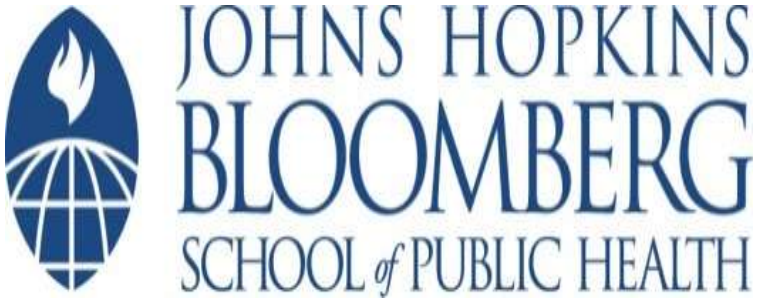


Integrating and Analyzing "Big Data" Across Sectors to Improve the Health and Wellbeing of Populations: Overview of Challenges, Opportunities, and Experiences to Date in the US:

A collaborative workshop coordinated by the Johns Hopkins Bloomberg School of Public Health for visitors from Israel

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

CENTER FOR POPULATION HEALTH
INFORMATION TECHNOLOGY

Overview of Big Data Issues in the US and Frontiers for Future Research and Development

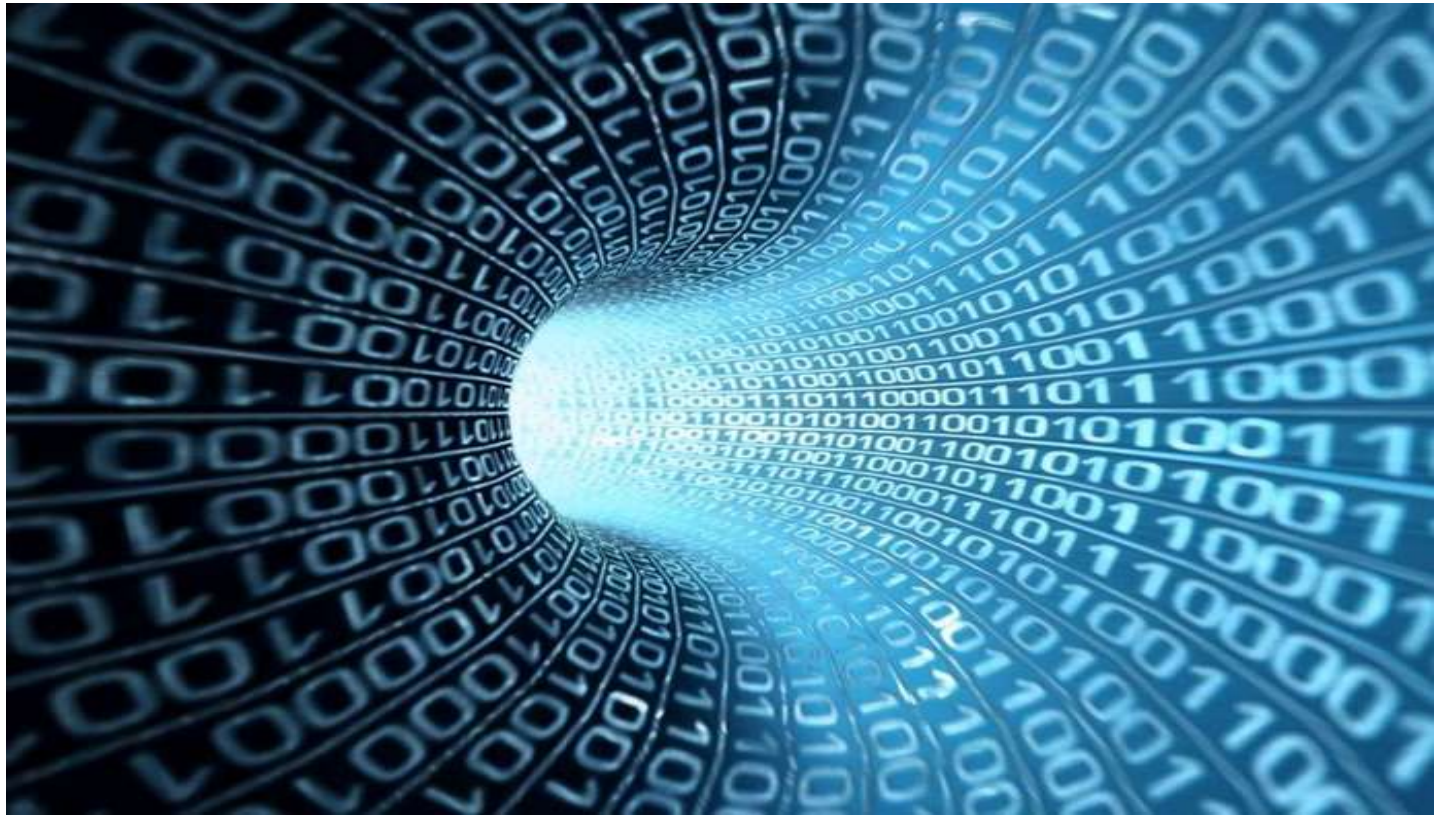
Jonathan Weiner, DrPH &

Hadi Kharrazi, MD, PhD

Our goals for this session are to:

- Help set the stage for this workshop by reviewing key “big data”, population health and social determinants frameworks and paradigms.
- To offer a brief overview of technical and strategic opportunities and challenges related to the application or large electronic database within the context of health and social welfare integration.
- To share a few examples of big data projects we have undertaken at the JHU Center for Population Health IT (CPHIT – “see-fit”) using very large linked digital databases.
- To begin to identify some potential “take aways” for our Israeli colleagues.
- To offer some potential ideas for US/JHU -- Israel/JDC collaboration in this domain.

The Big Data Revolution in Health and Health Care



Health Care and Population Health Digital Ecosystem:

