Importance of Addressing Social Determinants of Health in Achieving Health Equity

Sherita Hill Golden, MD, MHS
Hugh P. McCormick Family Professor of Endocrinology and Metabolism
Vice President & Chief Diversity Officer
Establishment of Task Force on Black and Minority Health 1983-1984

Association of Minority Health Professions School’s study “Blacks and the Health Professions in the 1980s: A National Crisis and a Time for Action”

- Dr. Louis Sullivan, Dean and President, Morehouse School of Medicine
- Dr. David Satcher, President of Meharry Medical College
- Dr. Walter Bowie, Dean of Tuskegee University
- Dr. M. Alfred Haynes, Dean of Charles R. Drew University of Medicine and Science

Shortage of minorities in health professions and health disparities among blacks

Secretary of the U.S. Department of Health and Human Services (1983-1985)
Heckler Report on Black and Minority Health 1985

- Health disparities accounted for 60,000 excess deaths each year
- Six causes of death accounted for more than 80% of mortality among Blacks and other minority populations
  - Cancer
  - Cardiovascular Disease
  - Diabetes (33% higher in Black compared to White population)
  - Homicide
- Outlined:
  - Recommendations to reduce health disparities
  - Need to improve data collection among Hispanics, Asian Americans, and American Indian/Alaska Natives
Additional Heckler Report Outcomes

• 1986—HHS Office of Minority Health
  – Led first by Dr. Herbert Nickens

• Other outcomes
  – Pivotal legislation, funding, policies, research and initiatives focused on minority health and health equity
  – NIH Office of Minority Health
  – Centers for Disease Control and Prevention
  – Health Resource and Services Administration
  – More inclusive data collection techniques
  – Establishment of institutions, centers, commissions, and state and local minority health offices

Gracia JD, *Health Affairs*, 2018
IOM report: “Unequal Treatment”
- Examine health system, provider, and patient factors
- Ethnic minorities → less access to preventive care, treatment and surgery → delayed diagnosis, advanced disease
- Persistence of race/ethnic disparities in health and healthcare
Of all the forms of inequality, injustice in health care is the most shocking and inhumane.

How Did We Get Here?

Historical Medical and Scientific Context of Health Disparities
HISTORICAL DISCRIMINATION AND RACISM DURING SLAVERY AND POST-CIVIL WAR

Medical and Scientific Contributors
• Eugenics Theory defining certain races and ethnicities as biologically inferior
• Closure of medical schools training black physicians in 1910s
• Experimentation on vulnerable groups without their consent

Dr. J. Marion Sims
• Experimented and repeatedly performed gynecological procedures on slave women without anesthesia
• Consent from owner, not slave
• Rewarded with international acclaim and presidency of American Medical Association

Healthcare Context
Poor access to care, ↓ quality of care, ↓ participatory decision-making in patient-provider relationships, ↓ health literacy

Golden et al. J Clin Endocrinol Metab, 2021
How Did We Get Here?

Historical Social Context of Health Disparities
HISTORICAL DISCRIMINATION AND RACISM DURING SLAVERY AND POST-CIVIL WAR

Medical and Scientific Contributors
- Eugenics Theory defining certain races and ethnicities as biologically inferior
- Closure of medical schools training black physicians in 1910s
- Experimentation on vulnerable groups without their consent

Trust in medical establishment

Healthcare provider bias toward minority patients

Language and communication barriers

Healthcare Context
- Poor access to care, ↓ quality of care, ↓ participatory decision-making in patient-provider relationships, ↓ health literacy

Social Conditions and Policies
- Redlining and predatory lending leading to racial residential segregation and housing insecurity
- Inadequate investment to maintain public works and school systems in minority neighborhoods
- Discrimination in access to high quality jobs with adequate health insurance

Federal housing loans refused to millions of Black, Asian, Hispanic, Jewish, and immigrant families

Golden et al. J Clin Endocrinol Metab, 2021
Redlining Baltimore City

Baltimore City 1910 Housing Ordinance:
“…to compel by law the separation of the white and black races in their places of residence; to prohibit the negro from intruding himself and his family as permanent residents in a district already dedicated to the white race, and equally, to present the white man from forcing himself upon a district given over to the negro.”
The Black Butterfly and the White L
HISTORICAL DISCRIMINATION AND RACISM DURING SLAVERY AND POST-CIVIL WAR

Medical and Scientific Contributors
- Eugenics Theory defining certain races and ethnicities as biologically inferior
- Closure of medical schools training black physicians in 1910s
- Experimentation on vulnerable groups without their consent

↓ Trust in medical establishment

Healthcare Context
- Poor access to care
- ↓ quality of care
- ↓ participatory decision-making in patient-provider relationships
- ↓ health literacy

↓ Healthcare provider bias toward minority patients

Language and communication barriers

↓ neighborhood stability, cleanliness, sidewalks, open space, parks

↓ access to healthy food

↓ affordable housing

Structural and institutional racism

Social Conditions and Policies
- Redlining and predatory lending leading to racial residential segregation and housing insecurity
- Inadequate investment to maintain public works and school systems in minority neighborhoods
- Discrimination in access to high quality jobs with adequate health insurance

Physical Context

Healthcare provider bias toward minority patients
HISTORICAL DISCRIMINATION AND RACISM DURING SLAVERY AND POST-CIVIL WAR

**Medical and Scientific Contributors**
- Eugenics Theory defining certain races and ethnicities as biologically inferior
- Closure of medical schools training black physicians in 1910s
- Experimentation on vulnerable groups without their consent

**Social Conditions and Policies**
- Redlining and predatory lending leading to racial residential segregation and housing insecurity
- Inadequate investment to maintain public works and school systems in minority neighborhoods
- Discrimination in access to high quality jobs with adequate health insurance

**Language and communication barriers**

**Trust in medical establishment**

**Healthcare provider bias toward minority patients**

**Healthcare Context**
- Poor access to care, ↓ quality of care, ↓ participatory decision-making in patient-provider relationships, ↓ health literacy

**Structural and institutional racism**

**Physical Context**
- ↓ neighborhood stability, cleanliness, sidewalks, open space, parks
- ↓ access to healthy food
- ↓ affordable housing

**Environmental Context**
- ↑ poverty
- Crowded multigenerational living conditions
- Mass incarceration
- ↑ service sector employment
- ↑ public transportation use

**↑ COVID-19 Exposure**

Golden et al. *J Clin Endocrinol Metab*, 2021
HISTORICAL DISCRIMINATION AND RACISM DURING SLAVERY AND POST-CIVIL WAR

Medical and Scientific Contributors
- Eugenics Theory defining certain races and ethnicities as biologically inferior
- Closure of medical schools training black physicians in 1910s
- Experimentation on vulnerable groups without their consent

