Population Health Informatics

Challenges, Opportunities, and Case Studies

Hadi Kharrazi, MHI MD PhD

Johns Hopkins University
School of Public Health – Department of Health Policy and Management
School of Medicine – Division of Health Sciences and Informatics

kharrazi@jhu.edu
Overview

- JHU CPHIT
  - JHSoM
  - JHSPH
  - CPHIT

- Population Health Informatics
  - Emerging Field
  - Data Sources
  - Data Types
  - Data Spectrum
  - Data Analytic
  - Drivers (incentives; mandates; standards)
  - Results of the National Population Health Informatics Workshop

- CPHIT Portfolio
  - EHR-based Utilization Prediction (eACG)
  - Predicting Elderly Falls
  - Geriatric Frailty & EHRs
  - VHA Obesity Trajectory
  - Population Health Metrics
  - Pop Health IT Curriculum and Workforce Training

- Discussion
  - Challenges & Opportunities
Johns Hopkins University Center for Population Health IT (CPHIT)
JHU → JHSoM (School of Medicine)

- Johns Hopkins, the Quaker merchant, banker and businessman, left $7 million in 1873 to create The JHU/Hospital to create new models and standards for medical education and health care.

- JHU/JHM/JHSoS
  - 6 hospitals + 35 outpatient clinics
  - $8 billion in operating revenues
  - 2800 full time faculty + 1200 part time faculty (40k with staff members)
  - 2.8 million-plus annual outpatient visits
  - 360,000-plus annual ED visits
  - 115,000-plus annual hospital admissions
  - $2 billion in total research and development spending based on NSF rankings (highest for 35 years) → $420 mil NIH grants to SoM
  - Managed care health plan: JHHC (368k lives)
JHU → JHSPH (School of Public Health)

- **JHU/JHSPM**
  - Johns Hopkins Bloomberg School of Public Health (founded 1916)
  - 670 faculty + 709 part time
  - Research in 130 countries
  - Research budget: $500+ mil
  - Degrees: 12 graduate including the largest MPH
  - Centers: 60+
  - Highlights:
    - First School of Public Health in the US
    - Largest school of public health in the world
    - Receives 20 percent of all grants and contracts awarded to the 57 accredited U.S. schools of public health
    - Ranked No. 1 by U.S. News & World Report since 1994
The Johns Hopkins Center for Population Health Information Technology
(CPHIT, or “see-fit”)

- The **mission** of this innovative, multi-disciplinary R&D center is to improve the health and well-being of populations by advancing the state-of-the-art of Health IT across public and private health organization.

- CPHIT **focuses** on the application of electronic health records (EHRs), mobile health and other e-health and HIT tools targeted at communities and populations.

- Director: Dr. Weiner

- Research Director: Dr. Kharrazi

www.jhsph.edu/cphit
Population Health Informatics
Population Health Informatics → Emerging Field

Triple Aims developed by the Institute for Healthcare Improvement (IHI)
Population Health Informatics → Emerging Field

Biomedical informatics methods, techniques, and theories

Population HIT

Bioinformatics

Imaging Informatics

Clinical Informatics

Consumer Health Informatics

Molecular Research

Health Research

Biomedical informatics as a basic science
Population Health Informatics → Data Sources

- Web Portals
- Public Health Systems
- IDS / ACO / Virtual Net
- Community / Population
- Practice Team
- Family and Care givers

- Physician
- Patient

- National Datasets
- HIE
- Claims MIS HIS
- EHR
- CPOE CDSS
- PHR
- mHealth apps
- Social HR data GIS
- Social Network
- Biomet. Tele-H.
- email and others
- email and others

Weiner, 2012  http://www.ijhpr.org/content/1/1/33
Population Health Informatics → Data Types

Outcomes
- Mortality
- Quality of Life
- Disparity
  - Race
  - Ethnicity
  - Gender
  - SES
  - Geography

Determinant
- Health Care
- Individual Behavior
- Social Environment
- Physical Environment
- Genetics

Population Health Data Types
- Clinical (e.g., claims, EHR, HIE)
- Surveys (e.g., HRA, ADL)
- Social data (e.g., SES)
- GIS-bound (e.g., census block)
- Genotyping

Health Determinants and Population Health Data Types
Population Health Informatics → Data Spectrum

Variety and Continuum of Information Systems in PopHI
Population Health Informatics → Data Analytic Cycle

**Population Health Database Development**
- Population Health Data Warehouse

**Data Preparation & Data Quality Checks**
- Quality
- Missing
- Transf.

**Extracting Knowledge by Modeling and Data Mining**
- **Base** (Year-0)
- **Outcome** (Year-1)
  - Demographics; Diagnosis; Medications; Cost and etc.
  - Cost; Mort.; ER-admit; Hospitalization; Readmit;
  - \( y (\text{binary, cont.}) \)