A framework JW first published in the Israeli J of Health Policy Research

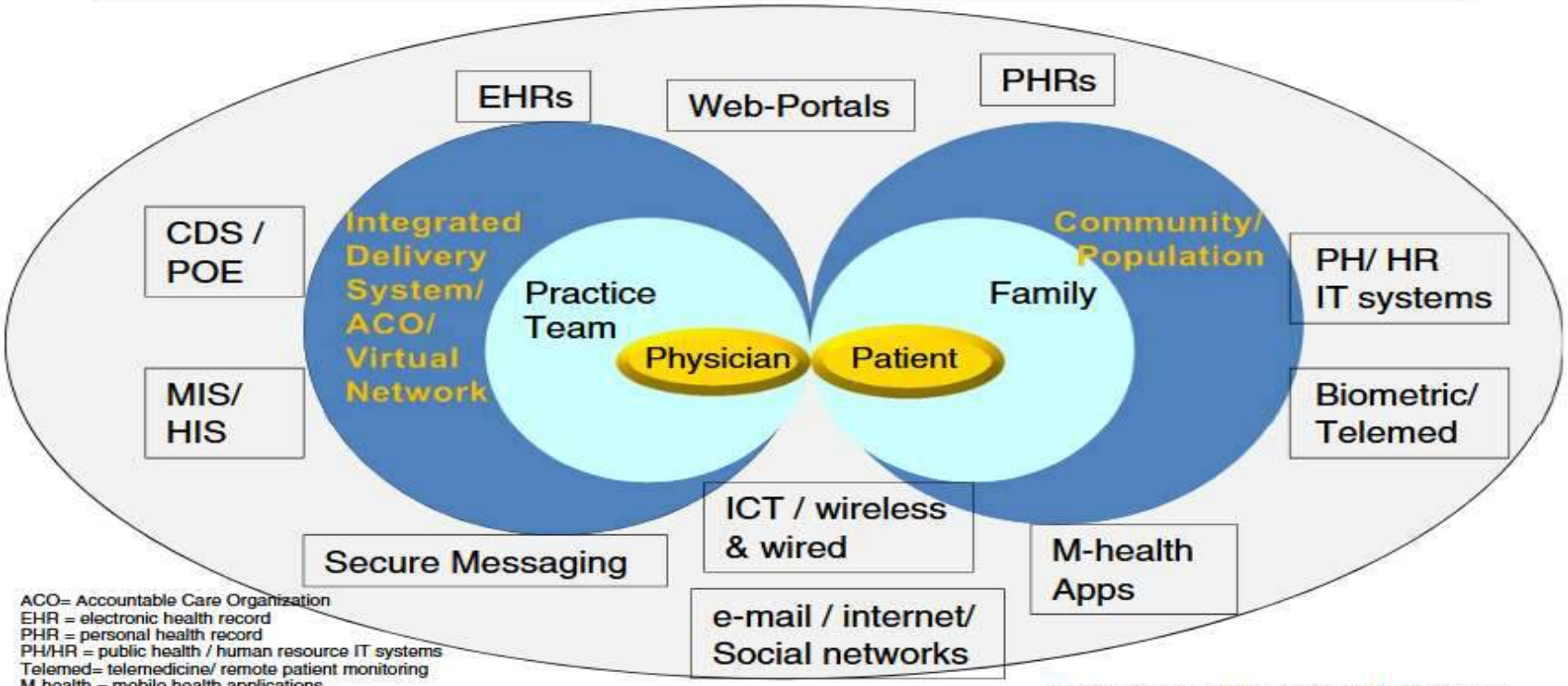


Figure 1 Physician/patient communication in the "e-health" context.

Weiner *Israel Journal of Health Policy Research* 2012, **1**:33
http://www.ijhpr.org/content/1/1/33

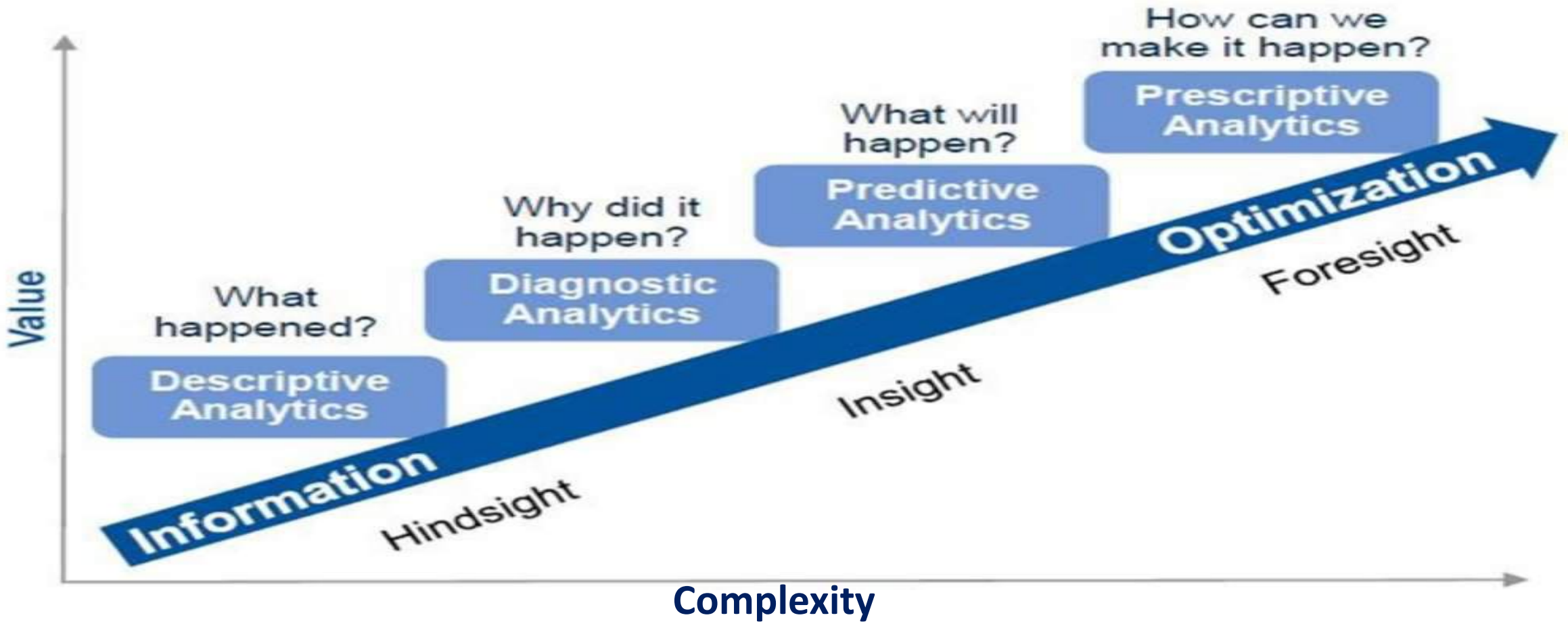
Israel Journal of Health Policy Research

