The CER Hub: An Informatics Platform for Collaborative Comparative Effectiveness Research using Comprehensive Electronic Medical Record Data

Brian Hazlehurst, PhD
Kaiser Permanente Northwest
Center for Health Research
CER requires LOTS of data

- Diverse populations, many topic areas
- Increasing adoption of EMR systems provides an emerging opportunity for developing large databases
- KP covers ~9M lives @ 4 encounters/yr, roughly 100,000 encounters per day captured in the EMR
  - A vast amount of this data is captured in unstructured (non-coded) text
Example clinical encounter records addressing family hx for cancer

Clinical note segment written

<table>
<thead>
<tr>
<th>Medical History: Asthma-Azmacort, Ventolin, rarely prednisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical History: neg</td>
</tr>
<tr>
<td>Family History: Fa-aodm, pgf colon ca, mgm bone marrow ca</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevant ICD9 dx code applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last mammogram: 1 yr ago. Previous Paps have been normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a strong family hx of breast cancer. (M, MGM, Aun)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevant ICD9 dx code applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROS: neg for exertional chest pain or pressure, shortness of breath, changes in bowel habits. Family History: + early MI, colon cancer-- sister in her 50s</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Relevant ICD9 dx code applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>V16.0 FAMILY HX MALIGNANCY GI TRACT</td>
</tr>
</tbody>
</table>
CER requires LOTS of data

- EMR adoption promises LOTS of data, but the data are heterogeneous both across and within institutions
  - EMR’s are variable (diverse representation of events)
  - Clinical practices are variable (diverse demands driving capture of events)
  - Patients are variable (diverse conditions and needs)

- Need scalable informatics solutions allowing assignment of consistent (and specific) meanings to highly heterogeneous data
The CER HUB

A web-based platform for collaborative development of study-specific, standardized, processors of comprehensive electronic medical record data.

1. Study team formed around specific questions and partners with participating “data providers”.
2. Data are extracted locally by each data provider (on demand for specific projects) in industry standard form (HL7 CDA).
3. A study-specific data processor of medical records (text and coded data) is created and validated using a set of online tools.
4. The data processor is downloaded and applied locally to generate standardized, study-specific, datasets.
5. These data are then pooled to answer targeted study questions centrally.
The CER HUB

- A web site with functions related to building, testing, sharing, study-specific processors of heterogeneous clinical data.
- A platform that hosts studies which use these processors to answer questions (asthma control, smoking cessation)
Building the data processor

- A set of tools for collaborative development of study-specific data processors.
  - Operationalizing study variables in terms of events identified in clinical records
  - These variables may involve events identified in text and/or structured data elements of clinical records.
  - Eg., “persistent asthma” can be operationalized in terms of sequences of asthma medication fills, exacerbation visits, and clinician assessment in the progress note.
Testing the data processor

- Developed processors can be run on the HUB against (de-identified) test datasets to evaluate performance.
- Allows for rapid development of the processor through iterative test-refine cycles.
- Validation study creates metrics that provide a “profile” about the data processor.
Sharing the data processor

- A web site hosting virtual communities of researchers with shared interests.
- Following conclusion of a study, the developed study-specific data processor becomes available for download to all CER HUB members.
- Researchers who join the HUB build out a “library” of processors through their activities using the HUB for conducting their research.
The CER HUB workflow

1) Develop a study protocol
   (define study measures and populations of interest)

2) Develop and validate a standardized data processor
   (operationalize study measures based on concepts in data)

3) Configure the processor for your site
   (define site-specific parameters for the processor)

4) Apply processor to local data

5) Pool standardized, sharable data for analysis

1a) Data extraction
   (extract data in a standard format)

Pop def
De-identified samples
Download processor
Study Protocol

Population and data element selection for encounter-based extraction

MediClass "knowledge module" provides configuration for specific application

Application specific extraction filter for study specific (and sharable) events

Study variables operationalized in terms of temporally located events

**CER HUB**

**EMRAdapter**

**MediClass Application**

**Events Extractor**

**EventsDataset Processor**

**Data Warehouse (XML)**

**Data Extraction**

**Event Identification**

**Study Analyses**

**Local Site**

**Data Coord. Ctr**

**CDA (XML)**

**CDA w/ MediClass Classifications (XML)**

**EventsDataset (Flat file)**

**Study Measures**
Informatics Tools for Evaluating Health and Healthcare

CER Hub is a web-based informatics platform for conducting healthcare research. Research projects using CER Hub technologies are formed as investigator-led communities focused on Comparative Effectiveness Research.

Learn More » View CER Projects »

Online collaboration for CER studies
The CER Hub is a web-based mechanism for conducting Comparative Effectiveness Research (CER) where researchers can collaboratively develop protocols to define and operationalize healthcare research questions and methods to answer these from electronic data.

Building clinical data infrastructure
Researchers can develop multi-institutional data sets using CER Hub’s centralized web-based services. These services provide automated tools and support for generating standardized data sets and allow analyses to answer CER questions.