Social Conditions and Policies
- Redlining and predatory lending leading to racial residential segregation and housing insecurity
- Inadequate investment to maintain public works and school systems in minority neighborhoods
- Discrimination in access to high quality jobs with adequate health insurance

↓ Trust in medical establishment
↓ healthcare provider bias toward minority patients
↓ healthcare context
- Poor access to care, ↓ quality of care, ↓ participatory decision-making in patient-provider relationships, ↓ health literacy

Language and communication barriers
↓ healthcare context
↓ stress, blood pressure, obesity, cholesterol, blood glucose, lung disease

↑ disparate health outcomes
Diabetes, Heart Disease, Hypertension, Asthma

↓ environmental context
- ↑ poverty
- Crowded multigenerational living conditions
- Mass incarceration
- ↑ service sector employment
- ↑ public transportation use
- Inadequate PPE
- ↑ COVID-19 exposure

↑ disparate health outcomes
COVID-19
Heckler Report Post-Mortem

• Recommendations focused on lifestyle choices and behaviors but not on:
  – Further desegregating medical institutions
  – Improving access to quality health care
  – Increasing investments in preventive medicine

• Focus on interventions to change behavior but not conditions

• Failed to discuss the plight of poor Whites and role of socioeconomic disadvantage in diabetes risk (e.g., Appalachia)
Social Determinants of Health Domains

- Income
- Occupation
- Housing
- Toxic Environmental Exposures
- Food Insecurity and Access
- Economic Stability
- Neighborhood and Built Environment
- Health and Health Care
- Social and Community Context
- Education
- Access
- Affordability
- Quality
- Social Cohesion
- Social Capital
- Social Support
Impact of Addressing Social Needs in Healthcare
Social Determinants of Health Interventions in Diabetes

Access
Affordability
Quality

SDOH

- Neighborhood and Built Environment
- Economic Stability
- Health and Health Care
- Education
- Social and Community Context
# Health Care Access, Affordability, and Quality

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| Access                        | ↑ insurance coverage for adults with diabetes (White, Black, Hispanic, low income, low educational attainment) in 2016  
                                  | ↑ Access to care, diabetes detection  
                                  | ↓ cost-related non-adherence in states with Medicaid expansion |
| Affordable Care Act           |                                                                          |
| Community Health Workers      | Better diabetes knowledge and self-care behaviors and improved glycemic control in underserved African American and Hispanic communities |
| Affordability                 | ↓ HbA1c and blood pressure                                               |
| Self-management interventions delivered directly to underserved |
| Quality                       | Improved quality of diabetes care among race/ethnic minorities           |
| Health IT                     |                                                                          |

Hill-Briggs et al, *Diabetes Care*, 2021
## Successful Interventions for Reducing Diabetes Health Disparities

<table>
<thead>
<tr>
<th>Level of intervention</th>
<th>Successful Components</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| Patient               | Interpersonal connections rather than computer-based  
• Face-to-face  
• Social networks  
• Family/peer support groups  
• Community health worker  
• Culturally tailored | Improved glycemic control and diabetes-related knowledge |
| Provider              | In-person feedback rather than computerized decision-support | Change in provider behavior and improved diabetes outcomes |

## Successful Interventions for Reducing Diabetes Health Disparities

<table>
<thead>
<tr>
<th>Level of intervention</th>
<th>Successful Components</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsystem/health care organization</td>
<td>Disease management • Identification of diabetes population (registries) • Practice guidelines • Health IT to track and monitor patients • Care management*</td>
<td>Improved diabetes outcomes</td>
</tr>
<tr>
<td>Community/health care system</td>
<td>• Culturally tailored patient education and empowerment • Community coalition building and advocacy • Community health workers • Provider audit and feedback • Quality improvement • Case management*</td>
<td>Improved minority health care Reduced racial and ethnic disparities in care</td>
</tr>
</tbody>
</table>

*Care management: Patient education addressing adherence barriers, ancillary services (labs), transportation*
Social Determinants of Health Interventions in Diabetes

- Housing
- Toxic Environmental Exposures
- Food Insecurity and Access

- Economic Stability
- Neighborhood and Built Environment
- Health and Health Care
- Social and Community Context
- Education

SDOH
## Physical and Food Environment

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Housing</strong></td>
<td>• Low poverty neighborhoods were associated with lower prevalence of extreme obesity and diabetes when compared to control</td>
</tr>
<tr>
<td><strong>Moving to Opportunity</strong> 4498 families randomized to 3 groups: (1) low-poverty voucher--required to use it in a census tract with a low poverty rate; (2) traditional voucher--no location restriction; (3) control --no new assistance</td>
<td></td>
</tr>
<tr>
<td><strong>Initiative to End Chronic Homelessness</strong> (placement in permanent housing and primary and mental health care)</td>
<td>↓ risk of new onset diabetes</td>
</tr>
<tr>
<td><strong>Built environment</strong> Policy to improve walkability and green space infrastructure</td>
<td>• Improved obesity</td>
</tr>
<tr>
<td><strong>Food environment</strong> Increased grocery store presence in low income neighborhoods</td>
<td>↓ Increase in diabetes prevalence</td>
</tr>
<tr>
<td>Diabetes-targeted food and self-management care at food banks and pantries</td>
<td>• Improved blood sugar, nutritional consumption, and food insecurity</td>
</tr>
</tbody>
</table>

Hill-Briggs et al, *Diabetes Care*, 2021
Solution: Advocacy and Policy
Affordable Care Act (ACA) 2010

- Increasing insurance coverage to expanding access to primary health care
- Prohibiting discrimination in health care
- Addressing diversity and cultural competency of the health care workforce
- Enhancing data collection and research
- Strengthening federal minority health infrastructure by augmenting authority of HHS Office of Minority Health
- Establishing offices of minority health within six HHS agencies
- Redesignating the National Center on Minority Health and Health Disparities to an NIH institute

Gracia JD, *Health Affairs*, 2018
Government and Community Affairs Health Disparities Briefings

- Legislative Black, Latino, and Asian Pacific Islander Caucuses
- Joint COVID Response Legislative Workgroup
- Montgomery County Delegation
- Legislative Black Caucus Town Hall
- Legislative Latino Caucus Town Hall
- House and Government Operations Committee, Maryland House of Delegates

Elizabeth Hafey, Esq
Assistant Director, State Affairs
Johns Hopkins University and Medicine

March 23, 2021
https://www.hopkinsmedicine.org/diversity/

Email: sahill@jhmi.edu
Challenges and Opportunities of Using EHR Data on Socio-Behavioral Needs in Patient Care, Research, and Population Health Management
Welcome Panelists!

