



Validating Population-Level Data Sources for Social Determinants of Health: A Three-Part Research Gap Analysis

4/23/2018

Sarah Gensheimer, MD/PhD Student

Julia Burgdorf, PhD Student

Zachary Predmore, PhD Student

Johns Hopkins Bloomberg School of Public Health



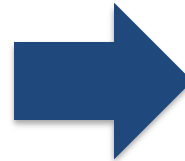
Background

Social Determinants of Health (SDH)

Social Determinants of Health
(e.g. education, income, employment)



Other Determinants of Health
(e.g. genetics, clinical care, behaviors)



**Health
Outcomes for
Individuals &
Populations**

(Adler et al., 2016)



Growing Interest in SDH

Calls for policymakers and health care providers to

- **Address** social determinants of health
- **Collect data** in a clinical setting (e.g. electronic health records)



SDH: Data Sources

Patient-reported data:

Direct from patient, collected during clinical encounter (e.g. office visit, pre-visit survey)

- + Patient-specific
- Collection burden on providers & patients, IT specialists, etc.

Population-level data:

Geographic averages (i.e. county, census tract), collected outside clinical encounter (e.g. U.S. Census Bureau)

- + Accessible using patient address
- Not patient-specific

SDH: Data Sources

Can population-level data predict patient-level outcomes, as well as patient-reported data?

Patient-reported data:

Direct from patient, collected during clinical encounter (e.g. office visit, pre-visit survey)

- + Patient-specific
- Collection burden on providers & patients, IT specialists, etc.

Population-level data:

Geographic averages (i.e. county, census tract), collected outside clinical encounter (e.g. U.S. Census Bureau)

- + Accessible using patient address
- Not patient-specific

Hypothesized Research Gap

There are **no projects** in the **HSRProj database** or the **published literature** that validate population-level data sources against patient-reported data sources for predicting patient-level outcomes.





Methods (3 Parts) & Findings

Exclusion Criteria & Gap Analysis

Exclusion Criteria:

- Did not address validity of *population-level* social determinants of health data sources for predicting *patient-level* outcomes
- Non-U.S.-based
- Pre-2000
- Qualitative methods

Gap analysis:

- Do any included studies validate population-level data sources against patient-reported data sources for predicting patient-level outcomes?



Part 1: HSRProj Search Strategy

- HSRProj available as both a downloadable Excel file and online interface (both were searched)
- Performed searches using abstract/keyword/MeSH and selected articles that met exclusion criteria

Abstract search terms:

- Social determinant
- Social factor
- Social + valid
- Education + valid
- Income + valid

Keyword search terms:

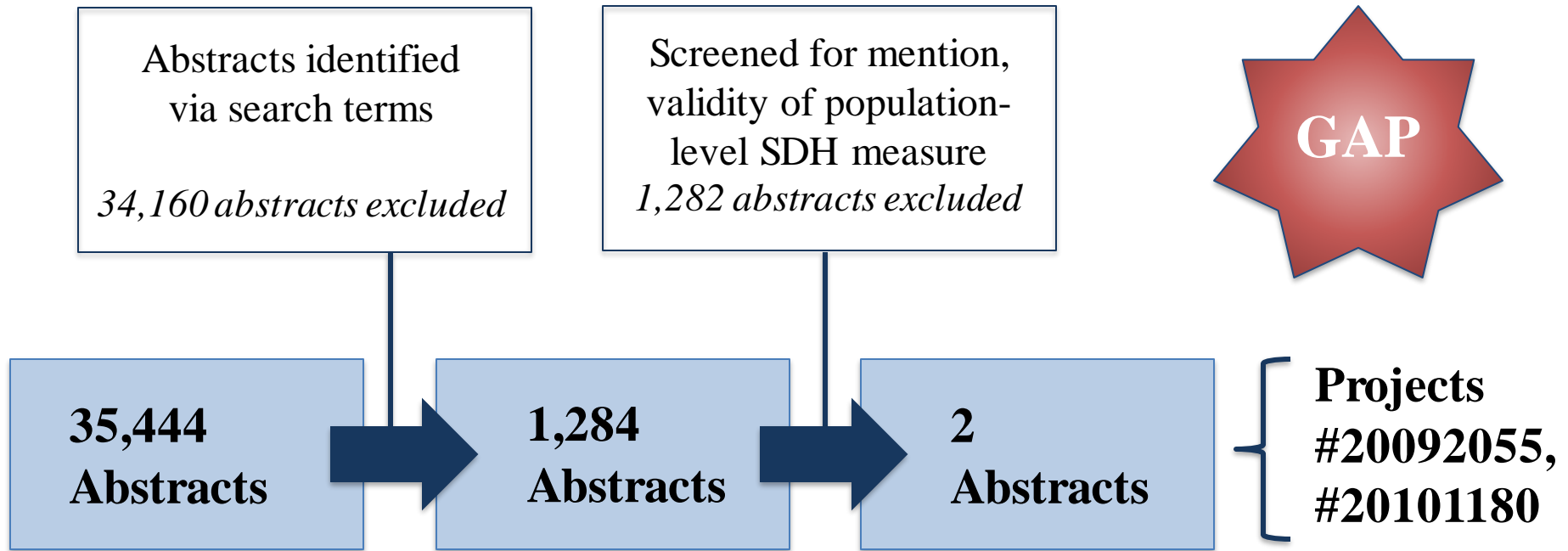
- Social determinant
- Social disparity
- Social + valid