Extracting EMR data
Standardized data processors built on the CER Hub make available natural language processing and knowledge-based systems technologies to automatically identify clinical events in all types of clinical data. Because the CER Hub uses an emerging standard for representing the complete medical record, data from any EMR implementation can be uniformly processed.
The CER HUB Project

- A consortium of researchers from 6 health systems
  - KPNW, KPGA, KPHI,
  - VA PugetSound,
  - Baylor HealthCare System,
  - OCHIN (consortium of FQHCs mostly on west coast)

- Developing and using the CER HUB to address effectiveness questions in asthma control therapy and smoking cessation services
# CER Hub “Possible Asthma” Population

<table>
<thead>
<tr>
<th>Study Sites</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>TOTAL DISTINCT PATIENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baylor</td>
<td>3166</td>
<td>6138</td>
<td>9836</td>
<td>6504</td>
<td>4850</td>
<td>30494</td>
</tr>
<tr>
<td>KP Georgia</td>
<td>9858</td>
<td>10224</td>
<td>10357</td>
<td>10917</td>
<td>5266</td>
<td>26756</td>
</tr>
<tr>
<td>KP Hawaii</td>
<td>12637</td>
<td>12182</td>
<td>12227</td>
<td>12756</td>
<td>12324</td>
<td>33349</td>
</tr>
<tr>
<td>KP Northwest</td>
<td>21342</td>
<td>22495</td>
<td>23677</td>
<td>24741</td>
<td>24731</td>
<td>64764</td>
</tr>
<tr>
<td>OCHIN</td>
<td>1997</td>
<td>3870</td>
<td>6204</td>
<td>11260</td>
<td>15306</td>
<td>26922</td>
</tr>
<tr>
<td>VA Puget Sound</td>
<td>1377</td>
<td>1668</td>
<td>1880</td>
<td>1972</td>
<td>2235</td>
<td>4667</td>
</tr>
<tr>
<td><strong>TOTAL DISTINCT PATIENTS</strong></td>
<td><strong>50377</strong></td>
<td><strong>56577</strong></td>
<td><strong>64181</strong></td>
<td><strong>68150</strong></td>
<td><strong>64712</strong></td>
<td><strong>186952</strong></td>
</tr>
</tbody>
</table>

© 2012, KAISER PERMANENTE CENTER FOR HEALTH RESEARCH
The 5 A’s of Smoking Cessation

<table>
<thead>
<tr>
<th>5A step</th>
<th>Operational definition</th>
<th>Example in free-text section of EMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask</td>
<td>Identify tobacco user status at every visit</td>
<td>“Patient smokes 1 ppd”</td>
</tr>
<tr>
<td>Advise</td>
<td>Advise all tobacco users to quit</td>
<td>“It is important for you to quit smoking now”</td>
</tr>
<tr>
<td>Assess</td>
<td>Determine patient’s willingness to make a quit attempt</td>
<td>“Patient not interested in quitting smoking”</td>
</tr>
<tr>
<td>Assist</td>
<td>Aid the patient in quitting</td>
<td>“Started patient on Zyban”</td>
</tr>
<tr>
<td>Arrange</td>
<td>Schedule follow-up contact, in person or via telephone</td>
<td>“Follow-up in 2 weeks for quit progress”</td>
</tr>
<tr>
<td>Study Sites</td>
<td>2006</td>
<td>2007</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Baylor</td>
<td>2707</td>
<td>12338</td>
</tr>
<tr>
<td>KP Georgia</td>
<td>17385</td>
<td>16067</td>
</tr>
<tr>
<td>KP Hawaii</td>
<td>19160</td>
<td>17849</td>
</tr>
<tr>
<td>KP Northwest</td>
<td>47202</td>
<td>47786</td>
</tr>
<tr>
<td>OCHIN</td>
<td>8489</td>
<td>14726</td>
</tr>
<tr>
<td>VA Puget Sound</td>
<td>10944</td>
<td>12334</td>
</tr>
<tr>
<td><strong>TOTAL DISTINCT PATIENTS</strong></td>
<td><strong>105887</strong></td>
<td><strong>121100</strong></td>
</tr>
</tbody>
</table>
Informatics Tools for Evaluating Health and Healthcare

CER Hub is a web-based informatics platform for conducting healthcare research. Research projects using CER Hub technologies are formed as investigator-led communities focused on Comparative Effectiveness Research.

Learn More »  View CER Projects »

Online collaboration for CER studies
The CER Hub is a web-based mechanism for conducting Comparative Effectiveness Research (CER) where researchers can collaboratively develop protocols to define and operationalize healthcare research questions and methods to answer these from electronic data.

Building clinical data infrastructure
Researchers can develop multi-institutional data sets using CER Hub's centralized web-based services. These services provide automated tools and support for generating standardized data sets and allow analyses to answer CER questions.

Extracting EMR data
Standardized data processors built on the CER Hub make available natural language processing and knowledge-based systems technologies to automatically identify clinical events in all types of clinical data. Because the CER Hub uses an emerging standard for representing the complete medical record, data from any EMR implementation can be uniformly processed.