- **Laura Gottlieb, MD, MPH**: Dr. Gottlieb is a Professor of Family and Community Medicine at the University of California, San Francisco. Her research explores health care sector programs and policies related to identifying and addressing social risk factors in the context of care delivery. She is the founding director of the [Social Interventions Research and Evaluation Network](#), a national research network that advances research on health care strategies to improve social conditions. Dr. Gottlieb is also Associate Director of the Robert Wood Johnson Foundation [Evidence for Action National Program Office](#).
Welcome Panelists!

- **Sarah C. DeSilvey, DNP, FNP-C:** Dr. DeSilvey balances practice as a rural family nurse practitioner with advocacy for addressing the social determinants of health in clinical and community practice. Her area of focus is developing the terminology to capture the process of caring for patients’ social needs. Her work has evolved into [The Gravity Project](#) – a national consensus initiative to develop terminology and data standards to address social needs. She serves as the Director of Clinical Informatics helping to weave the practice of social care, community-based approaches, and the literature of social risk into health care terminology.
• **Rachel L.J. Thornton, MD, PHD:** Dr. Thornton is Executive Director for Clinical Services in the Johns Hopkins Medicine Office of Population Health as well as an Associate Professor of Pediatrics at Johns Hopkins School of Medicine with a joint appointment in the Johns Hopkins Bloomberg School of Public Health. She’s also an Associate Director of both the Johns Hopkins Urban Health Institute and the Johns Hopkins Center for Health Equity.
Our Process

Panelists will each be asked each question but may elect to PASS

After each panelist shares their perspective, colleagues may offer a REJOINDER if they are so moved

To improve upon previous debates, we will refrain from interrupting each other or talking over one another 😊
YSDOHNEMR?

(How my voice recognition software translated the question “Why SDOH in EMR?”)
Describe your experience with SDOH measurement, collection and use in EMR as a movie title. What actor would play you in the movie?
What should be the social marketing plan for the “word of mouth” at the frontline? What do you say to the clinician who says, “this takes too long to collect, it puts patients off, and we’re not really going to use it?”
Tell your **favorite story** about SDOH in EMR that illustrates a major success story or a major challenge to be overcome.
Should we, can we, how can we use geo-data in EMR (to supplement/ in lieu of person-by-person data collection)?
Are you ever in conflict with yourself when wearing different “hats”? E.g., as a system-level Population Health Manager, do you have conflicts with your researcher self, your frontline clinician self, and your patient self, over what SDOH items you want to measure, collect, and access in EMR?
State of Socio-Behavioral Data in EHR and Examples of Use Cases

Elham Hatef, MD, MPH, FACPM

Center for Population Health IT
General Preventive Medicine Residency Program
Department of Health Policy and Management
March 12th, 2021
Introduction

- Development of **EHRs & Cross-Provider Regional HIE (CRISP)**
  - Opportunity for **Data-Oriented SDOH Assessment & Intervention**

  - The **point of care** (assessment & referrals of an individual with social need)
  - The **health delivery system level** (hiring a social worker in the clinic)
  - The **community** (building or strengthening community-based initiatives)
Integrating Social and Behavioral Determinants of Health into Population Health Analytics: A Conceptual Framework and Suggested Road Map

Zachary Predmore, Elham Hatef, and Jonathan P. Weiner

Published Online: 13 Mar 2019 | https://doi-org.proxy1.library.jhu.edu/10.1089/pop.2018.0151
Data Sources: Healthcare Systems

- Electronic Health Records
  - Structured
  - Unstructured (Free-Text)
Data Sources: Healthcare Systems

- We used Johns Hopkins EHR data from 2003 to 2018

- **Structured Data**
  January 2003 and June 2018 from 5,401,324 unique patients

- **Unstructured Data**
  July 2016 and May 2018 of 1,188,202 unique patients
# Collection Methods & Characteristics of Selected SBDH in Structured Data

<table>
<thead>
<tr>
<th>Common Collection Method</th>
<th>Completeness Rate</th>
<th>Collection Date</th>
<th>Facility Type</th>
<th>History and Details</th>
<th>Other Collection Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Address / Zip Code</td>
<td>~5.2 million patients (95%)</td>
<td>2003-Current</td>
<td>All facilities at the time of registration</td>
<td>~66% of patients’ address change records are available, with effective start and end dates to track address change over time</td>
<td>Billing Address, Claims Processing Address, Home Health Encounters and Episodes, Communications for Specific Encounters</td>
</tr>
<tr>
<td>Upon registration of each encounter. Documented as a street name &amp; number, an optional line for apartment or other information, a city, a state or province, and a zip code.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Ethnicity | Upon registration of each encounter | ~2.7 million patients (50%) | 2003-Current | All facilities at the time of registration | Ethnicity (Hispanic or Not Hispanic) captured separately from race | Transplant Organ Donors, Ethnicity Questionnaire, Ethnicity Origin Questionnaire |
| Race | Upon registration of each encounter | ~4.9 million patients (90%) | 2003-Current | All facilities at the time of registration | Patients can self-identify multiple races | Home Health, Transplant Organ Donors |
| Preferred Language | At the time of admission | 2,718,416 patients (50%) | 2003-Current | All facilities at the time of an encounter | The top preferred languages, by unique patient count: English (2,626,379, 48.6%) & Spanish (53,446, 0.9%) | Flowsheets, Questionnaires, Clinical Notes |
| Alcohol Use: “Alcoholic Drinks Per Week” | Social History portion of EHR during a patient encounter, whether in-person or not in-person encounters (telephone, MyChart, documentation) | 490,348 (9.08%) patients | 2013-Current | All facilities at the time of an encounter | Reports show having any value (including 0 alcoholic drinks per week) in social history | Flowsheets, Questionnaires, Clinical Notes |
| Smoking Status | Social History portion of EHR during a patient encounter, whether in-person or not in-person encounters (telephone, MyChart, documentation) | 1,728,749 (32%) patients reported having any value smoking status in social history | 2013-Current | All facilities at the time of an encounter | Smoking Quit Date is also populated, but only in 137,958 (2.6%) of encounters | Flowsheets, Questionnaires, Clinical Notes |
### Number of Patients with Selected SDOH Domains in EHR – Using Diagnoses-Based Query

