NCI Activities to Facilitate Residential History Research

Zaria Tatalovich, PhD
Dave Stinchcomb, MS
June 12, 2018

2018 NAACCR Annual Meeting
NCI Activities to Facilitate Residential History Research

Pilot Study – Residential History Data Sources
Pilot Study – Residential History Data Sources

- Environmental scan of possible sources
- Identified three commercial data providers
  - Data collected mainly for credit checks and background investigations
- Conducted a quality review
  - Collected self-reported residential histories from NCI and NIEHS volunteers
    - Total of 66 residential histories
    - Including 10 for deceased relatives
  - Submitted names and other identifiers to the 3 vendors
  - Compared the results to the self-reported data
Total Addresses per Year

- Vendor data start in the 1980s
- Vendor 1 has too many addresses
  - Twice as many in the 1990s
  - Many duplicates
  - Recent drop may be due to data quality improvements
- The Vendor 2 data
  - Increases more gradually starting in the 1980s
  - Has fewer addresses
- The Vendor 3 data
  - Starts around 1990
  - Also has duplicates
Constructing Residential Histories

- Algorithm to derive a residential history from vendor address information
  - Resolve duplicates and conflicts
  - Fill time gaps
- Basic steps:
  1. Match addresses within and across vendors
  2. Combine matched addresses from all of the vendors
  3. Decide on a time-frame for each address
     - Optionally trim time-frame based on vendor frequencies
  4. Weed out short duration addresses
  5. Build residential history working backwards from the most recent address
     - Use current start date as end date for the previous address
## Full Life Span Comparisons by Vendor Combo

<table>
<thead>
<tr>
<th>Vendors used</th>
<th>Percent time period coverage</th>
<th>Percent of covered time distance=0</th>
<th>Percent of covered time within 1 km</th>
<th>Percent of covered time within 5 km</th>
<th>Percent of covered time within 10 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>58.7</td>
<td>56.0</td>
<td>68.1</td>
<td>74.0</td>
<td>81.3</td>
</tr>
<tr>
<td>V2</td>
<td>35.4</td>
<td>49.1</td>
<td>56.7</td>
<td>70.0</td>
<td>76.3</td>
</tr>
<tr>
<td>V3</td>
<td>35.5</td>
<td>31.6</td>
<td>77.0</td>
<td>84.0</td>
<td>90.3</td>
</tr>
<tr>
<td>V1,V2</td>
<td>58.9</td>
<td>50.0</td>
<td>64.3</td>
<td>71.0</td>
<td>78.5</td>
</tr>
<tr>
<td>V1,V3</td>
<td>58.8</td>
<td>25.2</td>
<td>65.9</td>
<td>73.1</td>
<td>80.3</td>
</tr>
<tr>
<td>V2,V3</td>
<td>46.9</td>
<td>25.2</td>
<td>68.1</td>
<td>78.1</td>
<td>84.0</td>
</tr>
<tr>
<td>V1,V2,V3</td>
<td>59.1</td>
<td>28.4</td>
<td>64.2</td>
<td>72.4</td>
<td>80.1</td>
</tr>
<tr>
<td>Cur Res</td>
<td>NA</td>
<td>23.0</td>
<td>27.6</td>
<td>32.5</td>
<td>37.4</td>
</tr>
</tbody>
</table>
Summary of Results

- Commercial data can be used to generate residential histories for cases in cancer registries
  - An algorithm is needed to process vendor address data
  - Data is available for deceased individuals

- Residential histories derived from commercial data are reasonably complete and accurate
  - Substantially better than assuming that people never move
  - Data start in the mid 1980s
  - Primarily just U.S. addresses
  - Currently, Vendor 1 has the most complete and accurate data
  - Including data from Vendors 2 and 3 did not improve the results
NCI Activities to Facilitate Residential History Research

Funding Opportunity Announcement
Pilot Raised Methodological Questions

- Obtaining residential histories for specific populations:
  - with little credit or banking history
  - older individuals whose histories extend prior to the electronic data collection
- Obtaining automated residential histories for specific geographic areas
- Algorithms for optimizing the consolidation of conflicting addresses obtained through various data sources
- Refinements to methodologies for linkage to residential histories
Supplements for Validating the Use of Automated Sources of Residential Histories in Cancer Epidemiology Cohorts

(Administrative Supplement)

For the full text visit


Research Objective

• Ideally, studies would involve existing cohorts where high quality residential histories which had been obtained by interview or other means could be used as a “gold standard” for comparison to those obtained by automated processes
Integration of Individual Residential Histories into Cancer Research R01


Research Objective

• Encourage substantive investigation of the role of individual residential histories in the context of cancer research

Multidisciplinary Expertise

• This FOA is expected to engage scientists from a range of disciplines, including epidemiology, statistics, geospatial science, computer science, demography, decision science, and behavioral science, to effectively integrate individual residential histories into cancer research.
Research Questions

• Can residential history information help provide a more complete representation of the *exposome* of individuals?

• How do changes in place of residence affect probability of receiving cancer-related vaccines, chemoprevention, screening, treatment at critical times in one’s life?

• What is the relationship between geographic indicators of poverty and deprivation gleaned from residential histories and cancer risk and outcome
Integration of Individual Residential Histories into Cancer Research R21


Research Objective

• Develop new analytical methods for the use of individual residential histories in substantive research
Research Questions

• What is the best strategy to consolidate and reconcile conflicting residential history records in the epidemiological study of cancer risk and outcomes?

• What is the impact of varying amounts of geographic misspecification of residential histories in different types of epidemiological studies of cancer?

• What statistical or other analytic strategies can be developed to understand the relationships between exposures at various residences (including the length and intensity of the exposures) and the latency period between exposure and disease?
Facilitating Residential History for Research

• Researchers may elect to develop **novel methodologies** for collecting residential histories.

• Researchers may elect to follow the **standardized process**: using vendor supplied address information along with an **open-source algorithm and software** for consolidating residential histories - available from NCI website [https://gis.cancer.gov/tools/residential-histories.html](https://gis.cancer.gov/tools/residential-histories.html)