Using EHR Data to Conduct PBRN Research

Laura Mae Baldwin MD MPH – University of Washington, lmb@uw.edu
Alex Fiks MD – Children’s Hospital of Philadelphia, fiks@email.chop.edu
Alex Krist MD MPH – Virginia Commonwealth University, akrist@vcu.edu

Goals

• Understand how to get started with using EHR data and HIT across the PBRN research lifecycle.
• Be aware of the possible uses for EHR data and HIT in conducting PBRN research.
• Identify the facilitators and barriers to incorporating EHR data and HIT in your research projects.
• Develop strategies to ensure that PBRN partners benefit from EHR and HIT-based projects.

Introduction
How EHR Data and HIT Support PBRN Research Across the Translational Science Spectrum

<table>
<thead>
<tr>
<th>Research Lifecycle Stage</th>
<th>Examples of Support Provided by EHR Data and HIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Collaborations</td>
<td>Tailor HIT and EHR data queries to answer questions important to practices.</td>
</tr>
<tr>
<td>Identifying/Choosing the Problem</td>
<td>EHR data-based discovery.</td>
</tr>
<tr>
<td>Stating the Research Question</td>
<td></td>
</tr>
<tr>
<td>Developing a research approach</td>
<td>HIT-based intervention. EHR data may provide the research data – define outcomes, delineate mediator and moderator variables.</td>
</tr>
<tr>
<td>Select a sample</td>
<td>EHR can facilitate cohort discovery.</td>
</tr>
<tr>
<td>Collect high quality data</td>
<td>EHR can provide discrete data fields, standardized coding (e.g., ICD, CPT, SNOMED-CT).</td>
</tr>
<tr>
<td>Analyze and interpret data</td>
<td></td>
</tr>
<tr>
<td>Disseminate findings</td>
<td>EHR-based patient portals can communicate with patients.</td>
</tr>
</tbody>
</table>

Research New Functions and New Technology

Creating new technology or modifying existing systems is hard...

Think about partners for collaboration; create a multidisciplinary team; align with your EHR vendor; find developers who have done something similar.
Don't Forget Dissemination & Implementation Research!

Improving how we use EHRs and technology may be more important than what we use.

Frameworks exist for evaluating Dissemination & Implementation outcomes (e.g., RE-AIM.org).

Roles of the EHR in Research

Pilot Data – Obesity Study

- Demographics:
  - Eligible – 3,734
  - Female – 49.5%
  - Age – 9.7 years (6.6 – 12.8)
  - Time between visits – 395 days (322 to 552)

Longitudinal EHR Data Facilitates Further Analysis:

If "obese" at baseline, then at second visit:

- 91%
- 7%
- 2%

Remained same
Returned to healthy weight
Overweight
Obese

If "overweight" at baseline, then at second visit:

- 90%
- 9%
- 1%

Remained same
Returned to healthy weight
Overweight
Obese
Subject Recruitment
Patient lists delivered to research assistants

- Lists of appointments for potential subjects that favor sensitivity over specificity
- Sent each Friday to research team
- Study team can target specific dates/times to be on-site to maximize their recruitment potential

Subject Recruitment:
Extracting Data from Templates

Enrollment: EHR-Based Forms
Using the EHR as the Intervention: Decision Support

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Due Date</th>
<th>Next Doses</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV</td>
<td>6/1/2010</td>
<td>9/1/2010</td>
</tr>
<tr>
<td>Measles</td>
<td>6/1/2010</td>
<td></td>
</tr>
<tr>
<td>Mumps</td>
<td>6/1/2010</td>
<td></td>
</tr>
<tr>
<td>Varicella</td>
<td>6/1/2010</td>
<td></td>
</tr>
<tr>
<td>Influenza</td>
<td>10/1/2016</td>
<td></td>
</tr>
</tbody>
</table>


Feedback Report

- Performance feedback—clinicians made aware of their own vaccination rates and how they compare to their practice and care network.
- Generated from EHR data

<table>
<thead>
<tr>
<th>Visit Type</th>
<th>Human Papilloma Virus (HPV) Vaccine Given to Eligible Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>You</td>
</tr>
<tr>
<td>Work Visit</td>
<td>36</td>
</tr>
<tr>
<td>Sick Visit</td>
<td>38</td>
</tr>
</tbody>
</table>

*Note: this is actual data from a randomly selected study clinician

Example #1 – National EHR datasets, getting to “Big Data”
What is Big Data Anyway?

“...a broad term for data sets so large or complex that traditional data processing applications are inadequate.”

Data Provenance

The first law of informatics:

“Data shall only be used for the purpose for which they were collected.”