<table>
<thead>
<tr>
<th>SBDH Categories and Subtypes/Codes</th>
<th>Diagnoses Based Query</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Connection / Isolation</strong></td>
<td><strong>35,171 (0.64%)</strong></td>
</tr>
<tr>
<td>Z60.2 Problems Related to Living Alone</td>
<td>1222</td>
</tr>
<tr>
<td>Z60.4 Social Exclusion and Rejection</td>
<td>223</td>
</tr>
<tr>
<td>Z63.0 Relationship Problems (with spouse/ partner)</td>
<td>852</td>
</tr>
<tr>
<td>Z63.5 Family Disruption (separation/ divorce)</td>
<td>548</td>
</tr>
<tr>
<td>Z63.8 Other Primary Support Group Problems</td>
<td>2230</td>
</tr>
<tr>
<td>Z63.9 Unspecified Primary Support Group Problem</td>
<td>3247</td>
</tr>
<tr>
<td>Z65.9 Unspecified Psychosocial Circumstances</td>
<td>938</td>
</tr>
<tr>
<td>Z73.4 Inadequate Social Skills</td>
<td>81</td>
</tr>
<tr>
<td>Z91.89 Other Specified Personal Risk Factors</td>
<td>18,947</td>
</tr>
<tr>
<td>R45.8 Other Emotional State Symptoms and Signs</td>
<td>3340</td>
</tr>
<tr>
<td><strong>Housing Issues</strong></td>
<td><strong>10,433 (0.19%)</strong></td>
</tr>
<tr>
<td>Z59.0 Homelessness</td>
<td>7022</td>
</tr>
<tr>
<td>Z59.1 Inadequate Housing</td>
<td>120</td>
</tr>
<tr>
<td>Z59.8 Other Housing Problems</td>
<td>3291</td>
</tr>
<tr>
<td><strong>Income / Financial Resource Strain</strong></td>
<td><strong>3543 (0.06%)</strong></td>
</tr>
<tr>
<td>Z59.5 Extreme Poverty</td>
<td>68</td>
</tr>
<tr>
<td>Z59.6 Low Income</td>
<td>72</td>
</tr>
<tr>
<td>Z59.7 Insufficient Social Insurance and Welfare</td>
<td>46</td>
</tr>
<tr>
<td>Z59.8 Other Economic Circumstances Problems</td>
<td>3357</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>767,901</td>
</tr>
<tr>
<td>Male</td>
<td>549,434</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>299,220</td>
</tr>
<tr>
<td>35-44</td>
<td>205,697</td>
</tr>
<tr>
<td>45-54</td>
<td>199,700</td>
</tr>
<tr>
<td>55-64</td>
<td>232,253</td>
</tr>
<tr>
<td>65-74</td>
<td>195,356</td>
</tr>
<tr>
<td>75-84</td>
<td>115,758</td>
</tr>
<tr>
<td>85+</td>
<td>69,351</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>741,472</td>
</tr>
<tr>
<td>African American</td>
<td>341,439</td>
</tr>
<tr>
<td>Asian American</td>
<td>71,787</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>1,338</td>
</tr>
<tr>
<td>Other</td>
<td>103,903</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
</tr>
<tr>
<td>Baltimore County</td>
<td>252,957</td>
</tr>
<tr>
<td>Howard</td>
<td>152,471</td>
</tr>
<tr>
<td>Montgomery</td>
<td>259,383</td>
</tr>
<tr>
<td>Prince George's</td>
<td>82,953</td>
</tr>
<tr>
<td>Baltimore city</td>
<td>230,214</td>
</tr>
</tbody>
</table>
### Study Updates - Breakdown of SDOH Domains by Healthcare Utilization

<table>
<thead>
<tr>
<th></th>
<th>Outpatient</th>
<th>Emergency Department</th>
<th>Inpatient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Study Population</strong></td>
<td>m=3.458 (sd=8.73)</td>
<td>m=0.159 (sd=0.636)</td>
<td>m=0.061 (sd=0.241)</td>
</tr>
<tr>
<td><strong>Any Social Needs</strong></td>
<td>m=8.277 (sd=15.424)</td>
<td>m=0.495 (sd=3.038)</td>
<td>m=0.137 (sd=0.473)</td>
</tr>
<tr>
<td><strong>Financial Challenges</strong></td>
<td>m=16.299 (sd=25.126)</td>
<td>m=0.626 (sd=3.263)</td>
<td>m=0.313 (sd=0.572)</td>
</tr>
<tr>
<td><strong>Food Insecurity</strong></td>
<td>m=18.905 (sd=26.856)</td>
<td>m=0.881 (sd=4.309)</td>
<td>m=0.368 (sd=0.783)</td>
</tr>
<tr>
<td><strong>Housing Issues</strong></td>
<td>m=14.849 (sd=22.263)</td>
<td>m=1.963 (sd=6.76)</td>
<td>m=0.4 (sd=0.814)</td>
</tr>
<tr>
<td><strong>Transportation Issues</strong></td>
<td>m=7.436 (sd=14.029)</td>
<td>m=0.207 (sd=1.323)</td>
<td>m=0.096 (sd=0.356)</td>
</tr>
</tbody>
</table>

#### Patients with the 3 most coded social needs

- **Z59.8** (Other problems related to housing and economic circumstances)
  - m=15.956 (sd=24.012) in Outpatient
  - m=0.62 (sd=3.293) in Emergency Department
  - m=0.313 (sd=0.574) in Inpatient

- **Z59.0** (homelessness)
  - m=14.357 (sd=20.798) in Outpatient
  - m=5.129 (sd=11.267) in Emergency Department
  - m=0.662 (sd=1.159) in Inpatient

- **Z76.89** (encountering health services in other circumstances)
  - m=6.326 (sd=11.645) in Outpatient
  - m=0.161 (sd=1.317) in Emergency Department
  - m=0.068 (sd=0.306) in Inpatient
### Number of Patients with Selected SBDH Domains in EHR – Using Diagnoses-Based Query and Unstructured Data