MeSH search terms:

- Social determinants of health



Part 1: HSRProj Search Results



- #20092055: compares neighborhood-level versus individual housing data as measures of SES
- #20101180: evaluates the validity and reliability of a scale to measure urban neighborhood social, physical, resource environments

Part 2: HSRProj & Machine Learning

- **Machine learning**
 - Using computers to analyze data without explicit programming
 - Supervised vs unsupervised approaches
 - Increasing in popularity in health services research and in systematic reviews
- **Our approach**
 - Tokenizing abstract text (i.e. Unique word = 1)
 - Calculating “distance” between abstracts (i.e. # words in common)
 - Review of the “closest” (most similar) abstracts to the two identified in Part 1



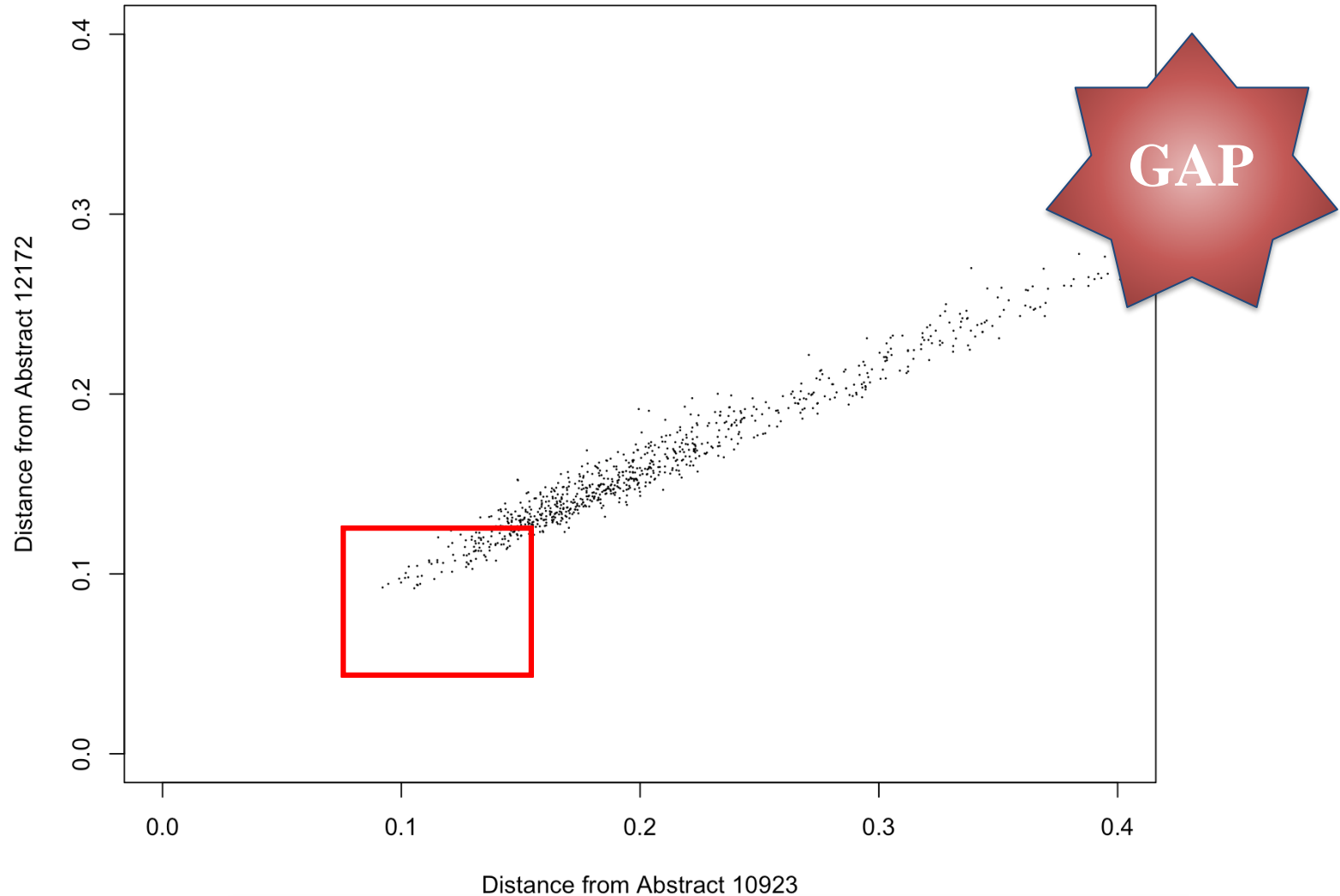
Part 2: “Distance” in practice

- Abstracts with similar words have smaller “distance”

	Word counts					
Abstract	neighborhood	education	urban	hospital	diabetes	Distance
Base	2	2	3	0	0	-
1	2	1	2	0	0	low
2	1	1	1	2	0	medium
3	0	0	0	1	3	high
4	0	0	0	0	0	high



Part 2: HSRProj & Machine Learning



Part 3: Literature Review

- **Consulted Librarian** at Welch Medical Library, Johns Hopkins University.
 - “Validate” = methodology
 - Snowball strategy
- **Searched Scopus** database (contains articles from PubMed/Medline & social sciences)
 - “social determinants of health”
or “SDOH” AND “data sources”
 - “social determinants of health”
or “SDOH” AND “valid*”



Part 3: Literature Review

Scopus Search
(n=247 titles)



3 articles*
included (n=126
references)



3 additional
articles^ included
from snowball

Final Included Articles

1. Kasthurirathne et al. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services (2018)*
2. Ash et al. Social determinants of health in managed care payment formulas (2017)*
3. Kansgara et al. Risk prediction models for hospital readmission: A systematic review (2011)*
