Validating the Dasymetric Areal Interpolation Method to Inform Health Policy

Chieko Maene, MS
Social Sciences Computing Services, University of Chicago

Monica E. Peek, MD, MPH, MSc
Section of General Internal Medicine, University of Chicago
Chicago Center for Diabetes Translation Research
County-level Estimates of Diagnosed Diabetes among Adults aged ≥ 20 years: United States 2007

Percent

0 - 6.5
6.6 - 8.0
8.1 - 9.4
9.5 - 11.1
> 11.2
County-level Estimates of Diagnosed Diabetes among Adults aged ≥ 20 years:
United States 2008

Percent

0 - 6.5
6.6 - 8.0
8.1 - 9.4
9.5 - 11.1
> 11.2

www.cdc.gov/diabetes
Diabetes mortality in Chicago
Potential years of life lost in Chicago
Avoidable diabetes-related hospitalizations
Health Information Exchanges

CHITREC Presents the LAUNCH Program

The Chicago Health Information Technology Regional Extension Center (CHITREC) is proud to announce the creation of LAUNCH, a program designed to support healthcare providers as they strive to reach health IT goals, including EHR adoption and Meaningful Use attestation.

LAUNCH offers comprehensive services and personalized education in assessment and planning, electronic health record (EHR) selection, Meaningful Use education and gap analysis, workflow redesign, incentive program registration and attestation, privacy and security guidance, and other customized services based on your needs.

Preparing to Attest?
CHITREC offers webinars to help you with the Medicaid EHR incentive program. Learn how to prepare for attestation and get a sneak preview of the attestation system.

Illinois Medicaid EHR Incentive Help Desk
Contact us for Attestation, Registration, and Meaningful Use answers.
855-MU-HELP-1
(855-684-3571)
hfs.ehrintensive@illinois.gov
Monday-Friday, 8:30am – 5:00pm

January Quiz
Do you know what meaningful use changes are coming in 2014?
CommunityRx: HealtheRx

The South Side is talking about HealtheRx

Patients and Neighbors

The community expert will know where to send me.

Because these places are all located near me, they'll be easy to get to.

The HealtheRx will be helpful between doctor visits to know where services are in the community.

What is HealtheRx?

- It is a true community partnership and a solution that benefits everyone. Together, we can significantly improve health, health care, and strengthen our communities at the same time.
- Doriane Miller, MD
  Associate Professor of Medicine
  Director, Center for Community Health and Vitality

Local Health Providers

As a doctor who treats patients on the South Side every day, I need HealtheRx. This new kind of ‘prescription’ goes beyond a diagnosis and medicine. It provides personalized information and support from community resource specialists to help patients stay healthy between clinic visits.

Tim Long, MD
Physician, Kamed Holman Health Center

- Fill your prescription! Lose weight! Eat healthier! Stop Smoking! All day long, we tell patients what we think they should do to be healthier, but we fail to make the connections to places and services they can use to stay well, live independently, and manage with disease. HealtheRx is the connection between health care and self-care.
- Stacy Lindau, MD, MAPP
  Associate Professor of Ob/Gyn and Medicine-Geriatrics
  Project Director, CommunityRx

For more information call (773) 834-2356 or visit www.healtherx.org

MAPS Corps and HealtheRx are innovations from CommunityRx, a flagship program of the South Side Health and Vitality Studies at the University of Chicago Medicine's Urban Health Initiative. CommunityRx is supported by grant #HHS09330097-02-00 from the Department of Health and Human Services, Centers for Medicare and Medicaid Services. Its contents are solely the responsibility of the authors and have not been approved by the Department of Health and Human Services, Centers for Medicare and Medicaid Services.

Stacy Lindau, MD, MA

Addressing Diabetes Disparities
Food Rx: Farmer’s Market partnership
Food Rx: Farmer’s Market partnership
Challenge

- Leveraging “big data” at zipcode level
Challenge

• Leveraging “big data” at zipcode level

• Meaningful at community level
Solution

• Leveraging “big data” at zipcode level

  Dasymetric Areal Interpolation

• Meaningful at community level
Chicago Public Health Department Collaboration

- Problem
- Methods
- Validation
- Conclusions
Challenge: “Modifiable Areal Unit Problem”

– Context: Public Health indicators in Chicago

– Research Question: What is the community-level variation in diabetes-related hospitalizations?