<table>
<thead>
<tr>
<th>SBDH Categories and Subtypes/Codes</th>
<th>Diagnoses-Based Query</th>
<th>Unstructured</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Connection / Isolation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z60.2 Problems Related to Living Alone</td>
<td>35,171 (0.64%)</td>
<td>30,893 (2.59%)</td>
</tr>
<tr>
<td>Z60.4 Social Exclusion and Rejection</td>
<td>1222</td>
<td>-</td>
</tr>
<tr>
<td>Z63.0 Relationship Problems (with spouse/ partner)</td>
<td>852</td>
<td>-</td>
</tr>
<tr>
<td>Z63.5 Family Disruption (separation/ divorce)</td>
<td>548</td>
<td>-</td>
</tr>
<tr>
<td>Z63.8 Other Primary Support Group Problems</td>
<td>2230</td>
<td>-</td>
</tr>
<tr>
<td>Z63.9 Unspecified Primary Support Group Problem</td>
<td>3247</td>
<td>-</td>
</tr>
<tr>
<td>Z65.9 Unspecified Psychosocial Circumstances</td>
<td>938</td>
<td>-</td>
</tr>
<tr>
<td>Z73.4 Inadequate Social Skills</td>
<td>81</td>
<td>-</td>
</tr>
<tr>
<td>Z91.89 Other Specified Personal Risk Factors</td>
<td>18,947</td>
<td>-</td>
</tr>
<tr>
<td>R45.8 Other Emotional State Symptoms and Signs</td>
<td>3340</td>
<td>-</td>
</tr>
<tr>
<td><strong>Housing Issues</strong></td>
<td>10,433 (0.19%)</td>
<td>35,646 (2.99%)</td>
</tr>
<tr>
<td>Z59.0 Homelessness</td>
<td>7022</td>
<td>-</td>
</tr>
<tr>
<td>Z59.1 Inadequate Housing</td>
<td>120</td>
<td>-</td>
</tr>
<tr>
<td>Z59.8 Other Housing Problems</td>
<td>3291</td>
<td>-</td>
</tr>
<tr>
<td><strong>Income / Financial Resource Strain</strong></td>
<td>3543 (0.06%)</td>
<td>11,882 (0.99%)</td>
</tr>
<tr>
<td>Z59.5 Extreme Poverty</td>
<td>68</td>
<td>-</td>
</tr>
<tr>
<td>Z59.6 Low Income</td>
<td>72</td>
<td>-</td>
</tr>
<tr>
<td>Z59.7 Insufficient Social Insurance and Welfare</td>
<td>46</td>
<td>-</td>
</tr>
<tr>
<td>Z59.8 Other Economic Circumstances Problems</td>
<td>3357</td>
<td>-</td>
</tr>
</tbody>
</table>
Characteristics of EHR’s Unstructured Data Containing SBDH – Stratified by Provider Role & Note Type
Collaborations with Other Health Systems - A Pilot Study to Improve the Use of Electronic Health Records for Identification of Patients with Social Needs: A Collaboration of Johns Hopkins Health System and Kaiser Permanente

• Conducted independently, in a parallel and coordinated framework across sites

• The validation assessment and NLP algorithm logic were identical across sites
  • The “gold standard” for assessment of algorithm validity differed according to data availability

• Population Studied
  • Beneficiaries ≥18 years of age during 2016 through 2019 who received care at JHHS, KPMAS, KPSCal
Collaborations with Other Health Systems - A Pilot Study to Improve the Use of Electronic Health Records for Identification of Patients with Social Needs: A Collaboration of Johns Hopkins Health System and Kaiser Permanente

<table>
<thead>
<tr>
<th></th>
<th>JHHS</th>
<th>KPMAS</th>
<th>KPScal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Population (Patient No.)</strong></td>
<td>~1,200,000</td>
<td>~1,600,000</td>
<td>~4,700,000</td>
</tr>
<tr>
<td><strong>NLP Validation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold Standard Method</td>
<td>SDOH Questionnaire</td>
<td>SDOH Questionnaire</td>
<td>SDOH ICD codes Manual Annotation</td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Patients/ Response No. (with/without residential Instability) | 1,000 (500+/ 500-) | 8,197 (833+, 7364-) | 300 (150+/150-)
| Clinical Note No.                    | 134,062  | 78,825   | 9,575    |
| **NLP Algorithm Performance**        |          |          |          |
| Sensitivity                          | 0.84     | 0.61     | 0.96     |
| Specificity                          | 0.96     | 0.87     | 0.97     |
Assessing the Impact of Social Needs and Social Determinants of Health on Health Care Utilization: Using Patient- and Community-Level Data

Elham Hatif, MD, MPH1,2 Xiaomeng Ma, MS1 Masoud Rouhizadeh, MS, PhD3 Gurmehar Singh, MS1 Jonathan P. Weiner, DrPH1 and Hadi Kharrazi, MD, PhD1,4
Use case: Predictive Risk Modeling

• A 3-year retrospective population-based study

• Assessing individual- and community-level housing needs and how the level of housing needs impacts health care utilization

• EHR data from July 2016 to May 2018
  • 1,187,956 unique patients – EHR structured and unstructured data

• 2017 U.S Census Data (American Community Survey)
  • Area Deprivation Index – A composite measure to rank neighborhoods based on their socio-economic challenges (higher ADI indicates more challenges)
### Key Findings

Logistic Regression Assessing Factors Associated with Healthcare Utilization Among Johns Hopkins Patients Between 2016-2018

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall Population</th>
<th>Medicaid Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Housing Issues</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homelessness</td>
<td>1.336</td>
<td>1.902</td>
</tr>
<tr>
<td>Housing Instability</td>
<td>1.489</td>
<td>1.473</td>
</tr>
<tr>
<td>Characteristics of the Building</td>
<td>0.888</td>
<td>0.847</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.001</td>
<td>1.010</td>
</tr>
<tr>
<td><strong>Sex (male as reference)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.437</td>
<td>1.563</td>
</tr>
<tr>
<td><strong>Race (whites as reference)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.959</td>
<td>0.794</td>
</tr>
<tr>
<td><strong>Neighborhood Characteristics (ADI National Rank, neighborhoods below the 10th percentile as reference)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between 11&lt;sup&gt;th&lt;/sup&gt; &amp; 89&lt;sup&gt;th&lt;/sup&gt; Percentiles</td>
<td>1.442</td>
<td>1.466</td>
</tr>
<tr>
<td>Above the 90&lt;sup&gt;th&lt;/sup&gt; Percentile</td>
<td>1.549</td>
<td>1.598</td>
</tr>
<tr>
<td><strong>Insurance Type (commercial insurance as reference)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicare</td>
<td>1.489</td>
<td>-</td>
</tr>
<tr>
<td>Medicaid</td>
<td>2.078</td>
<td>-</td>
</tr>
<tr>
<td><strong>Charlson Comorbidity Score (score of 0 as reference)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;= 3</td>
<td>55.444</td>
<td>38.497</td>
</tr>
</tbody>
</table>