**Generating & Integrate New Data from Knowledge**
- Research and Operations

**Store, Share and Use the Knowledge**
- \( \hat{y} = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \)

**Creating Generalizable Knowledge by Model Validation and Evaluation**
- Validity
- Reliability
- Goodness of Fit
- Consistency
- Parsimony
- Reproducibility

**Overall Population Health Knowledge Management Process**
Population Health Informatics \rightarrow Drivers

- **Incentives:**

  ARRA / HITECH \rightarrow
  
  (1) EHR adoption [Meaningful Use (MU) measures];
  (2) state-wide HIEs;
  (3) Beacon Communities

- **Mandates:**

  ACA \rightarrow Payment Reforms \rightarrow PCMH and ACO initiatives \rightarrow Value-based \rightarrow Capitated models \rightarrow Population health + MACRA

- **Facilitation:**

  ONC \rightarrow Data Standards, integration, and sharing \rightarrow Distributed models
Population Health Informatics → Drivers → EHR adoption

HIT-Enabled Health Reform

2009 HITECH Policies

2011 Criteria (Capture & Share)

2013 Criteria (Advanced Care)

2015 Criteria (Improved Outcome)

Drivers → ARRA → HITECH / MU of EHRs + HIEs
Population Health Informatics → Drivers → EHR adoption (cont.)

Percentage of EHR systems among office-based physicians
Big Data and Healthcare → Drivers → Value-base Care (ACOs) (cont.)

ACO Management

Provider
Hospitals
Inpatient

Provider
Ambulatory
Outpatient

Para
Pharm
Lab
Imaging

Patient
Population
Community
Public Health

Payer
Private
Public

© Hadi Kharrazi @ JHSPH-HPM
Population Health Informatics → Drivers → Data Standards
A proposed national research and development agenda for population health informatics: summary recommendations from a National Expert Workshop

Hadi Kharrazi1,2, Elyse C Lasser1, William A Yasnoff2,3, John Loonsk1, Annel Advani1, Harold P Lehmann2, David C Chin1, Jonathan P Weiner1,2

ABSTRACT

Objective The Johns Hopkins Center for Population Health IT hosted a 1-day symposium sponsored by the National Library of Medicine to help develop a national research and development (R&D) agenda for the emerging field of population health informatics (PopHL).

Material and Methods The symposium provided a venue for national experts to brainstorm, identify, discuss, and prioritize the top challenges and opportunities in the PopHL field, as well as R&D areas to address these.

Results This manuscript summarizes the findings of the PopHL symposium. The symposium participants’ recommendations have been categorized into 13 overarching themes, including policy alignment, data governance, sustainability and incentives, and standards/interoperability.

Discussion The proposed consensus-based national agenda for PopHL consisted of 18 priority recommendations grouped into 4 broad goals: (1) Developing a standardized collaborative framework and infrastructure, (2) Advancing technical tools and methods, (3) Developing a scientific evidence and knowledge base, and (4) Developing an appropriate framework for policy, privacy, and sustainability. There was a substantial amount of agreement between all the participants on the challenges and opportunities for PopHL as well as on the actions that needed to be taken to address these.

Conclusion PopHL is a rapidly growing field that has emerged to address the population dimension of the Triple Aim. The proposed PopHL R&D agenda is comprehensive and timely, but should be considered only a starting-point, given that ongoing developments in health policy, population health management, and informatics are very dynamic, suggesting that the agenda will require constant monitoring and updating.

JAMIA Paper Reporting on the Findings from the National Workshop
Population health comprises organized activities for assessing and improving the health and well-being of a defined population. Population health is practiced by both private and public organizations. The target “population” can be a specific geographic community or region, or it may represent some other “denominator,” such as enrollees of a health plan, persons residing in a provider’s catchment area, or an aggregation of individuals with special needs. The difference between population health and public health is subtle, and there is not always a full consensus on these definitions. That said, public health services are typically provided by government agencies and include the “core” public health functions of health assessment, assurance, and policy-setting. In the United States, most actions of public health agencies represent population health, but a considerable proportion, if not the majority, of population health services are provided by private organizations.
Population Health Informatics → National Workshop on Pop Health IT (cont.)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Population Health Informatics</th>
<th>Public Health Informatics</th>
<th>Clinical Informatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Intervention Targets</td>
<td>• Total population&lt;br&gt;• Target populations&lt;br&gt;• Provider organization&lt;br&gt;• Healthcare systems</td>
<td>• Total population</td>
<td>• Clinician&lt;br&gt;• Patient or consumer&lt;br&gt;• Provider organization&lt;br&gt;• Target population</td>
</tr>
<tr>
<td>Main Operational Goal</td>
<td>• Outreach and prevention&lt;br&gt;• Care integration&lt;br&gt;• Disease management</td>
<td>• Assessment&lt;br&gt;• Prevention</td>
<td>• Treatment&lt;br&gt;• Rehabilitation</td>
</tr>
<tr>
<td>Action Arm</td>
<td>• Population health organization&lt;br&gt;• Care management organizations</td>
<td>• Public health agencies&lt;br&gt;• Non-for-profit and non-governmental organizations</td>
<td>• Clinical organizations</td>
</tr>
<tr>
<td>Key Stakeholders</td>
<td>• Provider and payer systems&lt;br&gt;• Government and community</td>
<td>• Federal, state, and local governments</td>
<td>• Providers&lt;br&gt;• Consumers</td>
</tr>
<tr>
<td>Key Information Challenges</td>
<td>• Capturing non-medical info&lt;br&gt;• Information system interoperability across sectors</td>
<td>• Expanding public health IT systems&lt;br&gt;• Medical and public health interoperability</td>
<td>• Decision support&lt;br&gt;• EHR interoperability</td>
</tr>
</tbody>
</table>