COMMENTARY Open Access

Doctor-patient communication in the e-health era

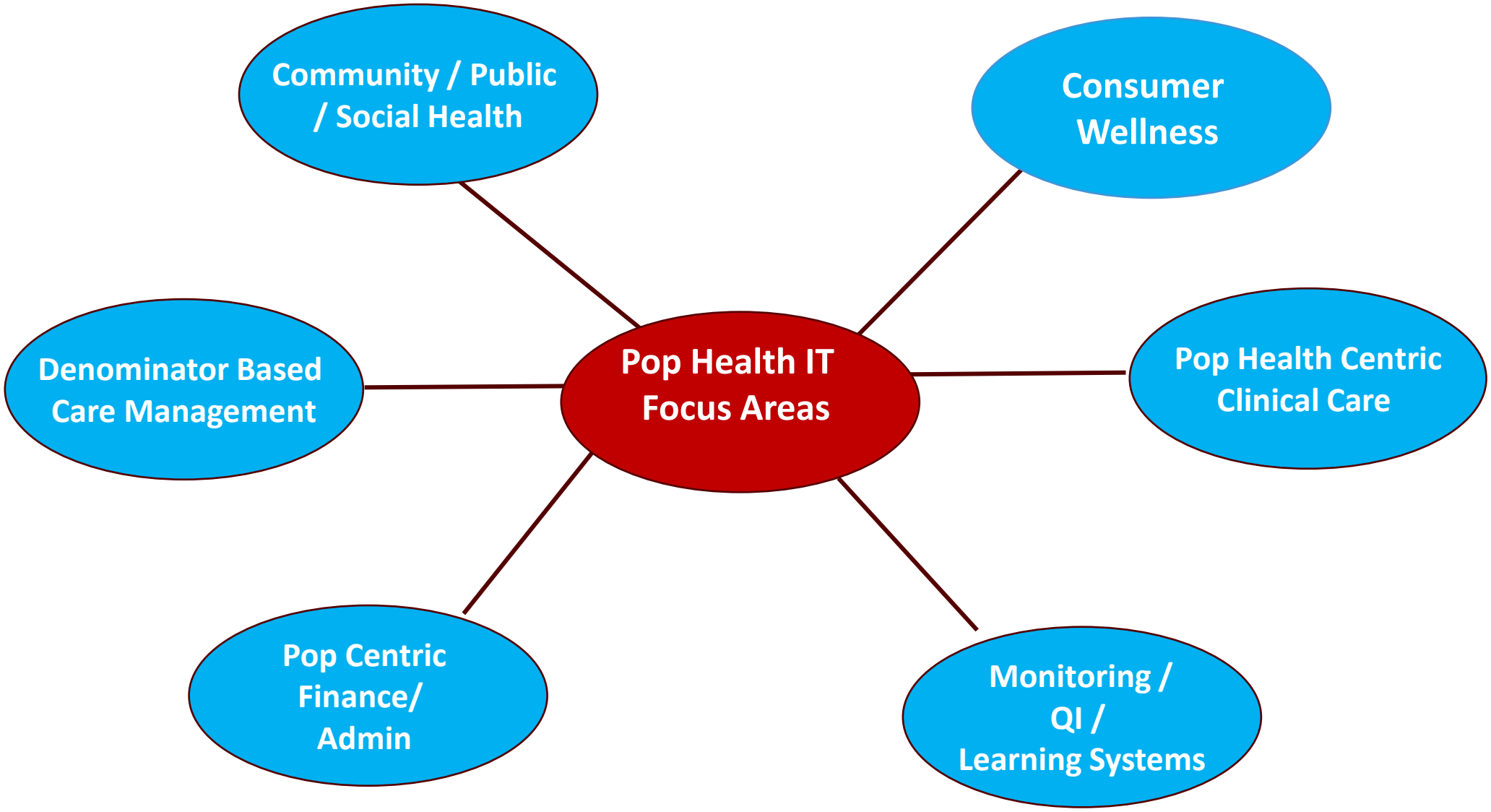
Jonathan P Weiner*

The "Maslow Hierarchy" of Data Analytics Within Organizations and Agencies

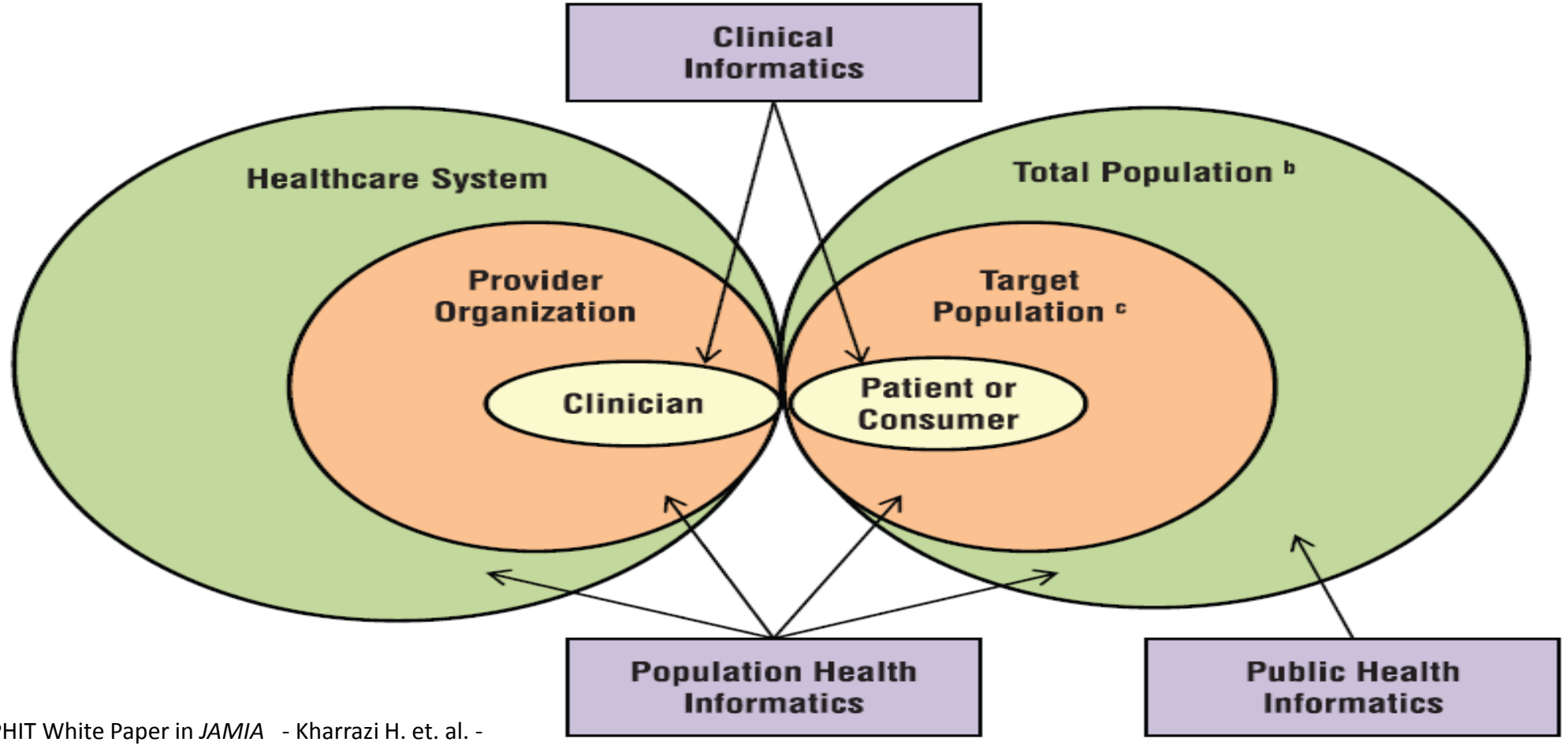


Source of Graphic: Gartner

Key Focus Areas of Big Data IT Supported Applications in Population Health



Population Health Informatics vs. Public Health Informatics & Clinical Informatics



Source: CPHIT White Paper in *JAMIA* - Kharrazi H. et. al. - https://www.ncbi.nlm.nih.gov/pubmed/27018264

Working Definitions

Population Health: A comprehensive framework for assessing and improving the health and well being of a defined population. Population health is practiced by private and public organizations that focus on communities, persons "enrolled" by a health care organization, or other groupings of individuals that comprise a specific cohort of interest.

Public Health: Societal (i.e., governmental) actions to improve health. In the US, the core public health functions relate to assessment, assurance and policy setting.

What is “BIG DATA” in Healthcare

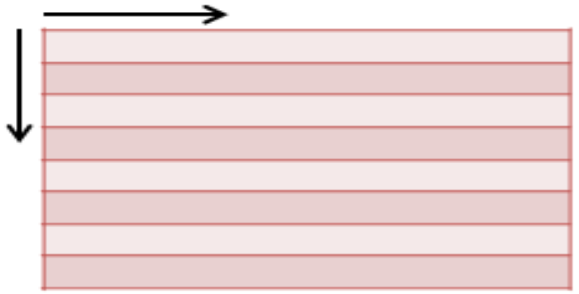
Definition: “Big Data” is a collection of data sets so large and complex that it becomes impractical to process using traditional database tools.

Technology: relational databases; web data; data warehouses; unstructured data; virtualization and parallel processing; in-memory databases; and cloud-based data storage and computing.

Background in the US:

- 1990: The Human Genome Project
- 2008+ Massive roll out of EHRs and Integration of various source of data (HIT/e-health/m-health)
- Healthcare data is expected to grow from ~500 petabytes in 2012 to more than 25,000 petabytes in 2020

