The use of residential histories in geospatial research of cancer: The Multiethnic Cohort Study

Iona Cheng
UCSF Department of Epidemiology & Biostatistics
June 12, 2018
Multiethnic Cohort

Purpose:

To examine lifestyle risk factors, especially diet and nutrition, as well as genetic susceptibility in relation to the causation of cancer.

- Funded by NCI in 1993
- Population-based study of > 215,000 men & women
  - Residents of Los Angeles and Hawaii
  - African-Americans, Japanese Americans, Latinos, Native Hawaiians, & Whites
- Ages 45-75 at enrollment
Data Collection

- 1993-1996: Baseline Questionnaire
- 1998-2000: Brief Follow-up Questionnaire
- 2003-2008: Repeat of Baseline Questionnaire
- 1996-2006: Biological Specimens
- 2010-2013: Brief Follow-up Questionnaire
- 2018-2020: Brief Follow-up Questionnaire

Self-reported Residential Addresses
Surveillance/Retention

- Passive Follow-up for cancer incidence & disease conditions
  - SEER Cancer Registries in HI and CA
  - National Death Index
  - Medicare

- Active Follow-up
  - Interval questionnaires
  - Annual newsletters (retention)
Geocoding

Use GIS software (ArcGIS, ESRI) and best available locator data (i.e., tax parcel and street centerline reference data) from government agencies and private vendors.

Of 337,141 residential addresses geocoded:
- 72% building or parcel
- 25% street segment
- 5% city, zip code, state
- <1% could not be geocoded

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Address</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>100004</td>
<td>342 Royal Dr</td>
<td>Los Angeles</td>
<td>CA</td>
<td>90010</td>
</tr>
</tbody>
</table>
Residential Address Data (1993-2016)

- 332,247 geocoded address records
- Average duration per address = 11 years
- Range in address duration = 10 years for Whites to 13 years for Japanese Americans
MEC participants: Los Angeles County

- African American
- Japanese
- Latino
- White

MEC participants: Oahu

- Native Hawaiian
- Japanese
- White

Leveraging geocodes & linkage for studies of cancer

- Neighborhood obesogenic environment & breast, colorectal, and prostate cancers
  - California Neighborhoods Data System: social & built environments
- Air pollution & breast cancer, lung cancer, COPD, benign and malignant brain cancer
  - Kriging interpolation: pollutants from regional sources
  - Land use regression model: pollutants from regional & local sources
  - Air pollution dispersion model: pollutants from local sources
- Opportunities for additional & historical addresses through commercial vendors
  - LexisNexis linkage
Approach: Multi-level & Integrative Data

- **Residential History & Geocodes**
- **Secondary data sources: CA Neighborhoods Data System, air monitoring & modeling data**
- **Questionnaire data: baseline & follow-up**
- **Biological data: GWAS, biomarkers**
- **Surveillance: Cancer and disease conditions**
Findings in the MEC

Contextual Impact of Neighborhood Obesogenic Factors on Postmenopausal Breast Cancer: The Multiethnic Cohort

Shannon M. Conroy, Christina A. Clarke, Juan Yang, Salma Shariff-Marco, Yurii B. Shvetsov, Song-Yi Park, Cheryl L. Albright, Andrew Hertz, Kristine R. Monroe, Laurence N. Kolonel, Loïc Le Marchand, Lynne R. Wilkens, Scarlett Lin Gomez, and Iona Cheng
Neighborhood Change

- **Two types of Neighborhood Change:**
  - **Movers:** Move to a different neighborhood environment
  - **Non-movers:** Change in the neighborhood environment

- **Measuring change:**
  - From baseline to current address (e.g. average, time-varying exposure)
  - Shift between categories of a neighborhood attribute (e.g., increase, decrease, no change)
  - Neighborhood trajectories across follow-up
Change in nSES and Prostate Cancer Incidence