Population Health Informatics vs Public Health Informatics vs Clinical Informatics
*(table of contexts and differences)*
Population Health Informatics → National Workshop on Pop Health IT (cont.)

Population Health Informatics vs Public Health Informatics vs Clinical Informatics
(Venn diagram of population denominator)
Population Health Informatics → Growing domain...
CPHIT Research Portfolio
CPHIT Portfolio (cont.)

- **Research Portfolio (selected list)**

  - **EHR-based Utilization Prediction (eACG):** Developing a wide range of EHR-based population focused predictive modeling tools
    - EHR ICD and e-prescribing Rx input (coordinated with claims)
    - Non-claims data: Lab Data; BMI and Vitals; and Social History

  - **Geriatric Frailty:** Developing a new geriatric/frailty “e-risk” score utilizing structured and unstructured EHR data and Claims

  - **Predicting Elderly Falls:** Developing analytics with Baltimore City Heath Department for regional collaboration to identify and predict elder’s fall injuries in the community using social, medical, and public health data
CPHIT Portfolio (cont.)

- **VHA Obesity**: Collaborating with the Veteran Health Administration
  - Developed population health analytic framework
  - Linking geo and social data for obesity trend analysis
  - Adding social/geo framework to PCMH (PACT) case finding program

- **Pop e-Measures**: Collaborating with Maryland State Health Department, HIE and Hospital Commission
  - Assist in the statewide pop health digital measurement infrastructure
  - Develop population health focused measures

- **Opioid**: Using multiple novel sources of data to develop predictive models to identify persons at risk for opioid overdose

- **Consumer Data**: Linking consumer/marketing data with medical data to identify health outcomes (e.g., consumer reports)
ACG system offers a unique approach to measuring morbidity that improves accuracy and fairness in evaluating provider performance, identifying patients at high risk, forecasting healthcare utilization and setting equitable payment rates.
Billions of dollars per year are now routinely exchanged using ACGs in almost every US State and in 16 + nations. Healthcare of as many as 100+ million patients is actively managed and monitored using ACGs on all continents. Over 700+ peer reviewed articles have been published that apply and evaluate ACGs.

Other EHR-only data: Lab results; Vital Signs; Medical and Family history...

New applications: Real time population health / community surveillance; Real time clinical action for individual consumer; Functional Status / Frailty
### CPHIT Portfolio →
EHR-based Utilization Prediction (eACG) (cont.)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Claims</th>
<th>EHR[^1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Reimbursement</td>
<td>Clinical care</td>
</tr>
<tr>
<td>Scope</td>
<td>All providers, including out of network providers, for a given patient</td>
<td>Network providers of a patient</td>
</tr>
<tr>
<td>Data consistency</td>
<td>High consistency across sources</td>
<td>Lower consistency across sources</td>
</tr>
<tr>
<td>Data structure</td>
<td>Most of data is structured</td>
<td>Considerable unstructured data</td>
</tr>
<tr>
<td>Coding standard</td>
<td>Strict adherence to coding systems</td>
<td>Variable adherence to coding systems</td>
</tr>
<tr>
<td>Provider coverage</td>
<td>All providers accepting the insurance</td>
<td>Limited to providers using same EHR</td>
</tr>
<tr>
<td>Coding limit</td>
<td>Limited to encoded data</td>
<td>Provides ability to enter free text</td>
</tr>
<tr>
<td>Member limitation</td>
<td>Limited to insured patients</td>
<td>Insured and uninsured patients</td>
</tr>
<tr>
<td>Coverage limitation</td>
<td>Non-covered items are missing</td>
<td>Includes data on non-covered items</td>
</tr>
<tr>
<td>Data type</td>
<td>Limited (mainly enrollment, Dx, Rx)</td>
<td>Additional data types (see below)</td>
</tr>
</tbody>
</table>