– Trial of dasymetric areal interpolation method
MAUP: Same Total, Different Aggregates

- **MAUP**
  - Modifiable Areal Unit Problem

- Interpretation of results can change depending on the choice of boundary

Total N=24
MAUP:
Same Total, Different Aggregates

• **MAUP**
  – Modifiable Areal Unit Problem

• Interpretation of results can change depending on the choice of boundary

Total N=24
Starting point
Diabetes hospitalization rates by ZIP

Convert to community area
ZIP Code & Community Areas

ZIP Codes (N=59)

Community Areas (N=77)
**ZIP Code & CA are similar**

<table>
<thead>
<tr>
<th></th>
<th>ZIP Codes (ZCTAs) (N=59)</th>
<th>Community Areas (N=77)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Area (square miles)</strong></td>
<td>0.09 – 16.60</td>
<td>0.61 – 13.34</td>
</tr>
<tr>
<td></td>
<td>4.43 77.8</td>
<td>3.00 65.7</td>
</tr>
<tr>
<td><strong>2010 Population</strong></td>
<td>493 – 133,916</td>
<td>2,876 – 98,514</td>
</tr>
<tr>
<td></td>
<td>6 47,143 56.5</td>
<td>35,008 63.9</td>
</tr>
</tbody>
</table>

*CV (Coefficient of variation) = the smaller, the less variability
Sources: 2010 US Census, the City of Chicago
Blocks are constituents of both Community Areas and ZIP Codes
ZIP Codes over Hyde Park
Why Dasymetric Areal Interpolation?

• **Geographical unit discrepancies:**
  – In Illinois, hospital discharge data comes with 5-digit ZIP Code only as patient address.
  – In Chicago, summary statistics are tabulated by Community Areas (i.e. Chicago neighborhoods).

• **Our solution:**
  – Estimate community level hospital discharge rate by allocating the # of discharges of a given ZIP Code to overlapping communities based on proportions of population and by ancillary information*.

* Ancillary information used was gender, race and age group
Dasymetric Interpolation Procedure (Stage 1)

1. Calculate for each ZIP code: male & female x 19 age groups x 4 race-ethnicity groups = 84 age-sex-race-specific rates

2. Apply rates to corresponding population for age-sex-race group in each census block to get case counts

3. Sum counts for each community area by age group

4. Calculate crude and adjusted rates
Results

Rates by ZIP

Rates by community area
# Results

## Actual vs. Interpolated

<table>
<thead>
<tr>
<th></th>
<th>ZIP Codes (N=58)</th>
<th>Community Areas (N=77)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Median</td>
</tr>
<tr>
<td>Discharges</td>
<td>0 – 393</td>
<td>109</td>
</tr>
<tr>
<td>Crude rate</td>
<td>0 – 57</td>
<td>18</td>
</tr>
<tr>
<td>Adjusted rate* (per 10,000)</td>
<td>0 – 57</td>
<td>19</td>
</tr>
</tbody>
</table>

Methodological Validation

- Need for validation study using case data geocoded to community area

Plan:
- Obtain raw hospital discharges from a single Chicago hospital (UCM)
- Compared the actual discharge rates to the estimated discharge rates from the dasymetric areal interpolation method

Evaluation measure:
- Statistical significance test:
  - One-tailed Chi-square test \( (p < 0.05) \)
Data: Validation Study

1. Univ. of Chicago Medicine Hospital Discharges
   - 84,942 cases
     • All discharges (patients), not just diabetes
     • From all departments within the UC Medical Center
   - Date range (1/1/2009-12/31/2011)
     • By discharge date
     • 2009 (n=29,239), 2010 (n=27,649), 2011 (n=28,054)
   - Variables
     • Location: residential street address
     • Demographic: DOB (age), sex, race, ethnicity
     • ICD-9 Diagnosis code (up to 10 per discharge)
2. 2010 Census TIGER/Line Shapefiles
   - http://www.census.gov/geo/maps-data/data/data/tiger.html
   - Census blocks
   - Census ZCTAs (ZIP Code Tabulation Areas)

3. 2010 Census summary file 1 (block-level, i.e. sumlevel=101)
   - http://www2.census.gov/census_2010/04-Summary_File_1/
   - Sex by age (P012), sex by age by race (P012H&I)
     • Race: NH White, Hispanic, others (= Total – NHW – Hisp.)