The Four "Vs" of Big Data in Health Care

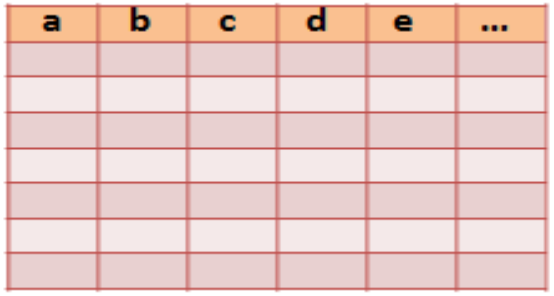


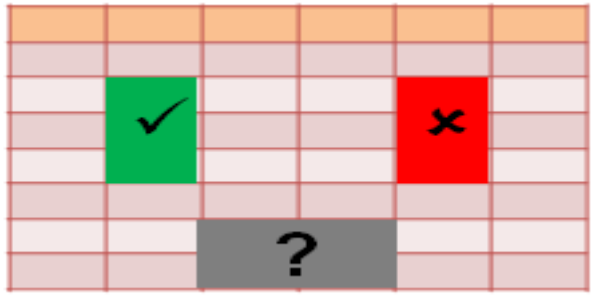
Volume

Velocity

Variety

Veracity



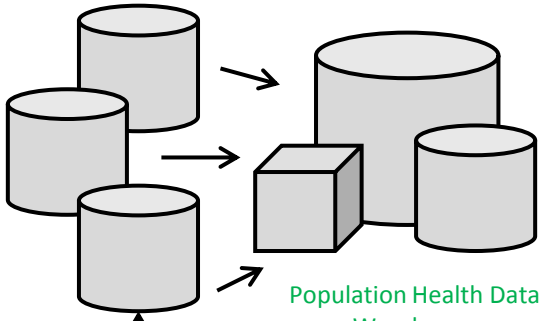


Source: JHU CPHIT – Dr. H. Kharrazi

The Key Stages of Health Analytics With Big Data

1

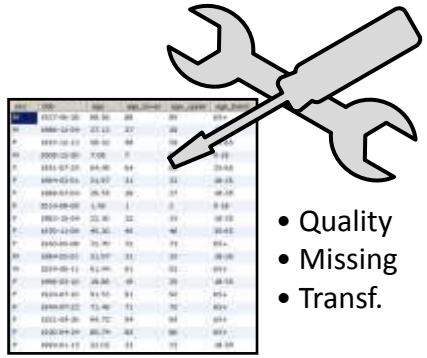
Database Development



Population Health Data Warehouse

2

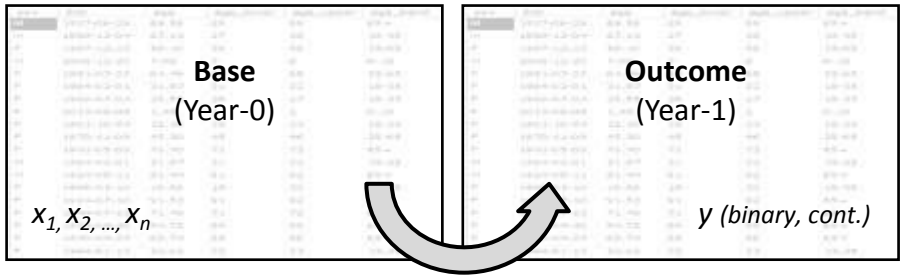
Data Preparation



- Quality
- Missing
- Transf.

3

Modeling and Data Mining



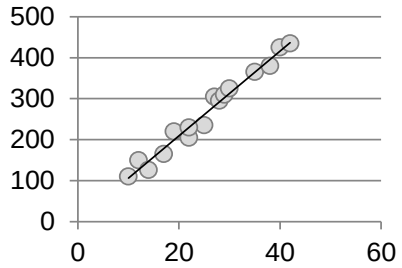
Demographics; Diagnosis; Medications; Cost and etc.

Predict

Cost; Mort.; ER-admit; Hospitalization; Readmit;

4

Model Validation and Evaluation

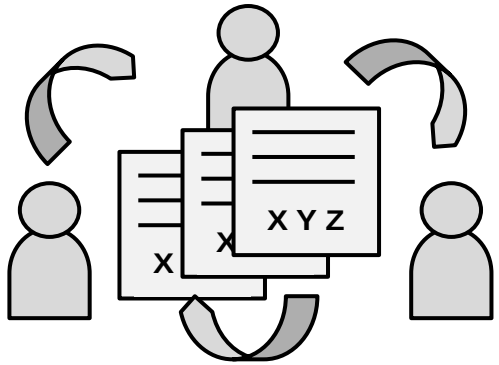


$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$

- Validity
- Reliability
- Goodness of Fit
- Consistency
- Predict Power
- Parsimony
- Reproducibility

5

Use of Knowledge



6

Generate New Data

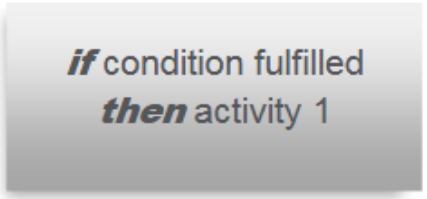
Research and Operational

Source: JHU CPHIT – Dr. H. Kharrazi

What is "Artificial Intelligence" (AI) and "Machine Learning" – Applications to Healthcare

fuzzy boundaries

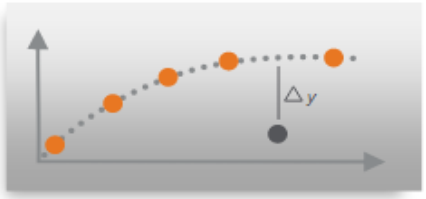
Rules-based Decision Making



Boolean Data
(yes or no)

Health care examples:
- Grouping claims into episodes of care
- Identifying gaps in care
- Identifying fraud

Statistical Reasoning



Numerical Data
allowing for curve fitting

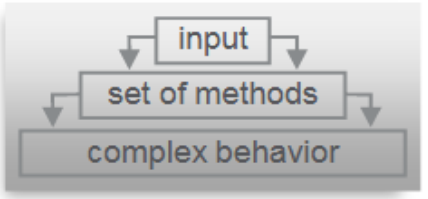
Health care examples:
- Estimating costs to serve a population
- Predicting medical spending for members

Machine Learning

Arbitrary Data
that needs to be abstracted into numbers

Health care examples:
- Identifying patients at risk for readmission
- Identifying patients who are at risk for using the ED inappropriately
- Determining prior authorization for medications

Artificial Intelligence



Arbitrary Data
autonomous selection of best methodology when presented with arbitrary data

Health care examples:
- Recommend "best fit" provider for a member
- Making diagnosis from patient symptoms, physical exam and laboratory values

Source: Paul Bleicher – Optum Labs

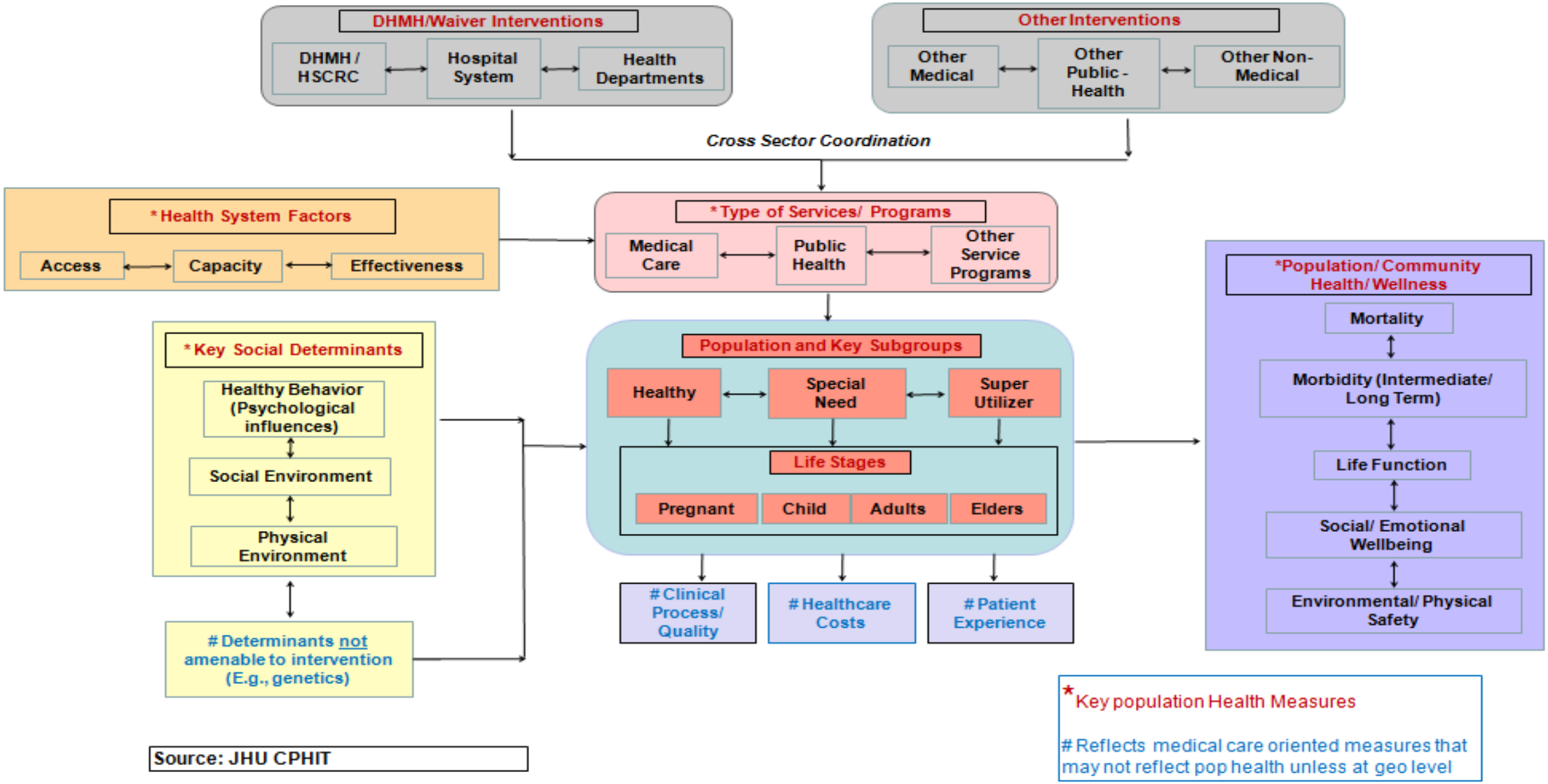
Big Data Challenges in Population Health

- Some - though not all - data relevant to population health are unstructured and “messy” (e.g., clinicians notes and social networks).
- Some data streams (imaging, sensors, genomics) are huge, but most others are reasonably sized (by today’s tech standards)
- Until interoperability (both within and external to care delivery) is surmounted, much data will be missing and difficult to link.
- So called “machine learning” / “AI” is a relatively small part of the solution and at times is oversold by vendors. Logic, evidence and “domain expertise” are still essential.
- Tools to share practical information with humans are key.
- Caveat emptor, there is a lot of hype and confusion out there.

