Introduction:
1. Survival analysis is an essential application of cancer surveillance data. Presumed alive assumption can be used for relative survival analyses.
2. The assumption makes the methodology sensitive to the date quality in incidence reporting, date of diagnosis (DX), date of last contact (DLC), and vital status. Pre-analysis data QC is important.
3. The study describes an informative way to use readily available historical incidence data to visualize unusual data patterns systematically and quickly, potentially before the data being submitted to NPCR.

Study Data:
1. NPCR data submissions from November 2016, 2017 and 2018 are used.
2. States with consistent patient IDs between two or more adjacent submissions

Method:
1. The study is based on the findings of two previous internal studies of NPCR: death status reporting delay causing overestimation of survival, and incidence reporting delay causing underestimation of survival.
3. In each set, cases are separated into 2 groups: existing – cases in prior and current submissions; new – cases in current but not prior submission. The groups are further categorized by vital status.
4. Cases in two adjacent submission are merged by patient ID and state.
5. Major reports for each set of analyses:
   - Existing deceased cases without DLC
   - Existing deceased cases that changed DLC
   - Existing deceased cases that became alive
   - Existing alive cases that became dead
   - New alive cases reporting patterns
   - New deceased cases reporting patterns
   - Corresponding patterns by site for each report
6. Patterns are assessed to establish acceptable normal pattern of each report.
7. Unusual patterns are evaluated with historical baseline to determine if the patterns are unique or persistent across submissions. Cases in the unusual patterns are extracted for further investigation
8. The result section demonstrates two pairs of patterns, normal, unusual. The ZONE FOR CHECKING defines a targeted area for detecting unusual patterns.

Results and Discussions (State level patterns for Submission 2017 vs. 2018):

Table 1a: Existing Deceased Cases Changed DLC in Submission 2017 vs. A State: Normal Pattern

<table>
<thead>
<tr>
<th>Zone for Checking</th>
<th>DLC Year in 2017</th>
<th>DLC Year in 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal Zone</td>
<td>0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 2 3 0 0 0</td>
</tr>
<tr>
<td>Total</td>
<td>0 0 1 1 1 2 1 0 3</td>
<td>0 0 3 0 85 0 85 0 85</td>
</tr>
</tbody>
</table>

Table 1b: Existing Deceased Cases Changed DLC in Submission 2018: Unusual Pattern

<table>
<thead>
<tr>
<th>Zone for Checking</th>
<th>DLC Year in 2017</th>
<th>DLC Year in 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal Zone</td>
<td>0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 9 13 0 0 0</td>
</tr>
<tr>
<td>Total</td>
<td>0 0 0 0 0 0 0 107</td>
<td>133 298 276 4 0 276</td>
</tr>
</tbody>
</table>

Results and Discussions:
1. As exhibited in Table 1 to 2, the method can facilitate quick visual detection of unusual data patterns and precisely outline the regions holding the cases of these patterns.
2. The historical baseline can be used to validate if the patterns are new or persistent to a state. The capability to provides boundary information of unusual patterns can be useful to researchers to extract these cases for further investigation. This may help resolve some data issues before data submission to NPCR.
3. In practice, multiple baselines can be obtained to assist in determining the onset and scope of some of the data issues which in turn may improve data collection practices.

Conclusions:
1. Using multiple years of NPCR submission data to monitor data quality for survival is a useful way to visually detect unusual case reporting patterns in areas that may impact survival estimates the most.
2. The method seems to be most effective if it is preformed before data submission to NPCR.
3. The method works well with a defined research goal, such as survival, where the subject knowledge helps to devise more meaningful data check algorithms. However, the method can potentially be applied beyond survival analysis to all data items for cancer data.
4. Further enhancement of this method can potentially benefit the states to use it as a tool for data quality control before submission.

Contact:
Xing Dong: xdong@icfi.com
Reda Wilson: dl0f@cdc.gov

National Center for Chronic Disease Prevention and Health Promotion
Division of Cancer Prevention and Control