The law of medical information: Berg and Goorman

• “The further information has to be able to circulate (i.e. the more diverse contexts it has to be usable in), the more work is required to disentangle the information from the context of its production. The question that then becomes pertinent is; who has to do this work, and who reaps the benefits?”

*Int J Med Informatics* 1999; 56:51–60
Comparative Effectiveness Research through Collaborative Electronic Reporting (CER2): Opportunities and Cautions as Data Gets Big

What is CER2 & what is its origin?

- CER2 is a collaboration of primary care informatics researchers studying pediatric care through EHR and related electronic data
- CER2 grew out of a need for large scale 21st century practice-based research
- CER2 joins existing EHR-based research networks into an electronic über-network

CER2 Unique Appeal

- Longitudinal electronic health record (EHR) database
  - Extensive follow-up time 2000-2014
  - Clinical data supplemented by administrative data
- Contains >1.2 million children, diverse practitioners, seen by practitioners in diverse settings from across the United States
- Supported by a unique team combining, health services, informatics and pharmacoepidemiology expertise
- EHR data from CER2 can be combined with data collected from pediatricians, parents, and children to conduct prospective interventional studies and provide a powerful 21st century research platform
CER² Participating States

222 Practice Sites in 27 States with 2119 Practitioners
Representing 1.2 Million Covered Lives


CER² Model

ePROS and
ENQUIRENet

DARTNet

CER²

Boston University

CER² Governance

• American Academy of Pediatrics “owns” the data and has data use agreement with all data contributors

• Aggregated HIPAA-limited dataset is stored on secure server at The Children’s Hospital of Philadelphia

• CER² partner members contribute and analyze data

• CER² affiliate members can help analyze data

• Non-member investigators can help analyze data

• The CER² team has begun to recruit additional partners, affiliates, and non-member investigators to write grants, contribute data and answer new questions.

Aggregated data in standardized (OMOP) format
Implausible Growth Data: Data Cleaning Method

- Challenge: CER² has growth data on ~1.2 million children
- Errors include:
  - Substitution of metric and English system values
  - Misplaced decimal points
  - Plain old mistakes
- How to sort these out?
- The “Daymont Method”
Missing Data on Race/Ethnicity

Missing data on race/ethnicity is common
• Could an adaptation of a method used previously in adults be helpful in pediatrics?
• We tested "Bayesian Improved Surname Geocoding" which accounts for US census geospatial and surname data in addressing missingness
• Bottom line: The new method more correctly race/ethnicity than traditional imputation, and reduced bias

Validity of Imputed Race/Ethnicity for Health Disparities Research

Table: Bias of different methods of estimate race/ethnic difference: true values and estimates by method

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>True value</th>
<th>Complete Case</th>
<th>Indicator scaled</th>
<th>Multiple imputation</th>
<th>Peto observed differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>1.0</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.0</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Evaluating the Impact of New Psychotropic Medication Guidelines

- In 2011, AAP ADHD practice guidelines provided, for the first time, guidance for the diagnosis and treatment of ADHD in preschoolers.

- To evaluate the impact on clinical practice, we examined changes in the diagnosis of ADHD and prescription of stimulants to children aged 4 through 5 years old following guideline publication.

Patterns were similar in sensitivity analyses that included all practices.

Recap: Big Data

- Data is being repurposed—this has inherent challenges
- Data cleaning is critical
- Once cleaned, pooled data may facilitate the elucidation of national practice patterns and the assessment of rare outcomes
- The CER2 team welcomes interested investigators to participate.
Example #2 – From regional data sharing infrastructure to small scale EHR data projects

WWAMI region Practice and Research Network

Data QUEST

WHO: A collaboration between the WPAN and CISA biomedical informatics program

WHAT: Technology – aligns data repositories across practices

• Identified clinical data repositories → at the practices
• De-identified data → in a central warehouse hosted at UW

WHY: Supporting research and quality improvement

BEYOND OUR REGION: Aligning with national efforts

• Practice-Based Research Networks → DARTNet Institute
• Patient-Centered Outcomes Research Institute (PCORI) PCORNet
• NIH Collaboratory Distributed Research Network
Data QUEST

- Strengths
  - Remarkable tool that has upgraded PRNN collaborations
  - 13 grants from small-scale studies to large-scale pragmatic clinical trials
  - 17 manuscripts on the subject of data sharing

- Challenges
  - Costs are significant for
    - Coordinating center personnel to manage and conduct operations, manage and maintain data quality
    - Implementation costs ~$7,000-10,000 per site, maintenance ~$2,500 annually
    - Data extraction costs ~$3500 for the first site, $750 each additional
  - Navigating governance structures in CCR/CCMGs can be formidable
  - CCMG organizations may change their structure and/or make updates that could disrupt the infrastructure
  - Verizon in EHR structure and how EHRs are used is substantial
  - Not all WPRN sites can participate in Data QUEST