---

### Variables

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Housing Issues</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homelessness</td>
<td>1.336</td>
<td>1.261, 1.416</td>
<td>&lt;0.00001</td>
<td>1.902</td>
<td>1.576, 2.296</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Housing Instability</td>
<td>1.489</td>
<td>1.380, 1.607</td>
<td>&lt;0.00001</td>
<td>1.473</td>
<td>1.227, 1.769</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Characteristics of the Building</td>
<td>0.888</td>
<td>0.818, 0.964</td>
<td>0.00469</td>
<td>0.847</td>
<td>0.640, 1.121</td>
<td>0.24500</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.001</td>
<td>1.001, 1.002</td>
<td>&lt;0.00001</td>
<td>1.010</td>
<td>1.008, 1.011</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Sex (male as reference)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.437</td>
<td>1.409, 1.467</td>
<td>&lt;0.00001</td>
<td>1.563</td>
<td>1.458, 1.675</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Race (whites as reference)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.959</td>
<td>0.937, 0.981</td>
<td>0.00039</td>
<td>0.794</td>
<td>0.734, 0.858</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Neighborhood Characteristics (ADI National Rank, neighborhoods below the 10th percentile as reference)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between 11&lt;sup&gt;th&lt;/sup&gt; &amp; 89&lt;sup&gt;th&lt;/sup&gt; Percentiles</td>
<td>1.442</td>
<td>1.404, 1.481</td>
<td>&lt;0.00001</td>
<td>1.466</td>
<td>1.239, 1.734</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Above the 90&lt;sup&gt;th&lt;/sup&gt; Percentile</td>
<td>1.549</td>
<td>1.474, 1.627</td>
<td>&lt;0.00001</td>
<td>1.598</td>
<td>1.325, 1.926</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Insurance Type (commercial insurance as reference)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicare</td>
<td>1.489</td>
<td>1.447, 1.532</td>
<td>&lt;0.00001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Medicaid</td>
<td>2.078</td>
<td>1.997, 2.162</td>
<td>&lt;0.00001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Charlson Comorbidity Score (score of 0 as reference)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;= 3</td>
<td>55.444</td>
<td>53.333, 57.639</td>
<td>&lt;0.00001</td>
<td>38.497</td>
<td>32.447, 45.675</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>
Challenges and Future Road Map

- Lack of standards, tools and best practices
- Cost burden on providers and health systems
- Data interoperability, confidentiality, and validity
Conclusions

- **Growing pressures from payers and policy makers** to achieve greater value for patients and beneficiaries

- **Clinicians, health plans, and provider organizations** must, in the near term, find ways to more effectively introduce social and behavioral factors into the medical care process

- The need for **evidence & best practices** derived from the social, informatics, and public health sciences will be essential

- Although **numerous technical, operational, and political challenges remain**, there is little question that a social and behavioral determinant-enabled, **approach to patient care and population health** will be necessary
Research Team & Collaborators

- **Center for Population Health IT**
  - Hsien-Yen Chang, PhD
  - Chris Kitchen, MA
  - Xiaomen Ma, MA
  - Kelly Searle, PhD
  - Elyse Lasser, MA
  - Hadi Kharrazi, MD, PhD
  - Jonathan Weiner, Dr.PH

- **Preventive Medicine Residency Program**
  - Marissa Tan, DO, MPH
  - Delaram Taghipoor, MD, MPH, MBA

- **Clinical Research Data Acquisition (CCDA)/ Institute for Clinical & Translational Research (ICTR)**
  - Masoud Rouhizadeh, PhD
Natural Language Processing Methods to Extract Socio-Behavioral Data from Unstructured EHR

Masoud Rouhizadeh, MA. MS. Ph.D
Structured data

- EPIC SBDH Wheel
- LOINC
- ICD
- HL7 Gravity
- EPIC CHR (Comprehensive Health Record)
Unstructured free text

• More than half of all data in an EHR system is unstructured.

  • Clinical notes, authored by providers (progress notes, surgical notes, discharge summaries, consults)

  • Radiology and pathology reports

  • Messages in patient portal and telephone conversations

• Manual coding is laborious and costly
From unstructured free text to structured data

• Natural language processing
• Sometimes referred to as text mining
• A subfield of artificial intelligence.
• Transforms unstructured data into structured data.
• Provides data for machine learning algorithms.
Extracting diagnosis, signs, symptoms, onsets and negation

1. The patient is a 56-year-old male diagnosed with SARS-CoV-2.

3. One week prior to presentation (D1: ~2/4) pt began experiencing generalized weakness and fatigue.

5. Day prior to admission pt had nausea, emesis, diarrhea.

7. No subjective fever.

9. Does endorse cough (non-productive) on day of admission (2/11).
Challenges

- Socio-Behavioral data extraction is difficult.
- Lack of uniform methods for documenting SBDH
  - Coding systems (ICD-10 and LOINC)
  - Taxonomies (SNOMED)
  - HL7 Gravity
Challenges

• Variations in linguistic expressions
  • Pt is homeless
  • Stays in a shelter.
  • Is looking for a place to stay.
  • Has trouble finding a place to stay/sleep
  • Does not have a permanent address
  • Sleeps in the bus station
  • Referred to the [...] Shelter
Challenges

• 9M encounters for over 1,000,000 patients
• Big data, has to be computationally efficient.
• Annotation human resources.
Search phrases

- Ontologies: LOINC, SNOMED, ICD10
- Public Health Surveys and Instruments
- Literature review

<table>
<thead>
<tr>
<th>Lack of Social Support (Social Isolation) or At Risk for Social Isolation (N=258)</th>
<th>In Need of Social Support Services (N=166)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Green Phrases (N=119)</strong></td>
<td><strong>Green Phrases (N=106)</strong></td>
</tr>
<tr>
<td>No household member renders care</td>
<td>Patient has no care provider</td>
</tr>
<tr>
<td><strong>Yellow Phrases (N=113)</strong></td>
<td><strong>Yellow Phrases (N=44)</strong></td>
</tr>
<tr>
<td>Alone</td>
<td>Patient should not live alone</td>
</tr>
<tr>
<td><strong>Red Phrases (N=26)</strong></td>
<td><strong>Red Phrases (N=16)</strong></td>
</tr>
<tr>
<td>Power of attorney</td>
<td>Persons encountering health services in other</td>
</tr>
<tr>
<td><strong>Lonely</strong></td>
<td><strong>It is not safe for patient to live alone</strong></td>
</tr>
<tr>
<td>lives alone</td>
<td><strong>Community living</strong></td>
</tr>
<tr>
<td>High expressed emotional level within family</td>
<td>No help</td>
</tr>
</tbody>
</table>
## Structured and Unstructured Data Coverage

<table>
<thead>
<tr>
<th>SBDH Category</th>
<th>Diagnoses-Based Query (~5.4M)</th>
<th>Unstructured (~1.1M)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnoses-Based Query (~5.4M)</strong></td>
<td>Patient Count (%)*</td>
<td>Patient Count (%)</td>
</tr>
<tr>
<td><strong>Housing Issues</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homelessness</td>
<td>10,433 (0.19%)</td>
<td>35,646 (2.99%)</td>
</tr>
<tr>
<td>Z59.0 Homelessness</td>
<td>7022</td>
<td>108,439 Notes</td>
</tr>
<tr>
<td>Inadequate Housing</td>
<td>Z59.1 Inadequate Housing</td>
<td>120</td>
</tr>
<tr>
<td>Housing Characteristics</td>
<td>Z59.8 Other Housing Problems</td>
<td>3291</td>
</tr>
</tbody>
</table>
Linguistic pattern assessment