4. Corrigan et al. Use of neighborhood characteristics to improve prediction of psychosocial outcomes: A traumatic brain injury model (2012)^
5. Franks et al. Including socioeconomic status in coronary heart disease risk estimation (2010)^
6. Comer et al. Incorporating geospatial capacity within clinical data systems to address social determinants of health (2011)^

GAP

*No articles
validate
population-level
sources against
patient-reported
sources*





Research Gap Significance

Implications for Researchers, Clinicians

If data sources ARE comparable: allows for identification of relevant SDH from more readily-available, population-level sources using patient address

- Reduces patient & provider burden during clinical encounter
- Avoids challenges for administrators, policymakers, researchers, payers, and IT specialists around establishing processes and requirements for collecting and using patient-reported data in EHR

If data sources ARE NOT comparable: confirms necessity of above processes, raises questions around results of research using population-level data

- New questions around SDH data in HSR studies





Proposed Solutions

Solution 1: Comparative Validity Study

- **Build a dataset:**
 - Patient-reported SDH
 - Population-level SDH
 - Clinical information (e.g. EHR data)
 - Utilization information (e.g. claims)
- **Compare predictive ability** of patient-reported & population-level data for patient-level outcomes.
- **Results guide researchers, clinicians, policymakers & IT specialists** in determining necessity of individual-level measures of SDH



Solution 2: Mobilize Research Community

- **Encourage grant-making institutions** to require researchers investigating social determinants use both population- and individual-level measures and compare their findings.
- **Leverage conferences** (e.g. Academy Health) and **publish a perspective** in a widely-circulated health services research journal to increase awareness of the research gap.