Social Determinants of Health

Graphic: USDHHS – Healthy People 2020

A Conceptual Model for Understanding Community Level Population Health in Maryland



Available at : E. Hatef et al 2017: https://www.ncbi.nlm.nih.gov/pubmed/29035630

A practical framework: amenable "social determinant" factors that can impact health

Economic Stability	Neighborhood and Physical Environment	Education	Food	Community and Social Context	Health Care System
Employment Income Expenses Debt Medical bills Support	Housing Transportation Safety Parks Playgrounds Walkability Zip code / geography	Literacy Language Early childhood education Vocational training Higher education	Hunger Access to healthy options	Social integration Support systems Community engagement Discrimination Stress	Health coverage Provider availability Provider linquistic and cultural competency Quality of care

Health Outcomes
Mortality, Morbidity, Life Expectancy, Health Care Expenditures, Health Status, Functional Limitations

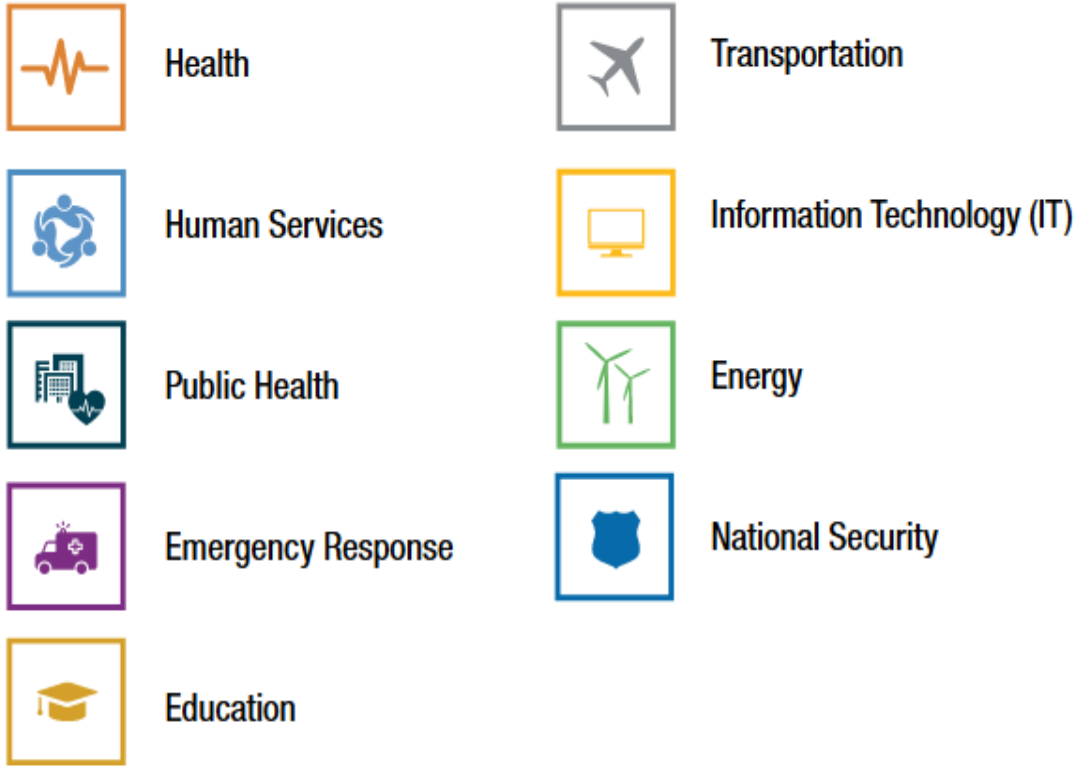
Source: Kaiser Family Foundation

Linking big data across medical and human service organizations:

Key challenges that need to be addressed and participating sectors

Nine Components of Interoperability

Nine Sectors



Source: National Interoperability Collaborative : https://nic-us.org/

The Johns Hopkins Center for Population Health IT (CPHIT *or "see-fit"*)

Mission: To improve the health and well-being of populations by advancing the state-of-the-art of Health IT across public and private health organizations.

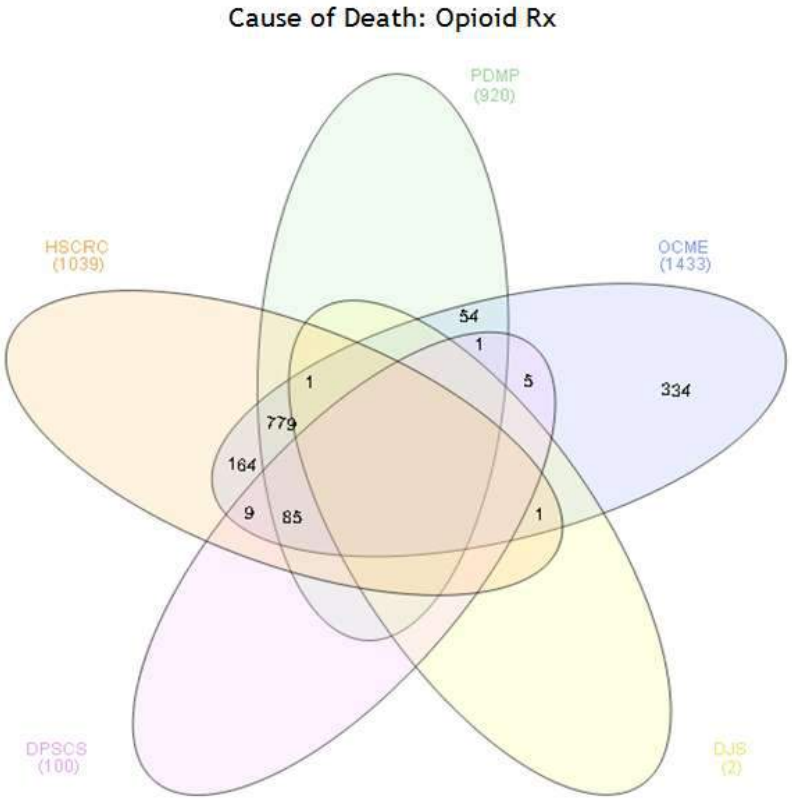
Focus: The application of electronic health records (EHRs), mobile health and other e-health and HIT tools targeted at communities and populations

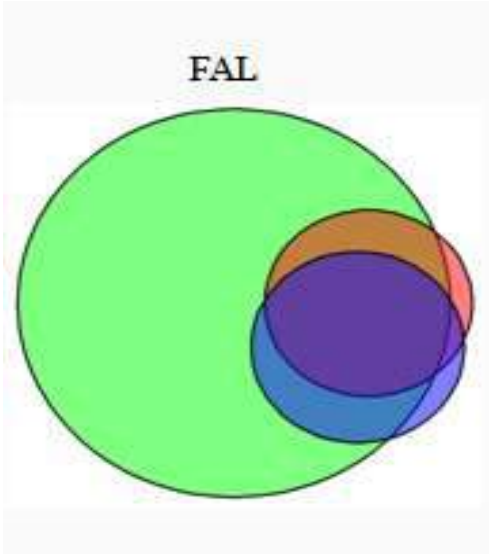
R&D for the Johns Hopkins ACG Predictive Modeling/Risk Adjustment System is based at CPHIT.

www.jhsph.edu/cphit
https://www.hopkinsacg.org/

CPHIT Projects Linking Medical and Social "Big" Data are Starting to Bear Fruit

To address the opioid crisis we are working with one State to link available data across data "silos" (e.g., PDMP/controlled Rx (PDMP), hospitals (HSCRC), coroner (OCME), Police/Corrections (DPSCS), juvenile services (DJS). Goal is to identify persons at risk. Of those who died from opioid RX, about 80% were identifiable from available data.