• Housing domain: 36% false positive.
Epic Flowsheets and Questionnaires

• Using semi-structured data to create a gold-standard dataset

• Is the patient homeless? (yes/no)

• Abuse/Neglect Screen Questionnaire:
  - 12k patient encounters (96% answered)

• Adult Admission General Intake Form:
  - 77k patient encounters (35% answered)
Data preparation

- Category indicative words:
  - home, house, homeless, place, shelter
- ±n word concordance window
- Removing redundant content
- Covering 99.8% of patients
NLP Methods to Extract Socio-Behavioral Data from Unstructured EHR

Machine Learning

Text Classification Task

Documents & Categories (Labels)

Machine Learning Algorithm

Model Training

New documents

Publish model

Trained model

Model Serving

Text Categories
Machine Learning

• Logistics Regression

• Support Vector Machines

• ngram features

• TF.IDF

• L1 regularization for dimensionality reduction
Identifying Homelessness Status

<table>
<thead>
<tr>
<th>Model</th>
<th>Pre</th>
<th>Rec</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi-RNN</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Logistics Regression - unigrams</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Logistics Regression - 1-3-grams</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Logistics Regression - 1-3-grams, TF.IDF</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>SVM - unigrams</td>
<td>0.90</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>SVM - 1-3-grams</td>
<td>0.90</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>SVM - 1-3-grams, TF.IDF</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Identifying Homelessness Status

3.2 Log odds ratio

Assuming \( p_s \) to be a probability of success, the odds ratio of \( p_s \) was defined as:

\[
\text{odds}(p_s) = \frac{p_s}{1 - p_s} \tag{9}
\]

We then defined odds ratio \( \Gamma \) between a term \( w \) and document \( d \) as the ratio between the odds of seeing \( w \) in \( d \) or \( P(w|d) \), and the odds of seeing \( w \) in all other documents or \( P(w|\overline{d}) \). Given the counts and probabilities defined above, we formulated \( \Gamma \) as follow:

\[
\Gamma = \frac{\text{odds}(p_1)}{\text{odds}(p_2)} = \frac{p_1(1 - p_2)}{p_2(1 - p_1)}
\]

\[
= \frac{P(w|d)(1 - P(w|\overline{d}))}{P(w|\overline{d})(1 - P(w|d))}
\]

\[
= \frac{\text{tf}_{wd}(N_c d c_w + \text{tf}_{wd})}{(c_w \text{tf}_{wd})(c_d \text{tf}_{wd})}(10)
\]

we then calculated the log \( \Gamma \). To smooth zero values, we added 0.5 to each multiplier in the final equation above.

3.3 Highly associated word sequences

Table 3.3 shows the most associated ngrams to the homelessness category. Most of these sequences contained actual housing-related terms which could in turn, be used for future text mining and keyword search.

<table>
<thead>
<tr>
<th>n-gram</th>
<th>log-odds</th>
<th>occurrence</th>
<th>occ. w homeless</th>
<th>#of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>unstable housing</td>
<td>4.0</td>
<td>275</td>
<td>275</td>
<td>54</td>
</tr>
<tr>
<td>blue shelter</td>
<td>4.6</td>
<td>486</td>
<td>486</td>
<td>64</td>
</tr>
<tr>
<td>is currently homeless</td>
<td>4.7</td>
<td>525</td>
<td>525</td>
<td>100</td>
</tr>
<tr>
<td>shelter and</td>
<td>4.7</td>
<td>533</td>
<td>533</td>
<td>110</td>
</tr>
<tr>
<td>homelessness</td>
<td>4.8</td>
<td>5457</td>
<td>5473</td>
<td>252</td>
</tr>
<tr>
<td>the shelter</td>
<td>4.8</td>
<td>579</td>
<td>579</td>
<td>108</td>
</tr>
<tr>
<td>intoxication</td>
<td>4.8</td>
<td>593</td>
<td>593</td>
<td>70</td>
</tr>
<tr>
<td>malingering</td>
<td>4.8</td>
<td>601</td>
<td>601</td>
<td>64</td>
</tr>
<tr>
<td>living situation homeless</td>
<td>4.9</td>
<td>634</td>
<td>634</td>
<td>94</td>
</tr>
<tr>
<td>of homelessness</td>
<td>4.9</td>
<td>648</td>
<td>648</td>
<td>84</td>
</tr>
</tbody>
</table>
A spectrum of approaches

complexity — accuracy — efficiency

• Expert-developed search patterns

• Machine Learning (Logistic Regression)

• Deep Learning (Convolutional Neural Networks)
Big Data for SBDH

- Methods are transferable and generalizable
- Minimal human annotation
- Questionnaires for gold-standard labels
- Operational use-cases
Big Data for SBDH

- Alcohol intake
- Substance abuse
- Smoking status
- Housing status
- Food insecurity
- Income
- Education
- Transportation
- Mental health
- Domestic abuse
- Physical activity
- Social isolation
Delivery methods

• HIPAA-compliant environments

• SAFE Desktop

• Precision Medicine Analytics Platform (PMAP)

• Any combination of:
  • Raw or processed data
  • Data analysis tools
Thank you.
Linking EHR and Population-Level Data to Address Social Determinants of Health

Symposium on Capturing SDoH in EHRs
Funded by JHMI ICTR-BSSS

Johns Hopkins School of Public Health
Johns Hopkins School of Medicine

Hadi Kharrazi, MD PhD FAMIA
Center for Population Health IT (CPHIT)
Mar 12, 2021
Summary

❖ Part I – Background about geo-derived SDoH
  ▪ Why geo-derived SDoH?
  ▪ What are the major types of geo-derived data?
  ▪ What are the common geo-derived SDOH data sources and variables?
  ▪ What are the different geo-levels?
  ▪ How geo-derived SDOH are linked with EHR data?
  ▪ What are composite SDoH scores?

❖ Part II – Maryland SDoH & JHMI patient population (Nexus project)
  ▪ Census maps (Maryland counties & Baltimore City tracts)
  ▪ JHMI maps (Maryland counties & Baltimore City tracts)
  ▪ Current status of geo-derived SDoH in JHMI Epic
Part I – Geo-derived SDoH
Why geo-derived SDOH factors?

- **Individual-level social needs**
  (e.g., ICD, LOINC, SNOMED codes)
  - **Pros**: high accuracy (e.g., clinical and social decision making)
  - **Cons**: incomplete; ad-hoc collection, but improving

- **Geo-derived social determinants of health**
  (e.g., neighborhood-level housing condition)
  - **Pros**:
    - Can be assigned for anybody with a permanent address (e.g., population level analysis)
    - Can easily be generated and linked with clinical data (e.g., incorporating in EHR’s data warehouse)
  - **Cons**:
    - Values are approximations, specially if based on survey data (e.g., margin of error)
    - Represents group associations and not individual-level associations (e.g., not for clinical judgement)
    - Requires special analytic methodology to address ecological fallacy (e.g., multi-level modeling)
What are the major types of geo-derived data?