### Data Availability

<table>
<thead>
<tr>
<th>Data Availability</th>
<th>Claims</th>
<th>EHR[^1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics[^2]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Limited</td>
<td>Limited</td>
</tr>
<tr>
<td>Diagnosis[^3]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Procedures</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Eligibility</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>Socioeconomic data</td>
<td>Zip-code derived</td>
<td>Coded and zip-code derived</td>
</tr>
<tr>
<td>Family history</td>
<td>Not available</td>
<td>Yes</td>
</tr>
<tr>
<td>Problem list</td>
<td>Not available</td>
<td>Yes</td>
</tr>
<tr>
<td>Procedure results</td>
<td>Not available</td>
<td>Yes</td>
</tr>
<tr>
<td>Laboratory results</td>
<td>Not available</td>
<td>Yes</td>
</tr>
<tr>
<td>Vital signs</td>
<td>Not available</td>
<td>Yes</td>
</tr>
<tr>
<td>Behavioral risk factors</td>
<td>Not available</td>
<td>Limited</td>
</tr>
<tr>
<td>Standardized surveys</td>
<td>Limited</td>
<td>Limited</td>
</tr>
</tbody>
</table>

**Claims vs EHR**
Comparing Population-based Risk-stratification Model Performance Using Demographic, Diagnosis and Medication Data Extracted From Outpatient Electronic Health Records Versus Administrative Claims

Background: There is an increasing demand for electronic health record (EHR)-based risk stratification and predictive modeling tools at the population level. This trend is partly due to increased value-based payment policies and the increasing availability of EHRs at the provider level. Risk stratification models, however, have been traditionally derived from claims or encounter systems. This study evaluates the challenges and opportunities of using EHR data instead of or in addition to administrative claims for risk stratification.

Discussion: The results show a promising performance of models predicting cost and hospitalization using outpatient EHR’s diagnosis and medication data. More research is needed to evaluate the benefits of other EHR data types (eg, lab values and vital signs) for risk stratification.

Key Words: risk stratification, predictive modeling, electronic health records, administrative claims, The Johns Hopkins ACG System

(Med Care 2017;55: 789–796)

Comparing Claims and EHR for Risk Stratification
Comparing Various Overlaps of Claims and EHR for Risk Stratification
Comparing Diagnostic Data Found in Claims (C) vs EHRs (E)
### TABLE 2. Comparison of Case Findings of Selected Conditions Based on Diagnostic and Medication Information Extracted From Administrative Claims, EHR Data, and a Combination of Both

<table>
<thead>
<tr>
<th>Measure</th>
<th>N (%) Population</th>
<th>Overlaps (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Claims</td>
<td>EHR</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Dx (EDC)</td>
<td>4047 (4.73)</td>
</tr>
<tr>
<td></td>
<td>Rx (RxMG)</td>
<td>3878 (4.53)</td>
</tr>
<tr>
<td></td>
<td>Both (and)</td>
<td>3364 (3.93)</td>
</tr>
<tr>
<td></td>
<td>Either (or)</td>
<td>4561 (5.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,152 (11.86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13,824 (16.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9232 (10.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14,744 (17.23)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Dx (EDC)</td>
<td>5451 (6.37)</td>
</tr>
<tr>
<td></td>
<td>Rx (RxMG)</td>
<td>15,814 (18.48)</td>
</tr>
<tr>
<td></td>
<td>Both (and)</td>
<td>4425 (5.17)</td>
</tr>
<tr>
<td></td>
<td>Either (or)</td>
<td>16,840 (19.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2458 (2.87)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>806 (0.94)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>478 (0.56)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2786 (3.26)</td>
</tr>
<tr>
<td>Depression</td>
<td>Dx (EDC)</td>
<td>4047 (4.73)</td>
</tr>
<tr>
<td></td>
<td>Rx (RxMG)</td>
<td>3878 (4.53)</td>
</tr>
<tr>
<td></td>
<td>Both (and)</td>
<td>3364 (3.93)</td>
</tr>
<tr>
<td></td>
<td>Either (or)</td>
<td>4561 (5.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,152 (11.86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13,824 (16.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9232 (10.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14,744 (17.23)</td>
</tr>
<tr>
<td>Cancer</td>
<td>Dx (EDC)</td>
<td>5451 (6.37)</td>
</tr>
<tr>
<td></td>
<td>Rx (RxMG)</td>
<td>15,814 (18.48)</td>
</tr>
<tr>
<td></td>
<td>Both (and)</td>
<td>4425 (5.17)</td>
</tr>
<tr>
<td></td>
<td>Either (or)</td>
<td>16,840 (19.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2458 (2.87)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>806 (0.94)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>478 (0.56)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2786 (3.26)</td>
</tr>
</tbody>
</table>

Cases found in EHR versus Claims:
Diabetes, Hypertension, Depression, Cancer
### CPHIT Portfolio → EHR-based Utilization Prediction (eACG) (cont.)