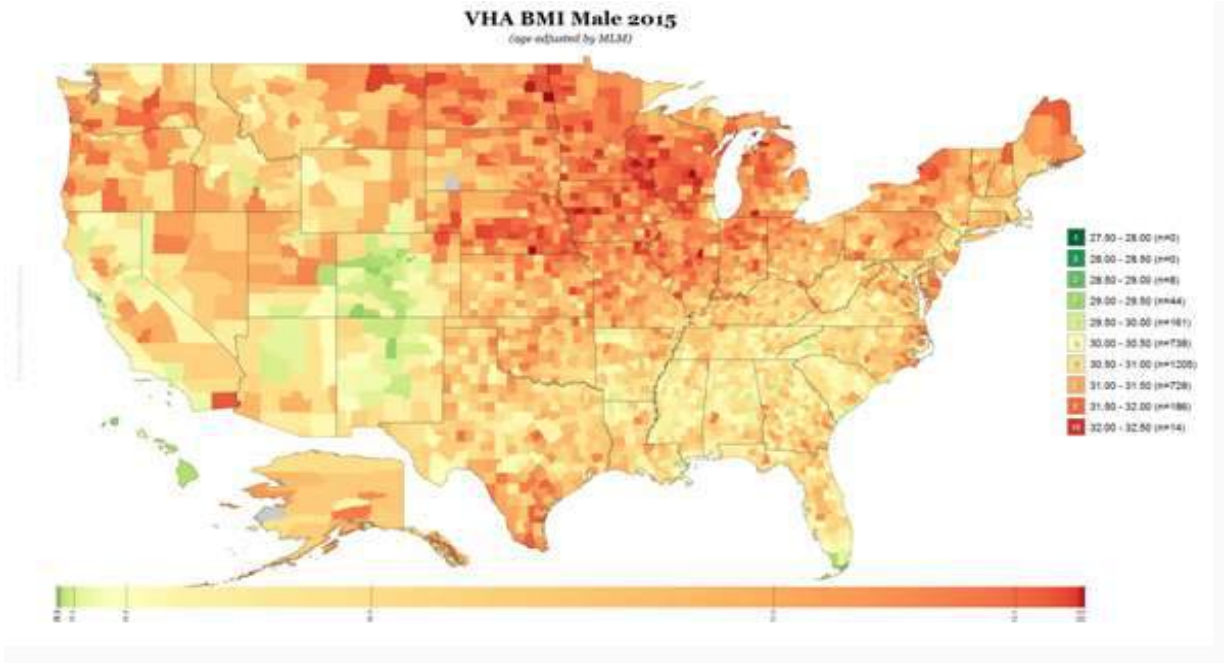




Identifying falls among elderly HMO cohort using NLP of millions of pages of MD/RN notes (green) vs. EMR (blue) & claims (orange)

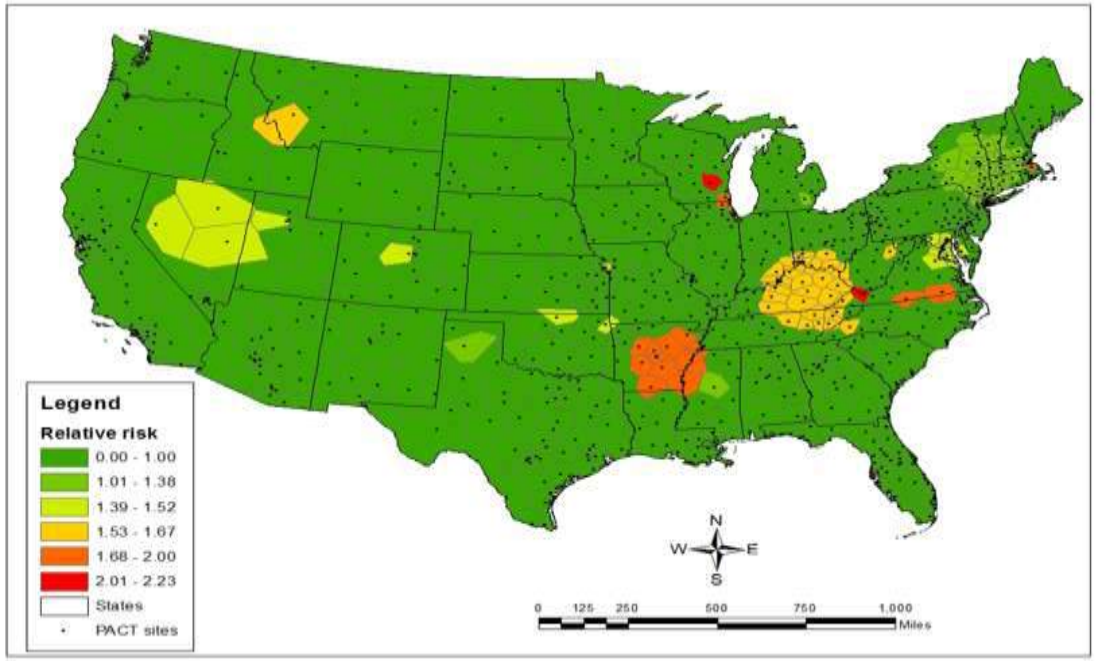
Source: In progress work from JHU CPHIT

CPHIT Big Data Projects are Starting to Bear Fruit – *Cont.*

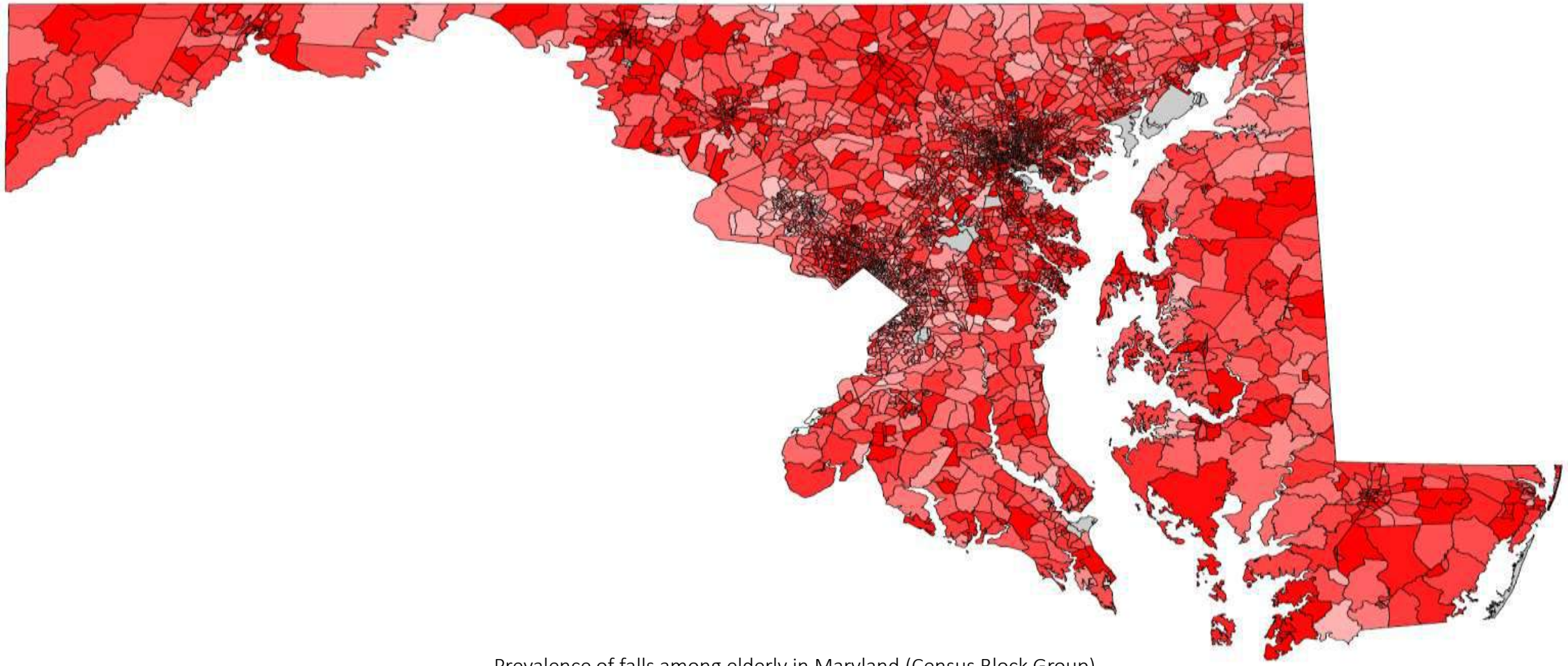


Obesity heat map of US counties based on 20+ M Veteran's BMIs from VA EHR records.