- **Points of interest**
  - Map use: Background map with point and polygon landmarks
  - Data source: Largely derived from other government sources
  - Representation: Points and polygons

- **Other Administrative Units**
  - Map use: To depict local, regional, state and federal districts
  - Data source: Derived from census administrative units
  - Representation: Polygons

- **Census administrative units**
  - Map use: To delineate census units such as blocks and tracts
  - Data source: Compiled with local government input
  - Representation: Polygons

- **Streets and Addresses**
  - Map use: Street network analysis and address location
  - Data source: Largely derived from other government sources
  - Representation: Lines for streets, tables for addresses

- **Census boundaries**
  - Map use: Linework from which other census features are generated
  - Data source: Compiled and partially derived from government sources
  - Representation: Lines for water, railroads, transmission, and other
What are the major types of geo-derived data? (cont.)

Count Data

Point Data
What are the common geo-derived SDOH data sources & vars?

❖ **Census** *(e.g., Decennial Census; ACS American Community Survey)*
  - Demographics: Age, Sex, Race, Ethnicity (Hispanic or Latino Origin); Ancestry
  - Residency: Citizenship Status; Year of Entry; Foreign Born Place of Birth; Place of Birth; Migration/Residence 1-Year Ago
  - Transportation: Commuting (Journey to Work); Place of Work
  - Household: Household Type; Family Type; Subfamilies; Relationship to Householder; Grandparents/children Characteristics
  - Marital: Marital Status; Marital History; Fertility
  - Education: School Enrollment; Educational Attainment; Undergraduate Field of Degree; Language Spoken at Home
  - Finance: Poverty Status; Disability Status; Income; Earnings; Food Stamps/SNAP
  - Jobs: Employment Status; Work Status Last Year; Industry, Occupation, and Class of Worker
  - Military: Veteran Status; Period of Military Service
  - Housing: Housing Characteristics; Group Quarters
  - Health: Health Insurance Coverage
  - Technology: Computer and Internet Use
  - Political: Citizen Voting-Age Population

❖ **Other Federal Agencies**
  - CDC: Health
  - USDA: Nutrition
  - IRS/SSA: Financial/Tax
  - FBI: Crime

❖ **Other Data Sources**
  - State databases
  - City/County databases
  - Commercial databases
  - Geographic
What are the different geo-levels?

- **Nation**
  - Zip Code
  - Census ZCTA

- **Region**
  - Country
  - Division

- **State**
  - School Dist.
  - Cong. Dist.
  - Subdivision

- **County**
  - Census Tract
  - Census Block Group
  - Census Block

- **Census Tract**
  - Census Block Group
  - Census Block

- **Census ZCTA**
  - Zip Code

- **Metro/Micro**
  - Tribal
  - Urban Area

- **Places**
  - Legislative

**Examples**

- **Nation**
  - Maryland
  - State = 1
  - County = 24
  - Tract = 1406
  - BG = 3926
  - ZCTA = 468
  - Zip = 603

- **Nation**
  - Nation
  - State = 56
  - County = 3233
  - Tract = 74,133
  - BG = 220,740
  - ZCTA = 33,144
  - Zip = 41,692
How geo-derived SDOH are linked with EHR data?
How geo-derived SDOH are linked with EHR data? (cont.)

- Geo-count data (geopolitical boundaries)
  - State/County
    - Pros
      - readily available in residential addresses
      - many data sources exists on this level
    - Cons
      - high level of aggregation reduces variation among geographies (i.e., very high approximation)
  - Zip code (3 and 5 digits)
    - Pros
      - readily available in residential addresses
      - several data sources exists on this level only
    - Cons
      - high level of aggregation
      - designed for frequency of mail, not residents
      - Zip5 is limited PHI (i.e., limited dataset; LDS)
How geo-derived SDOH are linked with EHR data? (cont.)

- **Geo-count data (cont.)**
  - Census boundaries (Tract & Block Group; FIPS or GEOID code)
    - **Pros**
      - more granular data represents higher variation among geographies (i.e., best approximation)
    - **Cons**
      - not readily available (geo-coding software needed to convert address; issues with conversion)
      - survey margin of error may be high (e.g., margin of error may be higher than the measurement)
      - data availability is limited on Census block-group
      - tract is considered limited PHI (LDS)
      - block-group data is considered full-PHI

- **Geo-point/location data (Latitude/Longitude)**
  - **Pros**
    - provides accurate results and includes a long list of data sources
  - **Cons**
    - needs GIS knowledge to calculate various geo-distances or counts
What are composite SDoH scores?

- Area Deprivation Index (ADI)
  - University of Wisconsin (link)

- Social Vulnerability Index (SVI)
  - CDC (link)

- Social Deprivation Index (SDI)
  - Robert Graham Center (link)

- Social Determinants of Health Index (SDOHi)
  - Boston University & Sharecare (link)

- County Health Ranking
  - RWJF (link)

- Disease/Condition Specific Indexes
  - COVID Community Vulnerability Index (CCVI)
Part II – Maryland SDoH & JHMI patient population
Census data – Population

Population (Census)

Left: Baltimore City (Tract) & Right: Maryland (County)
Census data – Household income

Median Household Income (Census)

*Left: Baltimore City (Tract) & Right: Maryland (County)*
Census data – Commuting to work by personal vehicle

Percentage of Car/Truck/Van Commuters (Census)

*Left: Baltimore City (Tract) & Right: Maryland (County)*
Census data – Houses with no plumbing

Percentage of Houses with No Plumbing (Census)

Left: Baltimore City (Tract) & Right: Maryland (County)
JHMI data – Population

Population (JHMI)

Left: Baltimore City (Tract) & Right: Maryland (County)
JHMI data – Financial

Financial Challenges (JHMI)

*Left: Baltimore City (Tract) & Right: Maryland (County)*
JHMI data – Transportation

Transportation Challenges (JHMI)

*Left: Baltimore City (Tract) & Right: Maryland (County)*
JHMI data – Utilization

Utilization – IP, ED and OP admissions (JHMI)

*Left: Baltimore City (Tract) & Right: Maryland (County)*
Current status of geo-derived SDoH in JHMI Epic

- JHMI Epic Data Warehouse
  - Area Deprivation Index
  - Census ACS Data (5 years) 2015 to 2018
    - Precalculated ratios
  - Ongoing updates to geo-derived SDoH data sources
Questions

kharrazi@jhu.edu
hkharrazi.com
jhsph.edu/cphit