#### TABLE 4. Comparison of Concurrent and Prospective Performance (Area Under the Curve) for Hospitalization and Top 1% Cost Using Claims, EHRs Data, or Both

<table>
<thead>
<tr>
<th>Outcome†</th>
<th>Year</th>
<th>Data Source and Model Variables*</th>
<th>Dem.</th>
<th>Claims</th>
<th>EHR</th>
<th>Claims or EHR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sex + Age</td>
<td>Dx</td>
<td>Rx</td>
<td>DxRx</td>
<td>Dx</td>
</tr>
<tr>
<td>Hosp.‡</td>
<td>Concurrent</td>
<td>0.640</td>
<td>0.864</td>
<td>0.804</td>
<td>0.875</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>Prospective</td>
<td>0.656</td>
<td>0.779</td>
<td>0.736</td>
<td>0.783</td>
<td>0.732</td>
</tr>
<tr>
<td>Top 1% cost§</td>
<td>Concurrent</td>
<td>0.649</td>
<td>0.925</td>
<td>0.864</td>
<td>0.938</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>Prospective</td>
<td>0.666</td>
<td>0.832</td>
<td>0.800</td>
<td>0.835</td>
<td>0.782</td>
</tr>
</tbody>
</table>

Model performance using EHR versus Claims
CPHIT Portfolio → Geriatric Frailty

- Claims
- EHR Free Text (NLP)
- EHR Structured
This study evaluates the added-value of EHR’s unstructured text in identifying geriatric risk factors. Claims and structured EHR data give an incomplete picture about several constructs that influence geriatric risk. There is a high possibility to miss patients with geriatric risk markers when excluding free-text from an analysis.

<table>
<thead>
<tr>
<th>Construct</th>
<th>C%</th>
<th>E%</th>
<th>NLP%</th>
<th>C+E%</th>
<th>C+E+NLP%</th>
<th>E+NLP%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFC</td>
<td>0.52</td>
<td>0.87</td>
<td>2.04</td>
<td>0.93</td>
<td>2.22</td>
<td>2.18</td>
</tr>
<tr>
<td>DEC</td>
<td>0.65</td>
<td>0.45</td>
<td>1.11</td>
<td>0.84</td>
<td>1.47</td>
<td>1.19</td>
</tr>
<tr>
<td>DEM</td>
<td>3.10</td>
<td>3.94</td>
<td>6.36</td>
<td>4.67</td>
<td>7.21</td>
<td>6.82</td>
</tr>
<tr>
<td>FAL</td>
<td>5.48</td>
<td>5.73</td>
<td>22.10</td>
<td>7.32</td>
<td>23.16</td>
<td>22.77</td>
</tr>
<tr>
<td>MAL</td>
<td>0.12</td>
<td>0.03</td>
<td>2.37</td>
<td>0.14</td>
<td>2.44</td>
<td>2.37</td>
</tr>
<tr>
<td>SSN</td>
<td>0.03</td>
<td>0.06</td>
<td>28.02</td>
<td>0.07</td>
<td>28.03</td>
<td>28.03</td>
</tr>
<tr>
<td>URC</td>
<td>0.55</td>
<td>0.97</td>
<td>3.10</td>
<td>1.09</td>
<td>3.38</td>
<td>3.30</td>
</tr>
<tr>
<td>VIS</td>
<td>0.53</td>
<td>0.76</td>
<td>4.81</td>
<td>0.85</td>
<td>4.94</td>
<td>4.92</td>
</tr>
<tr>
<td>WEI</td>
<td>4.68</td>
<td>5.60</td>
<td>9.11</td>
<td>6.65</td>
<td>11.96</td>
<td>11.18</td>
</tr>
<tr>
<td>WLK</td>
<td>8.39</td>
<td>7.88</td>
<td>32.45</td>
<td>10.80</td>
<td>34.93</td>
<td>34.02</td>
</tr>
</tbody>
</table>

Comparing rates of geriatric risk factors using claims and EHR data (structured and free text) within the larger population reference.
Added value of free text represented by the Venn diagram
Circle sizes represent the number of patients identified by each methodology/data-source

- **Green**: EHR Free Text
- **Blue**: EHR Structured
- **Red**: Insurance Claims
CPHIT Portfolio → Predicting Elderly Falls

- Baltimore Falls Reduction Initiative Engaging Neighborhoods & Data (B’FRIEND)
- B’FRIEND is a public-private partnership in Baltimore City based on innovative use of health data to decrease the rate of falls leading to an emergency room (ER) or hospital admission among elderly.
- Aim (1): Develop and validate a case identification methodology;
- Aim (2): Develop and validate a fall’s risk prediction model
- Aim (3): Evaluate the fall risk score and disseminate results

Prevalence of falls among elderly in Baltimore City (Census Block Group)
CPHIT Portfolio → Predicting Elderly Falls (cont.)