Linking, EHR, geo & social data to identify cohorts with potential hospital overuse within all VA primary care regions considering SES, race, and morbidity.



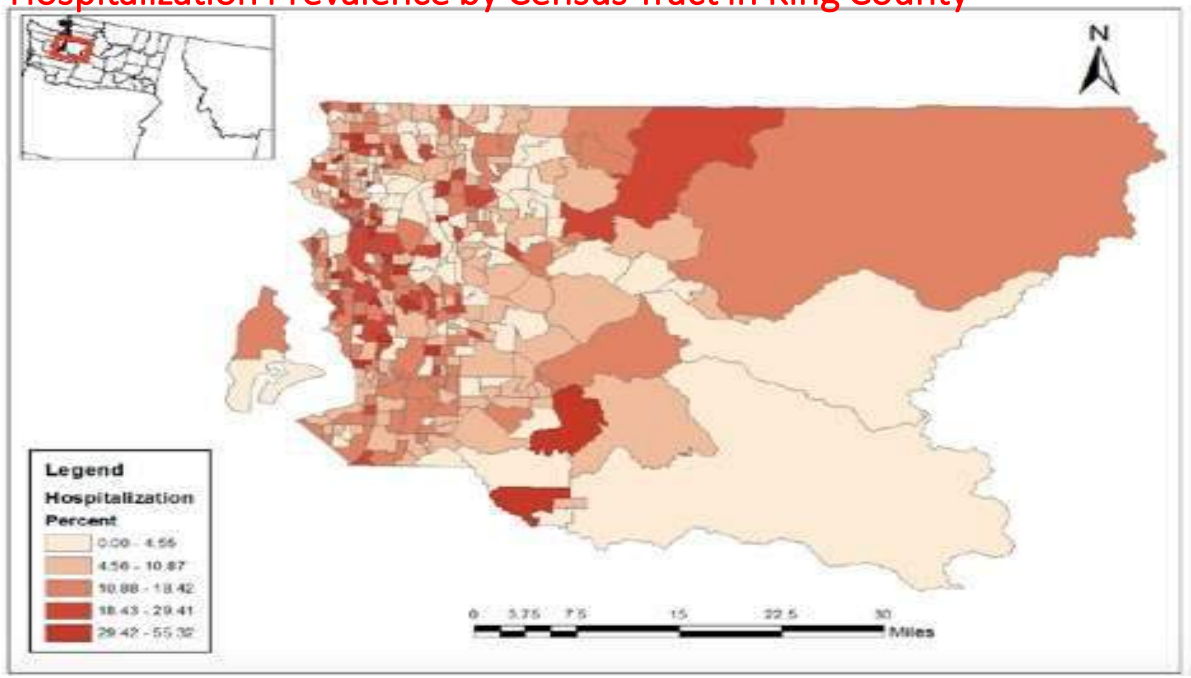
Linking Medical and Geographic Data To Predict Falls among the Older Adults Across Maryland



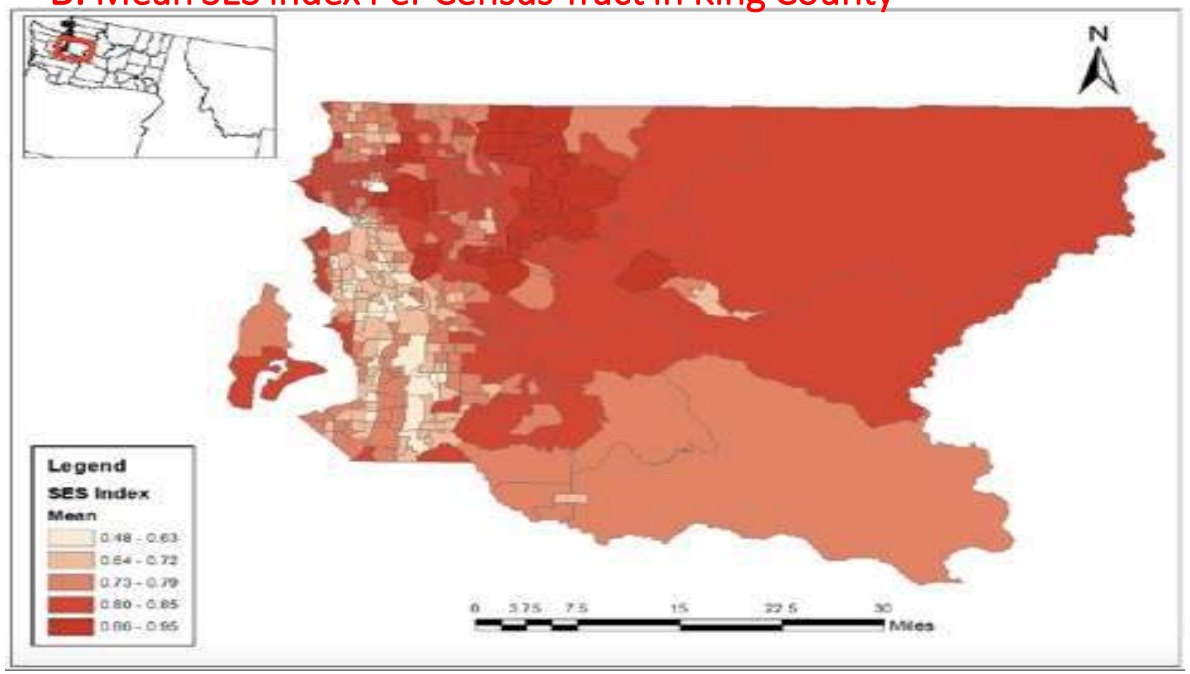
Prevalence of falls among elderly in Maryland (Census Block Group)