4. Chicago community area boundary file
   - https://data.cityofchicago.org/
Identified Data Issues

1. UC Medical Center discharge data
   - Missing values
     • age & sex (0.1%) -> excluded
     • race/ethnicity (13.9%) -> treated as “others” category
     • DX (1.4%) -> excluded

2. 2010 Census TIGER/Line Shapefiles
   - ZCTAs (ZIP Code Tabulation Areas) are generalized ZIP Code zones. They may include addresses associated with ZIP Codes that are not the same as the ZCTA.

3. 2010 Census summary file 1 (block-level)
   - Inability to identify “NH African-American” (47% of discharges)
   - Swapping (statistical disclosure avoidance technique)
     • “A small sample of households” “were swapped with data from other households that had identical characteristics on a certain set of variables but were from different geographic locations.”
UCM Diabetes Discharges

• Patient addresses were geocoded:
  – Software/data: ArcGIS 10.2/ESRI StreeMap Premium
  – 98.3% were geocoded at street address level.
  – Chicago residency, ZIP Code & Community Area were determined based on geocoded location.

• Diagnosis code selection:
  – Diabetes (ICD-9 250.x) discharges only.

• Result:
  – Chicago diabetes discharges with valid variables.
  – Total 6,534 discharges.
## UCM Diabetes Patients: Sociodemographic Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total (n=6,534)</th>
<th>NH White (n=461)</th>
<th>Hispanic (n=287)</th>
<th>Others (n=5,786)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All (%)</strong></td>
<td>100%</td>
<td>7%</td>
<td>4%</td>
<td>89%</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>2,581</td>
<td>279</td>
<td>167</td>
<td>2,135</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>3,953</td>
<td>182</td>
<td>120</td>
<td>3,651</td>
</tr>
<tr>
<td><strong>Age (mean, SD)</strong></td>
<td>61 (17)</td>
<td>64 (16)</td>
<td>62 (16)</td>
<td>61 (17)</td>
</tr>
<tr>
<td><strong>Distance Miles (mean, SD)</strong></td>
<td>3.83 (3.08)</td>
<td>7.3 (4.84)</td>
<td>7.14 (2.98)</td>
<td>3.39 (2.59)</td>
</tr>
<tr>
<td><strong>Length of Stay (mean, SD)</strong></td>
<td>4.33 (5.27)</td>
<td>4.44 (5.37)</td>
<td>4.18 (4.48)</td>
<td>4.33 (5.29)</td>
</tr>
</tbody>
</table>

Note: Patients might be repeated if they were hospitalized more than once, as our unit of analysis is a discharge, not a patient. Distance is a direct distance between patients’ residence and the UC Medical Center and measured in miles.
Diabetes Discharges Results

- Number of diabetes discharges by Chicago community areas.
- N=6,534
- Color in quintiles
- Patients are mostly from the Southside.
Interpolation: Step 1

• Start with a Census block data table with population & discharge counts by race, sex and age group

• Calculate ZIP Code level discharge rate by race, sex and age group.
  – Rate (weight) = Discharge # / Population for the ZIP/age/race/sex group

<table>
<thead>
<tr>
<th>ZIP Code</th>
<th>Race</th>
<th>Sex</th>
<th>Age group 1</th>
<th>Discharges</th>
<th>Population</th>
<th>Rate (weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60637</td>
<td>NH White</td>
<td>Female</td>
<td>35-44</td>
<td>33</td>
<td>2517</td>
<td>0.01311</td>
</tr>
</tbody>
</table>
Interpolation: Step 2

• Transfer the ZIP Code level discharge rate to the Census block table to estimate (interpolate) the number of discharges at the census block level by race, sex and age group.
  – Dasymetric count = Rate (weight) x Population

<table>
<thead>
<tr>
<th>ZIP Code</th>
<th>Block Code</th>
<th>Community</th>
<th>Race</th>
<th>Sex</th>
<th>Age group 1</th>
<th>Population</th>
<th>Rate (weight)</th>
<th>Dasymetric count</th>
</tr>
</thead>
<tbody>
<tr>
<td>60637</td>
<td>0364001007</td>
<td>Hyde Park</td>
<td>NH</td>
<td>Female</td>
<td>35-44</td>
<td>3</td>
<td>0.01311</td>
<td>0.03933</td>
</tr>
</tbody>
</table>
Interpolation: Step 3

- Aggregate the block level estimated discharges for all population (i.e. removing race, sex and age categories, except for four age categories for the subsequent age-adjustment) at a community level.

- Calculate crude discharge rates
  - Rate = Dasymetric count / Population
  - (Need to be age-adjusted next!)

<table>
<thead>
<tr>
<th>Community</th>
<th>Age group 2</th>
<th>Population</th>
<th>Dasymetric count</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyde Park</td>
<td>0-44</td>
<td>10168</td>
<td>295</td>
<td>0.02901</td>
</tr>
</tbody>
</table>
Interpolation: Step 4

- Adjust the crude discharge rates for age using the U.S. standard population population

<table>
<thead>
<tr>
<th>Community</th>
<th>Age-adjusted rate per 10,000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyde Park</td>
<td>47.26</td>
</tr>
</tbody>
</table>
Actual (L) vs. Interpolated (R) Hospital Discharge Rates

2010 Diabetes-related Discharge Age-adjusted Rate (4 groups) (Geocoded Location)

2010 Diabetes-related Discharge Age-adjusted Rate (4 groups) (Dysmetric By Race/Sex/Age)
Results: Differences

Rate Differences

Rate Differences (outside 95% CI)
Validation Results

• An estimated 6,544 hospitalizations were calculated using the dasymetric method, for a difference of 10 persons.
  – Raw N=6,534

• Variation in actual vs. estimated discharge rates by neighborhoods were not statistically significant, $X^2(76, N=6,534) = 54, p=0.97$. 
Conclusions

• Dasymetric Areal Interpolation an effective, validated approach

• Translate zipcode-level data to community-level data

• Inform local health policy and population health management
Acknowledgements

Co-Authors

• Roderick Jones, PhD, MPH, Northwestern University
• Emily Laflamme, MPH, Chicago Dept of Public Health
• Adam Pah, PhD, Northwestern University
• Nyahne Bergeron, MPH, University of Chicago
• Elbert S. Huang, MD, MPH, University of Chicago

Funding

• Chicago Center for Diabetes Translation Research (NIDDK P30 DK092949)
Supplemental Discussion

• In the racially diverse city, like Chicago, inclusion of race into calculation turned out to be crucial.

• To prove if inclusion of race in the method makes a difference we ran the same method using age and sex categories only.

• A measure we used for the comparison is root mean squared errors (RMSE) – see the next table.
## Results (supplement)

<table>
<thead>
<tr>
<th></th>
<th>Community Areas (N=77)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Range</td>
<td>RMSE</td>
</tr>
<tr>
<td>Community age-adjusted rates</td>
<td></td>
<td>0.10-84.68</td>
<td>(N/A)</td>
</tr>
<tr>
<td>Dasymetric (age, sex, race)</td>
<td></td>
<td>0.39-74.63</td>
<td>2.66</td>
</tr>
<tr>
<td>Dasymetric (age, sex)</td>
<td></td>
<td>0.38-67.23</td>
<td>4.22</td>
</tr>
</tbody>
</table>

* * The number of communities whose estimated rates are statistically different from the actual/observed rates at 95% confidence level.
Supplement: How to “group” age
Error comparisons from different age group uses

![Graph showing root-mean squared error (RMSE) and count difference](image-url)
Supplement: How to “group” age

<table>
<thead>
<tr>
<th>Number of categories</th>
<th>Minimum Interval</th>
<th>Categories</th>
<th>RMSE</th>
<th>Count difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>5</td>
<td>0, 5, 10, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85+</td>
<td>3.687</td>
<td>207</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85+</td>
<td>3.689</td>
<td>207</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80+</td>
<td>3.680</td>
<td>205</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75+</td>
<td>2.660</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>0, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75+</td>
<td>2.660</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>0, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75+</td>
<td>2.656</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0, 5, 10, 15, 25, 35, 45, 55, 65, 70, 75, 85+</td>
<td>2.700</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>0, 18, 25, 35, 45, 55, 65, 70, 75, 85+</td>
<td>2.696</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0, 18, 30, 45, 55, 65, 75, 85+</td>
<td>2.689</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0, 18, 30, 45, 55, 65, 75+</td>
<td>2.675</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0, 18, 30, 45, 65, 75+</td>
<td>2.685</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0, 18, 45, 65, 75+</td>
<td>2.682</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>0, 45, 65, 75+</td>
<td>2.716</td>
<td>11</td>
</tr>
</tbody>
</table>