Prevalence of falls among elderly in Maryland (Census Block Group)
Predictors and coefficients of the B’FRIEND model

| Predictors                          | Estimate | Std. error | z value | Pr(>|z|) | Significance | OR  | 2.50%  | 97.50% |
|-------------------------------------|----------|------------|---------|----------|--------------|-----|--------|--------|
| History of fall                     | 1.795    | 0.074      | 24.113  | <2e-16   | ***          | 6.02| 5.20   | 6.97   |
| Fracture                            | 0.604    | 0.104      | 5.821   | 5.85E-09 | ***          | 1.83| 1.49   | 2.24   |
| Substance Abuse                     | 0.520    | 0.082      | 6.364   | 1.96E-10 | ***          | 1.68| 1.43   | 1.97   |
| Parkinson                           | 0.337    | 0.178      | 1.895   | 0.058056 | .            | 1.40| 0.98   | 1.97   |
| Kyphoscoliosis                      | 0.322    | 0.153      | 2.102   | 0.035519 | *            | 1.38| 1.01   | 1.85   |
| Sex (female)                        | 0.173    | 0.046      | 3.736   | 0.000187 | ***          | 1.19| 1.09   | 1.30   |
| Depression                          | 0.146    | 0.068      | 2.141   | 0.032238 | *            | 1.16| 1.01   | 1.32   |
| Mental Illness                      | 0.128    | 0.065      | 1.980   | 0.047652 | *            | 1.14| 1.00   | 1.29   |
| Age                                 | 0.038    | 0.003      | 14.895  | <2e-16   | ***          | 1.04| 1.03   | 1.04   |
| Charlson Index                      | -0.053   | 0.009      | -5.711  | 1.12E-08 | ***          | 0.95| 0.93   | 0.97   |
| Vision                              | -0.211   | 0.057      | -3.689  | 0.000225 | ***          | 0.81| 0.72   | 0.91   |
| Obesity                             | -0.251   | 0.076      | -3.311  | 0.000931 | ***          | 0.78| 0.67   | 0.90   |
| Cardiovascular Disease              | -0.313   | 0.050      | -6.301  | 2.95E-10 | ***          | 0.73| 0.66   | 0.81   |
| Lower Urinary Tract Symptoms        | -0.345   | 0.074      | -4.656  | 3.23E-06 | ***          | 0.71| 0.61   | 0.82   |
| Hypertension                        | -0.357   | 0.050      | -7.080  | 1.44E-12 | ***          | 0.70| 0.63   | 0.77   |
| Cancer                              | -0.441   | 0.081      | -5.418  | 6.02E-08 | ***          | 0.64| 0.55   | 0.75   |
| Lower Back Pain                     | -0.495   | 0.067      | -7.368  | 1.73E-13 | ***          | 0.61| 0.53   | 0.69   |
| Joint Trauma                        | -0.526   | 0.197      | -2.674  | 0.007487 | **           | 0.59| 0.39   | 0.85   |
| Lower Extremity Joint Surgery       | -1.069   | 0.182      | -5.870  | 4.36E-09 | ***          | 0.34| 0.24   | 0.48   |
| (Intercept)                         | -4.372   | 0.197      | -22.249 | <2e-16   | ***          | 0.01| 0.01   | 0.02   |

Significance codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '+'.
CPHIT Portfolio → Predicting Elderly Falls (cont.)

**NEUROLOGICAL DISEASE**
- Dementia
- Parkinson’s
- Tremors
- Weakness
- Stroke
- MS

**CARDIOVASCULAR DISEASE**
- High blood pressure
- Chronic pain
- Liver Disease
- Kidney Disease
- Depression
- Syncope
- Chronic illness

**ENDOCRINE DISEASE**
- Diabetes
- Osteoporosis
- Hypothyroidism
- Vitamin D deficiency
- Lack of estrogen therapy

**MSK DISEASE**
- Joint pain/arthritis
- Gait disorder
- Fractures
- Hyperkyphosis
- History of falling

**URINARY SYMPTOMS**
- Incontinence
- Nocturia
- UTI

**POLYPHARMACY**
- Antidepressants
- Hypnotics
- Muscle relaxants
- Antipsychotics
- Anticonvulsants
- Antihistamines
- Antiemetics
- Narcotics/opioids

**URETERAL STONES**
- Use of assistive devices
- Walking without shoes
- Use of private vehicle
- Walking on uneven surfaces
- Participation in physical activity

**CANCER**
- Depression
- History of falling
- Fear of falling

**VISION LOSS**
- Hearing Loss

**URINARY TRACT INFECTION**
- Use of private vehicle

**B’FRIEND data sources**
CPHIT Portfolio → VHA Obesity

- **Aim (1):** Contextualize obesity factors within VHA’s population health framework
- **Aim (2):** Design a scalable pop-health ‘technical’ platform and develop a pilot for obesity
- **Aim (3):** Derive and evaluate “VHA’s Obesity Trajectory Population-based Risk Prediction Model” to measure the GIS-clustered population-based factors that affect the management of obesity

### Population % in each obesity category

- **Underweight** (BMI<18.5): 1.07%
- **Normal** (BMI>=18.5): 19.12%
- **Overweight** (BMI>=25.0): 35.96%
- **Obese Class I** (BMI>=30.0): 26.21%
- **Obese Class II** (BMI>=35.0): 11.55%
- **Obese Class III** (BMI>=40): 6.09%

**Prevalence of obesity among VHA population**

*Limited to 29,322 visits occurred in one day of 2013; generated using CDW data*
Ongoing Projects → VHA Obesity (cont.)