More on this project in the Afternoon

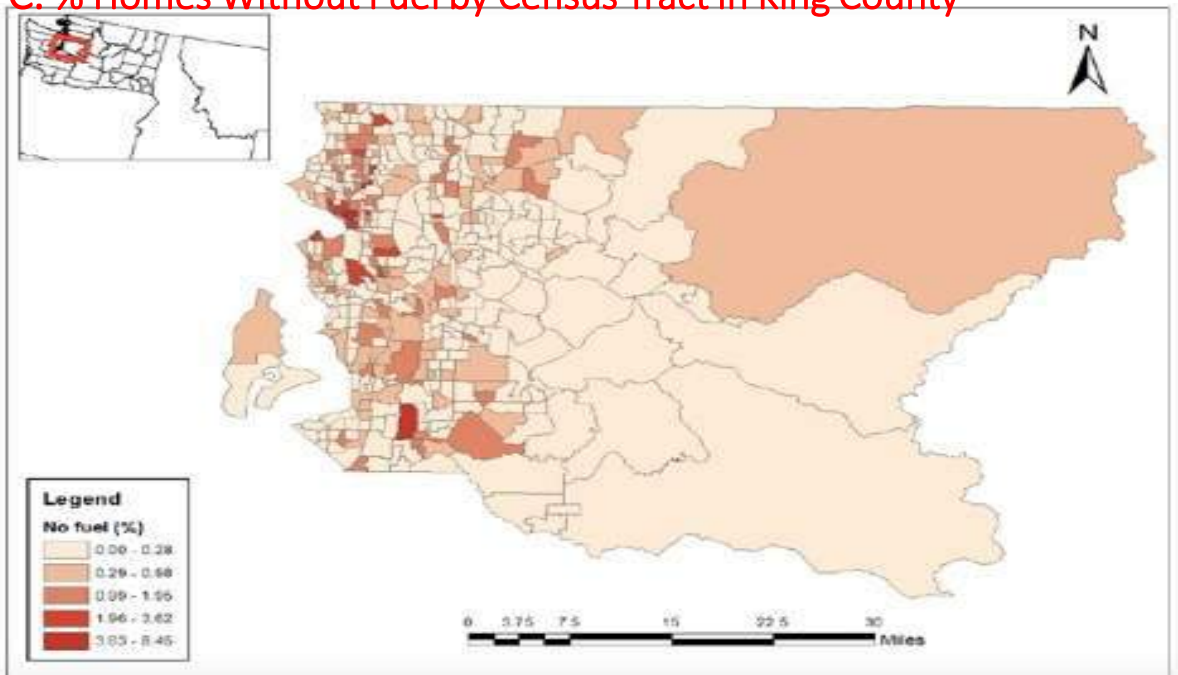
A. Hospitalization Prevalence by Census Tract in King County



B. Mean SES index Per Census Tract in King County



C. % Homes Without Fuel by Census Tract in King County



Comparison of VHA Primary Care Clinics Hospitalization Rates (A), Community Socio-Economic Status (B), and Measure of Housing Stock Quality (C) by Census Tract in Seattle Region

Some Big Data Challenges in Population Health

- Finding ways to integrate and link disparate data and identify "numerators" & "denominators" to define true populations and communities.
- Models and tools to help medical care systems move towards population perspectives and to better integrate social determinants.
- Advanced tools for extracting and analyzing unstructured data.
- Standards and interoperability frameworks for integrating across EHR and other IT vendors to achieve true community standards.

Big Data Challenges – *Cont.*

- New policy/legal frameworks and financial structures that support data integration.
- Privacy, confidentiality and security protections (and the consumer concerns associatde with this area).
- Closer collaboration between government, providers, payers/regulators, IT industry and academia.

Some Possible Lessons/Opportunities within US

- Many available tools and methods in analytics domain.
- Many innovations in academia, industry & various levels of government.
- Well developed population health analytics for those with insurance.
- Resurgence of interest within the “Social Determinants of Health” domain.
- As usual, the U.S. is so large and diverse you can find many excellent successful models (and many lessons of situations to avoid).

Some Potential Ideas for Collaboration

- Our HMO/population based medical care is quite similar and EHRs are now ubiquitous. Building on this commonality could be central.
- We, like you have problems in delivering care equally across all sub-groups. We like you are trying to address disparities. This could be a focus area.
- We both have advanced technologies, There could be synergies in developing standards and frameworks for integrating across EHR / IT vendors to achieve true community standards.
- Most US big data companies have large presence in Israel, they could collaborate and help support.
- We should think “big”, but start “small”.

Shadmi et al. BMC Public Health 2011, 11:609
http://www.biomedcentral.com/1471-2458/11/609

BMC Public Health

RESEARCH ARTICLE Open Access

Assessing socioeconomic health care utilization inequity in Israel: impact of alternative approaches to morbidity adjustment

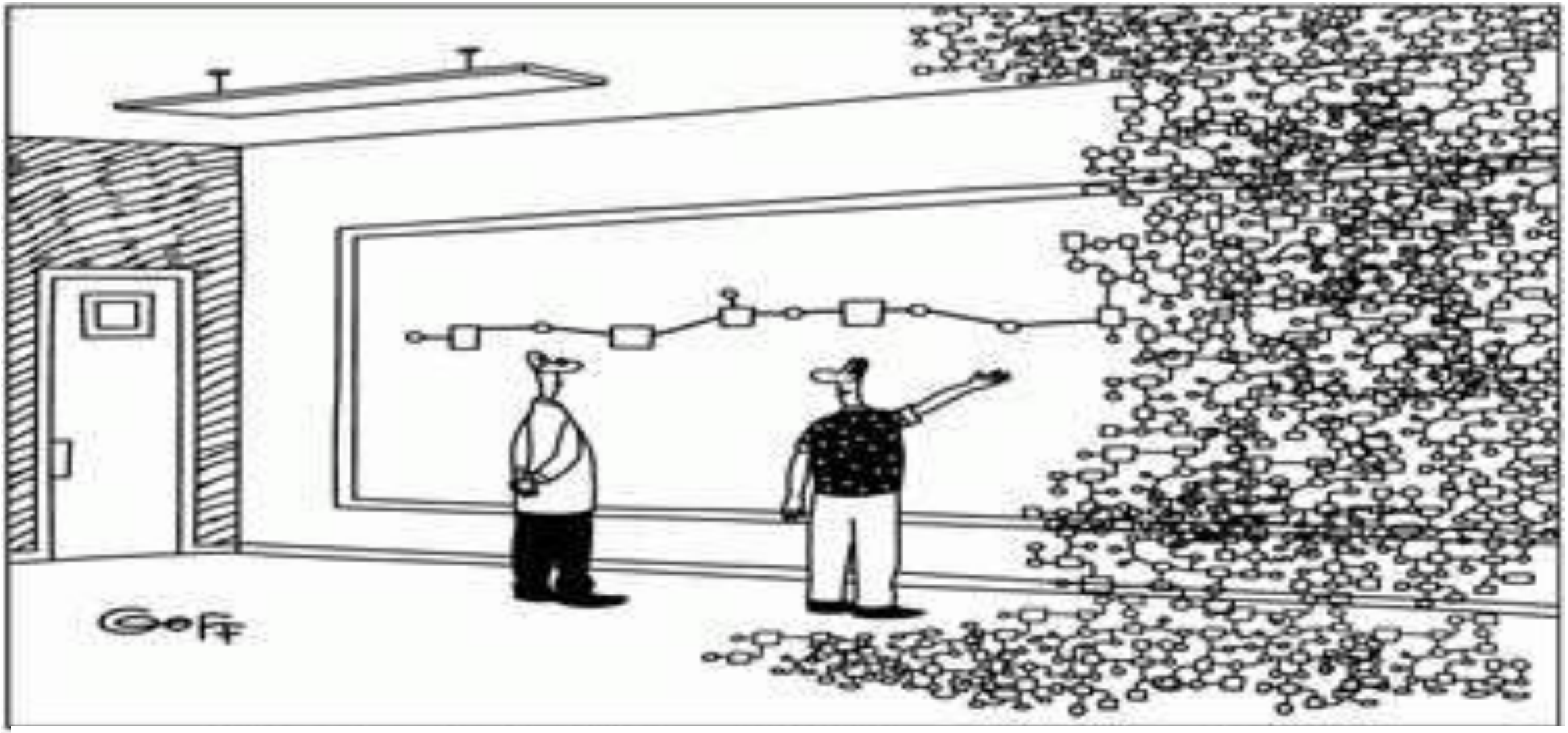
Efrat Shadmi[1,2†], Ran D Balicer[2,3*†], Karen Kinder[4], Chad Abrams[4] and Jonathan P Weiner[4]

Previous JHU/ Collaboration with Clalit "HMO" in Israel Using their Advanced Electronic Health Records

Table 3 Percent with high service use by socioeconomic status*

	Adults with Social Security Waiver	All other adults
Above average number of primary care visits	63%	34%
Above average number of specialist visits	42%	31%
Above average number of diagnostic tests	38%	28%
One or more hospitalizations	16%	7%

* p-value from chi square tests: $p < 0.001$ for all comparisons



This is where the idea of BIG DATA in health and social services gets a little complicated

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

CENTER FOR POPULATION HEALTH
INFORMATION TECHNOLOGY

For more information:

Prof. Jonathan Weiner
jweiner1@jhu.edu

Dr. Hadi Kharrazi
kharrazi@jhu.com

www.jhsph.edu/cphit

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH