Getting Your Questions Answered With the VDW

2007 HMORN Meeting
Ed Wagner
Gene Hart
Roy Pardee
Tyler Ross
The VDW

- A Virtual Data Warehouse
- Purpose
  - Increase efficiency – allow program written at one HMO to be run at other HMOs with little modification
  - Encourage consistency – document issues once
  - Faster turnaround time on requests
What the VDW is NOT

- NOT a Data Coordinating Center
- NOT a Centralized Data Warehouse
  - Raw data remains at each HMO
- NOT a Standard that will be enforced on all research projects
VDW Data Areas

- Enrollment
- Demographics
- Outpatient Pharmacy
- Encounters
  - Procedures and Diagnoses
- Tumor
- Vital Signs
- Census
VDW Standardized Methods

- Because the data structures are standardized, we can write important bits of code once and re-use them as needed.

- For example:
  - RxRisk chronic disease score
  - Continuous enrollment
Enrollment

Macros for:

- 2 Different types of continuous enrollment definitions (disregarding gaps of specified # of months)
- Finding date of first disenrollment after an index date

MRN
- enr_start
- enr_end
- ins_medicare
- ins_medicaid
- ins_commercial
- ins_privatepay
- ins_other
- drugcov
Demographics

- Follows SEER coding for race
- Standardized macro for calculating age

<table>
<thead>
<tr>
<th>MRN</th>
<th>birth_date</th>
<th>gender</th>
<th>race1-5</th>
<th>hispanic</th>
</tr>
</thead>
</table>

hmo research network
Tumor

- Generally a subset of NAACCR standard variables.
- Standard Macro for drawing samples of women with an Invasive Breast Cancer between specified dates.
Outpatient Pharmacy

Macros for:
- Pulling all fills for a given sample of people.
- Pulling all fills for a given list of National Drug Codes
- Producing counts of fills for a given list of NDCs

<table>
<thead>
<tr>
<th>Pharmacy</th>
<th>MRN</th>
<th>ndc</th>
<th>rxdate</th>
<th>rxsup</th>
<th>rxamt</th>
<th>rxmd</th>
</tr>
</thead>
</table>
## Vital Signs

- No need to impute weights for the BMI calculations—it’s already done!
- Standard macro for pulling all vital sign measures for a given sample of people.

<table>
<thead>
<tr>
<th>Vital Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRN</td>
</tr>
<tr>
<td>measure_date</td>
</tr>
<tr>
<td>ht</td>
</tr>
<tr>
<td>wt</td>
</tr>
<tr>
<td>bmi</td>
</tr>
<tr>
<td>days_diff</td>
</tr>
<tr>
<td>diastolic</td>
</tr>
<tr>
<td>systolic</td>
</tr>
<tr>
<td>position</td>
</tr>
</tbody>
</table>
Census

- Census Bureau reported demographics tied to individuals via geocoded addresses.
- Standard macro for pulling census for a given sample of people.
Encounters

- In- and Out-patient encounters
- Tons of macros for pulling people, procedures & diagnoses.
Sites Participating in VDW

Cancer Research Network Sites

- Group Health Cooperative Center for Health Studies
- Kaiser Permanente Northwest, The Center for Health Research/Northwest
- Kaiser Permanente Northern California Division of Research
- Kaiser Permanente Southern California Center for Research and Evaluation
- Kaiser Permanente Hawaii The Center for Health Research/Hawaii
- HP, HealthPartners Research Foundation
- Marshfield Clinical Research Foundation
- Kaiser Permanente Colorado, Clinical Research Unit
- Henry Ford Health System
- Fallon Community Health Plan, Meyers Primary Care Institute
- Geisinger Health System Center for Health Research and Rural Advocacy
- Kaiser Permanente Georgia, The Center for Health Research/Southeast

hmo research network
VDW Availability by HMO

- See the handout
General Principles of VDW Use

- Access is open to all HMORN projects
- Projects must pay their own way to use
- Transfer the minimum data necessary
VDW and the Future

- We have begun to standardize the data. We have more work to do
  - Improve data quality across HMOs
  - Add new data areas (Lab, Pathology, Radiology…)
  - Design better methods of communicating what the VDW can do
VDW and Multi-site Collaborations

- We have started to standardize the data
- The process of multi-site collaboration is not standardized
Process for Using the VDW

- Preparatory to Research Idea
- Using VDW for a funded grant
Using VDW: Approval Steps for all projects

- Find collaborating investigator at each interested site
- At each HMO the local site investigator need to get, or confirm IRB approval (mention Wed IRB talk)
Using VDW: Steps unique to funded research projects

- Explore necessity of a DUA with each site providing data
  - Consider who will be 1st authors of papers.
  - Where will analysis be done?
  - Will the final analytic dataset be shared at all sites?
Using VDW: Acquiring Data

- Project team develops specifications
- VDW program written at lead site
- Test VDW program at another site
- Distribute program to each participating site
- Program run at each site and results returned to lead site
- Analyse and publish
Real examples of VDW Projects

- A Prep to Research Count – Colorectal cancer counts
- Simple summarized data project – A.I.
- A project where individual level data are transferred CanCors Special Project
- A hybrid project (both VDW & not) BOW2
## Prep to Research Colorectal Cancer Counts

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>6/26</td>
<td>Tyler learns about the project</td>
</tr>
<tr>
<td>6/27</td>
<td>Tyler learns about modification to request. Tyler writes program and sends to sites</td>
</tr>
<tr>
<td>6/28</td>
<td>Karen (@HFHS) returns results. Investigator sees results and modifies specifications</td>
</tr>
<tr>
<td>6/29</td>
<td>Tyler learns of spec changes, modifies program, resends to all sites. Karen returns new results.</td>
</tr>
<tr>
<td>6/30</td>
<td>Jenn (@KPCO) returns results. Tyler assembles 3 sites results and send to investigator.</td>
</tr>
</tbody>
</table>
## Summarized data project - A.I.

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 2004</td>
<td>Formal study group forms</td>
</tr>
<tr>
<td>Mar 2004 – Jun 2005</td>
<td>Develop oncologist survey; sketches for identifying recurrence w/out abstraction, etc.</td>
</tr>
<tr>
<td>July 2005</td>
<td>Final table shells/NDC list developed.</td>
</tr>
</tbody>
</table>
| Sep 2005      | • Programming begins.  
|               | • Single-site output produced.                                              |
|               | • Candidate program sent to Site 2 for testing (2-day turnaround on results). |
### Summarized data project- A.I.

| Oct 2005          | • Combined output produced—provokes requests for tweaks & additional data.  
                   | • Candidate program revised & re-sent to Site 2 (1 day turnaround). |
|-------------------|-------------------------------------------------------------------------|
| Nov 2005          | • Biostat consult provokes further revision requests.                   |
| Dec 2005          | • Revised program sent to Site 2 for final test (same-day turnaround).  
                   | • Formal request made to 6 additional sites.                            
                   | • Usable data returned by Site 3, Site 4, and Site 5 (during holidays!).|
### Summarized data project- A.I.

<table>
<thead>
<tr>
<th>Date</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2006</td>
<td>• Data Issue: no ER marker data at Site 6.</td>
</tr>
<tr>
<td></td>
<td>• Data Issue: no DxDate at Site 7.</td>
</tr>
<tr>
<td></td>
<td>• Programming issue: Misleading program comment resulted in 2 sites missing fills from 2004—need to re-run at sites 4 &amp; 5</td>
</tr>
<tr>
<td></td>
<td>• Usable data returned by Site 8.</td>
</tr>
<tr>
<td>Feb 2006</td>
<td>• Re-runs completed at Sites 4 &amp; 5.</td>
</tr>
<tr>
<td>Mar 2006</td>
<td>• Site 6 cures ER marker issue and returns usable data.</td>
</tr>
<tr>
<td></td>
<td>• Final combined output distributed to group w/results from sites 1, 2, 3, 4, 5, 6 &amp; 8</td>
</tr>
</tbody>
</table>
Summarized data project- A.I.