Geographic distribution of obesity among VHA population
(Limited to 29,322 visits occurred in one day of 2013; generated using CDW data)
Ongoing Projects → VHA Obesity (cont.)

County BMI (using MLM adjustment) for Males 2000-2015
Ongoing Projects → VHA Obesity (cont.)

County BMI (using MLM adjustment) for Males 2000-2015
Ongoing Projects → VHA Obesity (cont.)

County BMI (using MLM adjustment) for Males 2015 (DC and Baltimore)
Ongoing Projects → VHA Obesity (cont.)

Kilometers

Spatial Intensity Male 2015
Ongoing Projects → VHA Obesity (cont.)

Spatial intensity of cases

Spatial intensity of non-cases

Spatial odds for obesity

Bandwidth: 50 km

VISN #19: VA Rocky Mountain Network (Colorado, Montana, Utah, Wyoming)
Ongoing Projects → VHA Obesity (cont.)
Ongoing Projects → VHA Obesity (cont.)

Interactive Web-based Real-time Geo-Temporal Exploration of Obesity Data
(Showing averages of 2014 for MD)
CPHIT Portfolio → Population Health Metrics

- **ACA → Triple aims → ACO/PCMH**

- **CMS → Value-based Purchasing**
  - Bundled models → Salaried models
  - Shared savings (MSSP) and/or Global payments (Pioneers)
  - Blended payments (Capitation + FFS)

- **MD’s DHMH → Maryland’s All-Payer In-Patient Waiver program**
  - Limit the growth of per capita hospital cost @ 3.58% (10 year compound MD’s GDP) → saving Medicare $330mil over the next 5 years
  - Shifting away from fee-for-service models into population-based payment models that reward providers for improving health outcomes → 95% of hospitals are on global budget
  - Immense need for population health management and IT infrastructure to operationalize it
Maryland's All-Payer Waiver Program – Toolkit to align financial incentives
### Global Payment Strategies

<table>
<thead>
<tr>
<th>Method</th>
<th>Services Included</th>
<th>Hospital Participating</th>
<th>Estimated Percent of Revenue at Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Patient Revenue (TPR)</td>
<td>All regulated services</td>
<td>10</td>
<td>~100%</td>
</tr>
<tr>
<td>Admission/Readmission Revenue (ARR)</td>
<td>All-cause readmissions for 30 days</td>
<td>31</td>
<td>~10%</td>
</tr>
<tr>
<td>Population Based Revenue (PBR) and Other Global Models</td>
<td>Core services for specific DRGs in hospital community</td>
<td>TBD</td>
<td>~30% to 100% (estimated)</td>
</tr>
<tr>
<td>Quality Programs with Revenue at Risk (Quality Based reimbursement and Maryland Hospital Acquired Conditions)</td>
<td>All inpatient regulated services <em>State will expand to all regulated services in future years</em></td>
<td>All</td>
<td>For each performance year, Maryland will place the same percentage of hospital revenue at risk as the national Medicare Value-Based Purchasing Program, Hospital Acquired Condition and Readmission Reduction programs a</td>
</tr>
<tr>
<td>Balanced Update Factors</td>
<td>All regulated services</td>
<td>All</td>
<td>N/A</td>
</tr>
<tr>
<td>Volume Controls</td>
<td>All regulated services under the models</td>
<td>All non-TPR/Global revenues</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Maryland’s All-Payer Waiver Program – Global Payment Strategies
On January 10, 2014, the Centers for Medicare and Medicaid Services (CMS) and the State of Maryland jointly announced the launch of a statewide model that will transform Maryland’s health care delivery system. Although some aspects of the new approach may be unique to Maryland and not applicable elsewhere, both the principles of this model and the process that led to its development in Medicare payment per hospital admission. This system has eliminated cost shifting among payers, more equitably spread the costs of uncompensated care and medical.
**CPHIT Portfolio → Population Health Metrics (cont.)**

### Maryland’s Population Health Informatics Backbone
- **Patient Experience Metrics**
- **Population Health Metrics**
- **Healthcare Cost**

### MD All-Payer Population Health Analytics Core
- **State-wide Population Health Data-warehouse**
- **Claims** (HSCRC, CMS)
- **National Data** (HCAHPS, CDC, QBR, PQI)
- **Local Metrics** (MD’s SHIP)
- **New Data Sources?**

### HIE (CRISP)

### Informatics Unit at HSCRC/DHMH
- A
- B
- C

### EHRs

1...n
CPHIT Portfolio → Population Health Metrics (cont.)

Inpatient Utilization By Census Tract – State Level View
CPHIT Portfolio → Population Health Metrics (cont.)

Inpatient Utilization By Census Tract – Neighborhood View (Capitol Heights Area)
CPHIT Portfolio (cont.)

- **National Pop Health IT Curriculum and Workforce Training:**
  Developing a national curriculum for Population Health IT and training more than 7000 healthcare professional incumbents

- https://www.healthit.gov/providers-professionals/health-it-education-opportunities

- http://learnhit.com
Discussion
Discussion → Challenges and Opportunities (cont.)

- **Data sources/types:**
  - How to compare data types and their added value?
  - What are the limits of each data type? What are we missing?
  - What can be used from unstructured data?

- **Data quality:**
  - How much juice is left in this data type (e.g., claims)?
  - Do objective measures have data quality issues (e.g., BMI)?
  - How can we measure the quality of subjective data?

- **Denominator:**
  - Are we excluding noise or signal?
  - Is this a too big of a cut or too narrow – sample size issues?
  - Patient attribution issues...
Discussion → Challenges and Opportunities (cont.)

• **Comparing models:**
  • What are the guidelines?
  • How to compare models across different subpopulations?

• **Feature reduction:**
  • How can we mix multiple approaches such as expert opinion + automated approaches to reduce the feature space?

• **Temporal data:**
  • What window is appropriate?
  • How to deal with large zero fills in temporal data?

• **Privacy and Security:**
  • Is HIPAA helping or hurting HPM’s data scientists?
Discussion → Challenges and Opportunities (cont.)

- **Big Data:**
  - Data volume/quantity (e.g., VHA data query)
  - Data veracity/quality (e.g., VHA BMI data)
  - Data variety/sources (e.g., EHR vs Claims)
  - Data velocity/timeliness (e.g., VHA vital signs)
  - Unstructured data → NLP and HIPAA De-identification issues
Discussion → Data Quality

Actual value: 200.6 lbs.

Measured (same day)
- Validity challenge
  198.9 | 198.9 | 198.9 lbs.
- Reliability challenge
  200.6 | 198.9 | 202.2 lbs.

Measured (diff. days)
- Missing data challenge
  NULL | 200.6 | NULL lbs.

Recorded value: 200 lbs.

User Typed (one entry)
- Typos
  200.6 lbs. → 20.06, 2006
- Mismatching units
  200.6 lbs. → 200.6 kg
- Assumptions/truncations
  200.6 lbs. → 200 lbs.
- Free-text additions
  200.6 lbs. → 200.6 pounds

Data warehouse value: 200 kg

DB Operations (one entry)
- Truncations/Rounding
  200.6 → 200.0
- Error conversions
  200.6 pounds → NULL
- Cleaning
  200+ lbs. → 200.0

Analytics (data points)
- Aggregation of data points
  200 | 0 → mean of 100
- Selecting a representative
  190 | 200 | 210 → 210 (first)
  190 | 200 | 210 → 200 (mean)
  190 | 200 | 210 → 210 (last)
- Removing outliers
  200 | 200 | 350 → 200 | 200 | NULL

Example of data quality issues across various data collection and use stages
Thank you!

Q & A

kharrazi@jhu.edu

www.jhsphs.edu/cphit
Appendix
CPHIT Portfolio → HIE-based Readmission Prediction

Health Information Exchange (HIE)
CPHIT Portfolio → HIE-based Readmission Prediction (cont.)

Federated Consistent HIE
Focus Areas:
- Query Portal Growth
- Direct Secure Messaging
- Encounter Notification System (ENS)
- Encounter Reporting System (ERS)
- Health Benefits Exchange integration

Progress Metric | Result
---|---
Organizations Live | 48
- Hospitals (Total 48)
- Hospital Clinical Data Feeds (Total 143 - Lab, Radiology, Clinical Docs)
- National Labs
- Radiology Centers (Non-Hospital)

Identities and Queries |  
---|---
- Master Patient Index (MPI) Identities ~4M
- Opt-Outs ~1500
- Queries (Past 30 Days) ~3500

Data Feeds Available |  
---|---
- Lab Results ~16M
- Radiology Reports ~5M
- Clinical Documents ~2M
CRISP Encounter Notification System + Readmission Risk Prediction Model (RRPM) integration

Copyright CRISP
CPHIT Portfolio → HIE-based Readmission Prediction (cont.)

Notice: Risk of readmission for patient is high.

Simulate a different outcome...

Drag sliders to see the estimated effect on readmission risk score.

Advance RRPM integration in CRISP Encounter Notification System