- The goal identify the true burden of HAIs
- Accurate data in NHSN allows for comparable data
Importance of Data Validation

• Non punitive validation
• Engages health care facilities in accurate data collection methods
• The goal identify the true burden of HAIs
• Accurate data in NHSN allows for comparable data
• Opportunity for facilities and validators to discuss HAI prevention and response
SHARPPPS Program Data Validation

• The North Carolina Surveillance for Healthcare-Associated Resistant Pathogens Patient Safety (SHARPPPS) Program has been performing data validation HAIs since 2015

• SHARPPPS performs data validation without funding

• Since 2015 CLABSI, CDI, CAUTI, and MRSA have been validated
Selecting a Sample

- There 93 Acute Care Hospitals in North Carolina
- CDC recommends 18 facilities be selected for states that have 21-149 hospitals
- Want to select hospitals that represent the state
- Selecting those that would benefit the most from data validation
- Must select hospitals without introducing bias
Selection Bias

- Want to avoid asking facilities to self-select
- Want to select representative facilities
- Want to target facilities that would benefit the most
CDC methodology
SIR Report for HAI

Sort by number of predicted events
SIR Report for HAI

Sort by number of predicted events

Tertile 1

Tertile 2

Tertile 3
SIR Report for HAI

Sort by number of predicted events

Tertile

Group A
SIR Above Median

Group B
SIR Below Median

Group C
SIR = 0

Tertile 2

Tertile 3
SIR Report for HAI

Sort by number of predicted events

Tertile

Group A
SIR Above Median

Group B
SIR Below Median

Group C
SIR = 0

18 Facilities

Tertile 2

5% Random Sample

4 Facilities

Tertile 3
Results

The majority of facilities were:

- All 18 facilities were in the top tertile
  - *Highest number of predicted events*
- In urban areas
  - *North Carolina is 80% rural*
  - *67 of 93 hospitals are in rural counties*
- Trauma centers
  - *Affiliated with major medical schools*
  - *Experience high volume of higher acuity patients*
# CDC Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td></td>
</tr>
</tbody>
</table>
### CDC Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td></td>
</tr>
<tr>
<td>• Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR</td>
<td></td>
</tr>
</tbody>
</table>
## CDC Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td>• Excludes facilities with &lt; 1 Predicted Event</td>
</tr>
<tr>
<td>• Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR</td>
<td></td>
</tr>
</tbody>
</table>
## CDC Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td>• Excludes facilities with &lt; 1 Predicted Event</td>
</tr>
<tr>
<td>• Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR</td>
<td>• Excludes smaller facilities</td>
</tr>
</tbody>
</table>
**CDC Methodology Review**

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td>• Excludes facilities with &lt; 1 Predicted Event</td>
</tr>
<tr>
<td>• Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR</td>
<td>• Excludes smaller facilities</td>
</tr>
<tr>
<td></td>
<td>• Weighted selection of facilities (Top Tertile only)</td>
</tr>
</tbody>
</table>
## CDC Methodology Review

### Positives
- Focuses on high-burden facilities
- Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR

### Considerations
- Excludes facilities with < 1 Predicted Event
- Excludes smaller facilities
- Weighted selection of facilities (Top Tertile only)
- SIR doesn’t estimate the absolute burden of HAIs on a facility because it is a ratio of observed to predicted infections
### CDC Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focuses on high-burden facilities</td>
<td>• Excludes facilities with &lt; 1 Predicted Event</td>
</tr>
<tr>
<td>• Acknowledges potential for over- and under-reporting within the top third of facilities by stratifying by Median SIR</td>
<td>• Excludes smaller facilities</td>
</tr>
<tr>
<td></td>
<td>• Weighted selection of facilities (Top Tertile only)</td>
</tr>
<tr>
<td></td>
<td>• SIR doesn’t estimate the absolute burden of HAIs on a facility because it is a ratio of observed to predicted infections</td>
</tr>
<tr>
<td></td>
<td>• Relies on accurate risk-adjustment of facilities</td>
</tr>
</tbody>
</table>
Cumulative Attributable Difference Methodology

CAD = Observed # HAIs – (Predicted # HAIs * SIR Goal)

- Calculated even if the number of predicted events is < 0 (Unlike SIR)
- Represents the number of infections needed to be prevented to reach SIR goal
- The CAD can be used to identify facilities that would benefit the most from data validation
- NOT used for interfacility comparison

SOURCE:
DATA VALIDATION PRES | SEPTEMBER 28, 2018 | V1
CAD methodology
93 Acute Care Hospitals (North Carolina)

Stratum 1
(No CLABSIs reported)

Stratum 2
(> 0 CLABSI reported)
93 Acute Care Hospitals (North Carolina)

Stratum 1 (No CLABSIIs reported)

Stratum 2 (> 0 CLABSI reported)