Positive change
- All: HR & 95% CI = 0.96
- African American: HR & 95% CI = 0.92
- Japanese American: HR & 95% CI = 1.02
- Latino: HR & 95% CI = 1.18
- White: HR & 95% CI = 0.89

Negative change
- All: HR & 95% CI = 0.89
- African American: HR & 95% CI = 0.84
- Japanese American: HR & 95% CI = 1.05
- Latino: HR & 95% CI = 0.83
- White: HR & 95% CI = 1.01

Non-Movers: HR & 95% CI
- All: HR & 95% CI = 0.95
- African American: HR & 95% CI = 0.97
- Japanese American: HR & 95% CI = 0.84
- Latino: HR & 95% CI = 0.94
- White: HR & 95% CI = 0.82

Movers: HR & 95% CI
- All: HR & 95% CI = 0.95
- African American: HR & 95% CI = 0.87
- Japanese American: HR & 95% CI = 1.70
- Latino: HR & 95% CI = 0.88
- White: HR & 95% CI = 1.18

UCSF
Opportunities with Commercially-Obtained Address

- LexisNexis®, electronic database, that includes more than 65 billion records from more than 10,000 sources such as public, private, regulated, and derived data (~1980-present)

- NCI/Westat pilot study developed an algorithm that converts vendor-supplied set of addresses into a sequential residential history (Stinchcomb & Roeser, 2016).

Questions:

- What is the accuracy of commercially obtained addresses among diverse racial/ethnic groups, older age groups, low SES groups?

- Could there be misclassification of long-term geospatial exposures (relevant for cancer development) and/or short-term exposures (windows of susceptibility)?
Validating Residential Histories in Ethnically & Socioeconomically Diverse Populations

**Aim 1:** To determine the agreement of residential histories obtained by a commercial vendor in comparison to collected prospective residential histories of MEC participants.

**Aim 2:** To evaluate whether the agreement of commercially obtained residential histories in comparison to MEC collected residential histories differs by demographic and socioeconomic factors, the availability of linkage variables, and time periods.

**Aim 3:** To compare estimates of long- and short-term geospatial exposures based on residential histories collected by a commercial vendor and the MEC.
Future Directions

- A life-course perspective
- Integration of secondary (objective) and self-report (perception) neighborhood data
- Neighborhood environments at work
- Intersectional, interactions among multiple social determinants of health
- Integration with biological measures
  - Stress biomarkers, inflammation, disease biomarkers, epigenetics, genetic susceptibility
- Integration with policy measures
- Analytical approaches and tools, including multilevel, geostatistical/geospatial tools
- Causal inference methods
Study Teams & Funding

UCSF: Iona Cheng, Juan Yang, Jennifer Jain, Shannon Conroy, Mindy DeRouen, Sarah Li, Pushkar Inamdar, Salma Shariff-Marco, Scarlett Lin Gomez

UH: Yurii Shvetsov, Cheryl Albright, Loic Le Marchand, Lynne Wilkens

USC: Anna Wu, Chiuchen Tseng, Dan Stram, Wendy Setiawan, Lani Park, Lianfi Li, Florence Hoffman, Thomas Chen, Scott Fruin

UCI: Jun Wu, Shahir Masri

UCLA: Beate Ritz

U of WA: Timothy Larson

Stanford: Alice Whittemore

U of CO: Jon Samet

UT SW: Sandi Pruitt

Westat: David Stinchcomb

Funders

- NCI R01 CA154644
- NCI P01 CA138338
- NCI U01 CA164973
- NCI R03 CA117324
- SEER N01-PC-35136
- NIEHS R01ES026171
- Susan G. Komen Foundation IIR13262718
- Brain Lung Tumor Air Pollution Foundation 85837591
Thank You
“Asking a patient “where do you live?” could be used to identify geographic exposures associated with cancer risk and outcomes.”
MEC Southern California: Baseline Addresses

African Americans (n=33,247)

Latinos (n=45,263)