Leveling the Playing Field:
Using a Small Scale EHR Data Project to Increase Quality and Research Capacity in the WPRN

- Topic Selection
- Site Champions
- Pilot-scale data
- Coordinating Center
- Coordination of study
- Site Champions
- Selected EHR topic

- Parameter Development for EHR Data Extraction
- The Coordinating Center
- Guiding parameters for selected EHR topics
- Collected from site review and stakeholder feedback
- The Coordinating Center
- Standardized the parameters

- Participant Training
- The Coordinating Center
- Trained the site staff on the EHR topics
- The Coordinating Center
- Provided feedback during training

- Data Collection
- Sites pulled data
- Provided the Coordinating Center

- Results Dissemination
- Coordinating Center
- Disseminated the results
- Site puller sites and all site members
Returning Results to Sites is Critical

- One page study summary sheet
- Presented anonymously to the WPRN site champions at our annual meeting.
- Sites knew who they were – could compare their rates to similar practices

Small Scale EHR Data Project

- **Strengths**
  - More WPRN sites could participate
  - Sites were highly engaged in the science, data collection, and results interpretation
  - No IRB approval required!

- **Challenges**
  - No process for assessing data quality or validation
  - Results can be used for preliminary studies, but no individual data for research
  - Limited resources to support follow-on quality improvement efforts
Lessons Learned

- EHR-based projects are a powerful tool for engaging PBRN sites in research
  - Small scale EHR projects
    - Highly collaborative, responsive to PBRN site interests
    - Demonstrated site capabilities for EHR queries to sites themselves and to the PBRN Coordinating Center
  - Large scale data sharing
    - Requires a greater commitment from sites – must provide information on data origins (provenance)
    - Engages them as collaborators with academic investigators on pragmatic clinical trials, implementation and dissemination research

Example #3 – Integrating patient reported data into the EHR

Informed Decision-Making Module

- Focus on three cancer screening decisions
- Specific aims:
  - **Aim 1** – to reach patients outside the clinical setting – before clinical encounters – to explore their preferred approach to decisions about cancer screening
  - **Aim 2** – to follow patients into the clinical encounter to study the assistance offered by clinicians, its congruence with the patient’s stated preferences, and the effect of website exposure on the conversation and decision outcomes
How Information Technology Could Help

Patients Facing Cancer Screening Decisions (n=72,000 portal users)

<table>
<thead>
<tr>
<th></th>
<th>Any</th>
<th>Breast</th>
<th>Colon</th>
<th>Prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>11,094 (13.5%)</td>
<td>3,615 (4.4%)</td>
<td>6,115 (7.5%)</td>
<td>1,364 (1.7%)</td>
</tr>
<tr>
<td>Phase 1 (6 weeks)</td>
<td>1,010</td>
<td>297</td>
<td>542</td>
<td>171</td>
</tr>
<tr>
<td>- Prompt when using</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 2 (14 weeks)</td>
<td>610</td>
<td>171</td>
<td>354</td>
<td>85</td>
</tr>
<tr>
<td>- Invite before visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 3 (12 weeks)</td>
<td>9,436</td>
<td>3,220</td>
<td>5,136</td>
<td>1,080</td>
</tr>
<tr>
<td>- invite outside of visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Module Use by Cancer Screening and Phase

P ≤ 0.001 (start and complete)
- Breast vs. colon
- Breast vs. prostate
- Colon vs. prostate

P ≤ 0.001 (start and complete)
- Phase 1 vs. phase 2
- Phase 1 vs. phase 3
- Phase 2 vs. phase 3
Impact of the Module on the Visit

Getting HIT to Work for this Research Study

Facilitators
- We controlled the design and programming of the patient portal
- Patient and clinician engagement supported the research
- HIT automated a repetitious process that occurred frequently in practice

Barriers
- We had limited control over how the module interacted with the EHR
- More time is needed to cause the culture change needed to redesign care
- Sometimes it is difficult to balance research versus clinical data collection

The Benefits of Patient Reported Data
- Can be used to better engage patients in their care and wellbeing
- Can extend care outside of traditional office visits
- Can serve as a new data source – the answers to which only patients know
The IOM Has Prioritized Capturing Patient Reported Data in Electronic Health Records

Thank you for your interest

Key References

- Electronic health record functionality needed to better support primary care. *J Am Med Inform Assoc.* 2014; 0:1-8. [PMID 24421225].
- [https://dataquest.iths.org](https://dataquest.iths.org)