Evaluation of HHS Protect data

Mathew R P Sapiano, PhD
DHQP, Lantana Consulting Services, LLC
November 2020
A HHS Protect denominator dataset has been generated, but it is incomplete—Does not include some facilities that have reported data

The denominator is a mix of CCNs and facility orgids from NHSN—Includes 6537 CCNs and 178 other IDs—6537 CCNs mix of single- and multi-site—Unclear if CCNs represent correct number of inpatient beds...

The denominator dataset is missing 28 facilities that appear in Unified Prioritized Timeseries since Nov 1

1. Inaccurate and incomplete denominator data make inference regarding data completeness impossible
2. Ancillary facility information is not available
Are CCNs being used correctly? One example...

- Example of potential issue with assignment of CCNs
  - Avera St Luke’s (Aberdeen, SD): CCN=430014; 70 inpatient beds – \textit{included in prioritized dataset}
  - Avera St Luke’s hospital (Aberdeen, SD): CCN=435014; 10 inpatient beds – \textit{excluded in prioritized dataset; CCN not valid!}

- Is this a multi-site CCN? This facility has many units (including inpatient mental health)
  - Why is this other data being reported daily to TeleTracking?

<table>
<thead>
<tr>
<th>hospital_name</th>
<th>ccn</th>
<th>address</th>
<th>state</th>
<th>zip</th>
<th>collection_date</th>
<th>inpatient_beds</th>
<th>inpatient_beds_used</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVERA ST LUKES</td>
<td>430014</td>
<td>305 S STATE ST POST SD</td>
<td>SD</td>
<td>57401</td>
<td>11/16/2020</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td>Avera St. Luke's Hospital</td>
<td>435014</td>
<td>305 S State Street</td>
<td>SD</td>
<td>57401</td>
<td>11/16/2020</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>AVERA ST LUKES</td>
<td>430014</td>
<td>305 S STATE ST POST SD</td>
<td>SD</td>
<td>57401</td>
<td>11/15/2020</td>
<td>70</td>
<td>44</td>
</tr>
<tr>
<td>Avera St. Luke's Hospital</td>
<td>435014</td>
<td>305 S State Street</td>
<td>SD</td>
<td>57401</td>
<td>11/15/2020</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>AVERA ST LUKES</td>
<td>430014</td>
<td>305 S STATE ST POST SD</td>
<td>SD</td>
<td>57401</td>
<td>11/14/2020</td>
<td>70</td>
<td>44</td>
</tr>
</tbody>
</table>

Note: HHS Denominator lists 109 “CMS” beds
Comparisons of hospital data

Comparison of three timeseries:

- NHSN National Estimates (V4 from April 1 to July 14, 2020)
- HHS Protect hospital data (from July 15, 2020; prioritized facilities only)
  - Show unweighted sum of reported facilities
  - HHS “Estimates” – assumed to be a weighted national/state estimate
  - Created a simple weighted sum [no imputation; no error estimates]
    - Weighted to the HHS Protect sampling frame (uses CCNs)
- COVID Tracking Project: compilation of state estimates, which are being used by media and others
HHS Protect National numbers seem too low

Inpatients, United States

Total estimated inpatient beds is lower in HHS Protect than NHSN.

As a result: simple weighted estimate is lower than NHSN Estimate.

Note trend in unweighted data.
The HHS Estimates seem too low

HHS Estimates ->

Methods are unclear, but presented as a national weighted estimate

Not widely available – WH taskforce and HHS only
HHS Protect National numbers seem too low

Inpatients, United States

HHS Estimate in-line with unweighted... Shouldn’t it be higher?
HHS Protect National numbers seem too low

Inpatients, United States

HHS Estimate in-line with unweighted... Shouldn’t it be higher?
What is the correct number of inpatient beds?

Inpatients, Florida

Inpatients, Massachusetts

Inpatients, Arkansas

Inpatients, Pennsylvania

Legend:
- NHSN Estimated (95% CI band)
- Unweighted HHS Protect beds occupied
- Simple weighted HHS Protect beds occupied
The unweighted HHS Protect timeseries is mis-leading
HHS Protect and NHSN are close, but not the same
States may use different definitions of COVID inpatients

Inpatients with COVID-19, Maryland

Maryland reported 707 COVID-19 hospitalizations on 11/9
### Summary of issues

- **HHS Protect hospital surveillance data has major flaws**
  - Reading and using the data is unnecessarily difficult
  - Use of CCNs makes epi and statistical analyses difficult (national and county estimation)
  - Doubts that all constituent units/facilities within each CCN are included in reporting

- **HHS Protect data has bias due to non-response**
  - As reporting increased from Jul-Nov, the unweighted estimate also increased...
  - *Are percentage metrics also biased?*

- **HHS Protect “Estimates” seem too low – they are about the same as the unweighted estimates... are they actually being weighted? To what?**

- **Newspapers and states are using hospital data collected by states**
  - Seeks too low in some states... likely that they are including only confirmed cases
  - *Mixed definitions lead to a definite undercount of hospital cases!*
  - *Shows that a coordinated, transparent, public reporting Federal system is needed*
Additional material
Source of HHS Protect data by state

- Prioritized reporting source in HHS Protect by state
  - Y-axis is number of facilities
  - X-axis runs from Apr 1 to Nov 9
- NHSN was the priority until July 14