Japanese Americans (n=12,385)

Whites (n=14,330)
Number of Businesses and Obesity

Conroy, et al. Cancer Causes Control, 2018
nSES and Breast Cancer Risk

• No association with four components of the neighborhood obesogenic index: Urban, Unhealthy Food, Mixed Land Use, Parks

• Independent association between nSES and BC, adjusting for demographics, breast cancer risk factors, all obesogenic indices variables

<table>
<thead>
<tr>
<th>nSES</th>
<th>All (cases=2341)</th>
<th>ER+ or PR+ (cases=1542)</th>
<th>ER- and PR- (cases=361)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5-High</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Q4</td>
<td>0.92 (0.79-1.07)</td>
<td>0.88 (0.74-1.05)</td>
<td>0.91 (0.62-1.35)</td>
</tr>
<tr>
<td>Q3</td>
<td>0.88 (0.75-1.02)</td>
<td>0.81 (0.68-0.97)</td>
<td>0.70 (0.47-1.06)</td>
</tr>
<tr>
<td>Q2</td>
<td>0.89 (0.75-1.04)</td>
<td>0.81 (0.67-0.98)</td>
<td>1.02 (0.67-1.56)</td>
</tr>
<tr>
<td>Q1-Low</td>
<td><strong>0.79 (0.65-0.95)</strong></td>
<td><strong>0.73 (0.58-0.91)</strong></td>
<td><strong>0.91 (0.62-1.35)</strong></td>
</tr>
<tr>
<td>P trend</td>
<td><strong>0.02</strong></td>
<td><strong>0.01</strong></td>
<td><strong>0.55</strong></td>
</tr>
</tbody>
</table>

Model 1 demographics, smoking, alcohol, diet, menopause, comorbidities, NSAD use, neighborhood obesogenic factors, clustering by block group
Model 2 also adjusted for BMI
nSES and Breast Cancer Risk

- No association with four components of the neighborhood obesogenic index: Urban, Unhealthy Food, Mixed Land Use, Parks
- Independent association between nSES and BC, adjusting for demographics, breast cancer risk factors, all obesogenic indices variables

<table>
<thead>
<tr>
<th>nSES</th>
<th>All (cases=2341)</th>
<th>ER+ or PR+ (cases=1542)</th>
<th>ER- and PR- (cases=361)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5-High</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Q4</td>
<td>0.92 (0.79-1.07)</td>
<td>0.88 (0.74-1.05)</td>
<td>0.91 (0.62-1.35)</td>
</tr>
<tr>
<td>Q3</td>
<td>0.88 (0.75-1.02)</td>
<td>0.81 (0.68-0.97)</td>
<td>0.70 (0.47-1.06)</td>
</tr>
<tr>
<td>Q2</td>
<td>0.89 (0.75-1.04)</td>
<td>0.81 (0.67-0.98)</td>
<td>1.02 (0.67-1.56)</td>
</tr>
<tr>
<td>Q1-Low</td>
<td>0.79 (0.65-0.95)</td>
<td>0.73 (0.58-0.91)</td>
<td>0.91 (0.62-1.35)</td>
</tr>
<tr>
<td>P trend</td>
<td>0.02</td>
<td>0.01</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>African Americans</td>
<td>Japanese Americans</td>
<td>Latinos</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Kriging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM$_{10}$</td>
<td>1.02</td>
<td>1.26</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM$_{2.5}$</td>
<td>1.04</td>
<td>1.54</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_x$</td>
<td>1.19</td>
<td>1.28</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LUR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_x$</td>
<td>1.26 (p=0.04)</td>
<td>1.42 (p=0.02)</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CALINE4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_x$</td>
<td>1.02</td>
<td>1.15</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Kriging PM$_{10}$ & PM$_{2.5}$ per 10ug/m$^3$, LUR & Kriging NO$_x$ per 50 ppb, CALINE4 NO$_x$ 10 ppb*