The background features a dark blue gradient. On the left side, there is a faint, light blue graphic of a shield. Inside the shield, the upper portion shows a stylized flame, and the lower portion shows a grid representing a globe. A horizontal light blue band is positioned across the middle of the slide, containing the word "Conclusions" in white, bold, serif font.

Conclusions

Our Findings

- Identified a research gap using 3 strategies:
 - HSRProj search
 - Machine learning
 - Scopus search
- Proposed two possible solutions:
 - Comparative validity study
 - Mobilization of research community



Limitations

- Part 1: Relying on HSRProj keywords/MeSH words to be comprehensive and accurate
- Part 2: Search was based on word counts, did not involve semantic analysis
- Part 3: Challenging to search Scopus for methodology, unable to use building block approach



Acknowledgements

- Albert Wu, MD, MPH (Advisor)
- Claire Twose, MLIS (Informationist)
- Support: NRSA & MSTP at Johns Hopkins



Sources (1)

Adler N, et al. (2016). Addressing social determinants of health and health disparities: A vital direction for health and health care. National Academy of Medicine. <https://nam.edu/wp-content/uploads/2016/09/Addressing-Social-Determinants-of-Health-and-Health-Disparities.pdf>

Alley D, et al. (2016.) Accountable health communities – Addressing social needs through Medicare and Medicaid. *New England Journal of Medicine*, 374:8-11. doi: 10.1056/NEJMp1512532

Ash AS, Mick EO, Ellis RP, Kiefe CI, Allison JJ, Clark MA. Social determinants of health in managed care payment formulas. *JAMA Internal Medicine*. 2017; 177(10):1424-1430

Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1. (2014). Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records; Board on Population Health and Public Health Practice; Institute of Medicine. Washington (DC): National Academies Press (US); 2014 Jun.

Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. (2015). Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records; Board on Population Health and Public Health Practice; Institute of Medicine. Washington (DC): National Academies Press (US); 2015 Jan.

Comer KF, Grannis S, Dixon B, Bodenhamer DJ, Wiehe, SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. *Public Health Rep*. 2011; 126(Suppl 3):54-61

Corrigan JD, Bogner J, Pretz C, Mellick D, Kreider S, Whiteneck GG, Harrison-Felix C, Dijkers MP, Heinemann AW. Use of neighborhood characteristics to improve prediction of psychosocial outcomes: A traumatic brain injury model. *Archives of Physical Medicine and Rehabilitation*. 2012; 93(8):1350-1358

Franks P, Tancredi DJ, Winters P, Fiscella K. Including socioeconomic status in coronary heart disease risk estimation. *Annals of Family Medicine*. 2010; 8(5):447-453

HSRProj (Health Services Research Projects in Progress). National Institutes of Health: US National Library of Medicine. https://wwwcf.nlm.nih.gov/hsr_project/home_proj.cfm



Sources (2)

Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk prediction models for hospital readmission: A systematic review. *JAMA*. 2011; 306(15):1688-1698

Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *Journal of the American Medical Informatics Association*. 2018; 25(1):47-53

National health workforce accounts: a handbook. (2017.) Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

National Quality Forum. (2014). Risk adjustment for socioeconomic status or other sociodemographic factors. http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx

Patient-Centered Outcomes Research Institute. *Users' Guide for Integrating Patient-Reported Outcomes into Electronic Health Records*. <http://www.pcori.org/document/users-guide-integrating-patient-reported-outcomes-electronic-health-records>

Clip Art:

- <http://vitrinebr.info/presentation-images-clip-art/presentation-images-clip-art-free-clipart-for-presentations-presentation-clipart-engineering-download/>
- <http://www.clipartpanda.com/categories/green-dollar-sign-clipart>
- <https://openclipart.org/detail/201030/primary-spreadsheet>





Thank You