Organize Strata by Absolute CAD and select 9 largest

9 ACHs

9 ACHs

18 ACHs
## Cumulative Attributable Difference Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Captures facilities with &lt; 1 predicted event</td>
<td></td>
</tr>
</tbody>
</table>
### Cumulative Attributable Difference Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Captures facilities with &lt; 1 predicted event</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Includes equal sample of facilities with 0 HAI events and &gt; 0 HAI events</strong></td>
<td></td>
</tr>
</tbody>
</table>
## Cumulative Attributable Difference Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Captures facilities with &lt; 1 predicted event</td>
<td></td>
</tr>
<tr>
<td>• Includes equal sample of facilities with 0 HAI events and &gt; 0 HAI events</td>
<td></td>
</tr>
<tr>
<td>• CAD accurately reflects absolute HAI Burden on a hospital</td>
<td></td>
</tr>
<tr>
<td>Positives</td>
<td>Considerations</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>• Captures facilities with &lt; 1 predicted event</td>
<td></td>
</tr>
<tr>
<td>• Includes equal sample of facilities with 0 HAI events and &gt; 0 HAI events</td>
<td></td>
</tr>
<tr>
<td>• CAD accurately reflects absolute HAI Burden on a hospital</td>
<td></td>
</tr>
<tr>
<td>• Potential to identify facilities with excellent prevention</td>
<td></td>
</tr>
</tbody>
</table>
## Cumulative Attributable Difference Methodology Review

<table>
<thead>
<tr>
<th>Positives</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Captures facilities with &lt; 1 predicted event</td>
<td></td>
</tr>
<tr>
<td>• Includes equal sample of facilities with 0 HAI events and &gt; 0 HAI events</td>
<td></td>
</tr>
<tr>
<td>• CAD accurately reflects absolute HAI Burden on a hospital</td>
<td></td>
</tr>
<tr>
<td>• Potential to identify facilities with excellent prevention</td>
<td></td>
</tr>
<tr>
<td>• Relies on accurate risk-adjustment of facilities</td>
<td></td>
</tr>
</tbody>
</table>
Choosing a methodology

• The current method (SIR) has its pros but there are several considerations

• The CAD method:
  – addresses the considerations of the SIR method
  – selected representative facilities of North Carolina
  – captured both under and overreporting facilities
  – method was chosen as the selection method
CDI Results

• 20 Facilities Validated
  – 13 ACHs
  – 2 LTACHs
  – 5 IRFs

• 1542 records validated

• 1 validator per record

• 95 % Facility and Validator Agreement

• 5% (79 records) not reported in NHSN that should have been
  – 87% (69 records) of these records were community onset
CLABSI Results

• 12 Facilities Validated
• 293 Records Reviewed
• 2 validators per record
• 98% Agreement between facility and validators
• 94% Agreement between validators
• 2% (6 records) were discrepant
  – 1 record was misclassified as not a CLABSI by the facility
  – 6 records were misclassified as CLABSIs by the facility
    • 3 records Secondary to other infections
    • 2 records were not in reporting locations
    • 1 record had no central line
The figure shows the prevalence of different organisms in positive blood cultures reviewed. The top left chart lists the organisms along with their prevalence:

- S. aureus: 43%
- CNS: 17%
- E. coli: 12%
- Candida spp: 7%
- E. faecalis: 7%
- K. pneumoniae: 5%
- P. aeruginosa: 5%
- Viridans strep: 3%
- S. marcescens: 2%

The top right chart shows the prevalence of different CLABSI organisms:

- Candida spp: 31%
- CNS: 17%
- E. faecalis: 14%
- S. aureus: 14%
- E. coli: 10%
- P. aeruginosa: 7%
- S. marcescens: 7%
In Summary

• Primary goal is to capture generalizable and representative data for the state

• The high agreement between facilities and validators suggests a thorough understanding of the NHSN surveillance definitions for CDIs and CLABSIs

• Future validations would be beneficial for all HAIs
Acknowledgements

The North Carolina Surveillance for Healthcare-Associated Resistant Pathogens Patient Safety Program would like to acknowledge and appreciate all participating healthcare facilities in North Carolina.
Savannah Carrico
HAI Epidemiologist
North Carolina Division of Public Health
Medical Consultation Unit
SHARPPS Program
Savannah.Carrico@dhhs.nc.gov
Questions??
Wrap-Up

- Next Quarterly Call scheduled for Friday, January 11, 2019 from 2-3pm EST

- Is there anyone else we should invite? Please forward their name and email to Bonnie Norrick ojd8@cdc.gov.

- If you are interested in sharing your validation experience on a Quarterly Validation Call, please reach out to the NHSN HAI Validation Team
Thank You!

Please Join us for the Next

NHSH Quarterly Validation Call for HAI Coordinators

Friday, January 11, 2019  2:00pm—3:00pm EST

For Questions Email NHSN@cdc.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.