Enhancing the completeness of birthplace data through linkage to death certificate data: an assessment from the California Cancer Registry Database

Andrea Sipin, MS, CPHI, CTR1 · Cheryl Moody, CTR2 · Deepa Padmanaban2 · Kathleen Davidson-Allen, CTR3 · Kyle L. Ziegler, BS, CTR4
1Los Angeles Cancer Surveillance Program, Los Angeles, CA, 2California Cancer Registry, Sacramento, CA, 3Cancer Prevention Institute of California, Fremont, CA, 4Cancer Registry of Greater California, Sacramento, CA

**Background**

Birthplace is a strong predictor of cancer risk with the potential to reveal new insights into etiology and prevention. In our commitment to providing high-quality data to support cancer surveillance and population-based research, the Cancer Prevention Institute of California, the Los Angeles Cancer Surveillance Program, and the Cancer Registry of Greater California, who make up the Regional Registries of the California Cancer Registry (CCR), conduct an annual linkage between the California Department of Vital Statistics and the statewide cancer registry database. This linkage initiates the death clearance process, utilizing death certificates to enhance the data quality of vital status information in the statewide database and capturing unreported cancer cases. Historically, capture of birthplace data from traditional cancer reporting sources has been a challenge, but leveraging death data through Vital Statistics linkage will allow us to enhance patient sociodemographic data for surveillance and research.

**Purpose**

- Determine the completeness of birthplace before and after linkage
- Determine the rate of agreement
- Discuss how the characteristics of patients differ between those who have birthplace in both files versus those missing birthplace from the CCR database
- Enhance completeness of the statewide database through linkage to death certificate birthplace data

**Methods**

Our approach involves conducting a death record re-linkage of cancer patients who expired in 2014 and 2015 in California. The purpose of the initial linkage of death certificates was to update vital status information of patients reported in the database. In this re-linkage project, we attempt to enhance the database with birthplace data by evaluating a total of 70,764 cases and calculating percentage of missing/unknown birth state overall, by sex, and race before and after linkage with death certificate data. A total of 172,884 cases were evaluated to determine percentage of missing/unknown birth country by the same categories.

Birth state and birth country in the database and the death certificate were compared for matching status to determine rate of agreement, limiting the comparison to patients with available values in both the database and death certificate. Percent agreement was used to indicate confidence in the linkage reliability.

**Results**

<table>
<thead>
<tr>
<th>Race</th>
<th>CCR Before</th>
<th>CCR After</th>
<th>DC Before</th>
<th>DC After</th>
<th>CCR-DC After</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH White</td>
<td>77,103</td>
<td>77,103</td>
<td>77,103</td>
<td>77,103</td>
<td></td>
</tr>
<tr>
<td>NH API</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
<tr>
<td>NH Black</td>
<td>7,170</td>
<td>7,170</td>
<td>7,170</td>
<td>7,170</td>
<td></td>
</tr>
<tr>
<td>NH Hispanic</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
</tbody>
</table>

- No significant difference between males and females when applying death certificate (DC) data
- Non-Hispanic Whites and non-Hispanic Blacks would benefit most from a DC linkage statewide; non-Hispanic American Indian/Alaska Native counts were too few to draw conclusions
- Percentage of cases missing birth state varied across the three registries, with Cancer Prevention Institute of California yielding the greatest reduction of missing/unknown values after a DC linkage

<table>
<thead>
<tr>
<th>Race</th>
<th>CCR Before</th>
<th>CCR After</th>
<th>DC Before</th>
<th>DC After</th>
<th>CCR-DC After</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH White</td>
<td>77,103</td>
<td>77,103</td>
<td>77,103</td>
<td>77,103</td>
<td></td>
</tr>
<tr>
<td>NH API</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
<tr>
<td>NH Black</td>
<td>7,170</td>
<td>7,170</td>
<td>7,170</td>
<td>7,170</td>
<td></td>
</tr>
<tr>
<td>NH Hispanic</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td>7,572</td>
<td></td>
</tr>
</tbody>
</table>

- No significant difference between males and females when applying DC data
- Statewide, the percentage of missing birth country was greater for non-Hispanic Whites and non-Hispanic Blacks, and these groups benefited the most from the DC linkage
- Percentage of cases missing birth country varied across the three registries and was significantly greater than percentage of cases missing birth state

**Conclusions**

Our assessment of death certificate data leads us to believe that a linkage to this data source is necessary to improve completeness of birthplace data in our cancer registry database. Birthplace data is not readily available or documented in medical records at all reporting facilities and therefore is not consistently collected and submitted by the Certified Tumor Registrar when reporting the cases to the central registry. Overall, the death certificate is able to increase birth state data completeness by 1.13%-14.68% statewide. Birth country data completeness improves more significantly at rates between 31.01%-55.46%.

Percent agreement between the CCR database and death certificate data exceeds 94% across all registries when comparing birth state data and exceeds 96% when comparing known birth country values across all registries, giving a high degree of confidence that the death certificate data is valuable and should be routinely linked to the California dataset.

**Future Directions**

- Make recommendations on improving data collection of birthplace information
- Assess reliability of social security number on the death certificate to determine the value of an extended linkage

**Acknowledgements**

We would like to thank Lihua Liu, PhD, Juanjuan Zhang, and Daphne Lichtensztajn, for their expertise in analysis of the data presented here. We appreciate the support of Deepa Padmanaban for her work relinking and generating the initial files for the regional registries to perform their analysis.

**California Cancer Registries**

- California Central Cancer Registry, UC Davis IPHI, Sacramento, CA
- Cancer Surveillance Program, Los Angeles, CA
- Cancer Prevention Institute of California, Fremont, CA
- Cancer Registry of Greater California, PHI, Sacramento, CA