- **Scientific work products:**
  - Am. Soc. Of Preventive Oncology Poster
  - Manuscripts
    - JCO reviewers required reanalysis of just adjuvant use—not possible w/the summarized data produced.
    - BCR&T accepted

- **Lesson Learned by PI**
  - May be worth jumping through the extra IRB hoops to get individual-level data.
Raw data transfer project – CanCors Special Project

- Biggest issue was that standardized data structures does not assure standardized data across HMOs
  - Coding not uniform across sites.
    - 2 HMOs had homegrown PROC codes
  - Currency of data varies across sites
  - IRB scope varied by HMO
Hybrid Project – BOW2

- Required VDW data and non-VDW data.
- Non-VDW data:
  - Lab (hg A1C + LDL Cholesterol)
  - Flu immunizations
  - Distinguishing screening from diagnostic mammograms

- All these required custom site programming.
## Hybrid Project – BOW2

<table>
<thead>
<tr>
<th>Date</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/16</td>
<td>Roy learns of the project—shell tables provided.</td>
</tr>
<tr>
<td>8/28</td>
<td>Single-site output generated for feedback</td>
</tr>
<tr>
<td>8/29</td>
<td>Program sent to Site 2 for testing</td>
</tr>
<tr>
<td>9/6</td>
<td>Data sent to Site 4 for testing</td>
</tr>
<tr>
<td>9/7</td>
<td>Data back from Site 2; Program sent to Sites 3 and 5</td>
</tr>
</tbody>
</table>
## Hybrid Project – BOW2

<table>
<thead>
<tr>
<th>Site</th>
<th>Date</th>
<th>Issue Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>8/29</td>
<td>No problems</td>
</tr>
<tr>
<td>Site 2</td>
<td>9/7</td>
<td>Testing site</td>
</tr>
<tr>
<td>Site 3</td>
<td>10/24</td>
<td>IRB delay (also: required that only site-aggregated results would be disseminated)</td>
</tr>
<tr>
<td>Site 4</td>
<td>11/20</td>
<td>Data availability delay (no tumor registry)</td>
</tr>
<tr>
<td>Site 5</td>
<td>12/19</td>
<td>Data availability delay</td>
</tr>
</tbody>
</table>

*Source: hmo research network*
Hybrid Project – BOW2

- Take-home:
  - Allow extra time for requests with non-VDW components.
  - Be sure and communicate how & to whom you intend to disseminate your results. The Site 3 requirement that PI not see their site’s data unaggregated came as a surprise.
Barriers to VDW use

- Determining which sites had available data
- Determining range of time data were available by site
- Determining what data were available by site
- Inconsistent coding across sites
- Documentation poor
- Blanket IRB approval not a reality at all sites
Facilitators of VDW Use

- VDW data documentation
- Good network of contacts at the sites
- A programmer who understands the VDW
- Tailor questions to data I knew were available
What do you know now that you wish you knew when you were 1st using the VDW?

- Tailor the idea to the limited nature of the data
- Better to get IRB approval for individual level data rather than summarized data
- There has not been much data validation across sites
Where to go for Information

❖ Site Data Manager at your site
❖ Online resources:
  ✦ Programmer’s Wiki
  ✦ Cancer counter
  ✦ Other counters
    • Diagnosis
    • Procedure
    • Pharmacy (coming soon?)
  ✦ CCSN handout?
Characteristics of Successful Collaborations – Sarah Greene’s Poster at this meeting

- Communication, communication, communication
- Leadership that is attentive, respectful and solicits input appropriately
- Shared opportunities
  - 1st authored papers
  - new grants
- Publication discussions early & often
- Transparent decision-making
- Identified mentors for junior scientists
- Data use agreements (DUAs) and related policies anticipate collaborative approaches to analysis & publication