

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

Enrollment

MRN

enr_start
enr_end
ins_medicare
ins_medicaid
ins_commercial
ins_privatepay
ins_other
drugcov

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

Demographics

Demographics

MRN

birth_date

gender

race1-5

hispanic

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

Tumor

Tumor

MRN

dxdate
staging vars...
tumor vars...
treatment vars...
etc.

- ❖ Generally a subset of NAACCR standard variables.
- ❖ Standard Macro for drawing samples of women w/an Invasive Breast Cancer between specified dates.

Outpatient Pharmacy

Pharmacy

MRN

ndc

rxdate

rxsup

rxamt

rxmd

❖ 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

Vital Signs

Vital Signs

MRN

measure_date

ht

wt

bmi

days_diff

diastolic

systolic

position

- ❖ 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.

Census

Census

MRN

block

blockgp

county

state

tract

zip

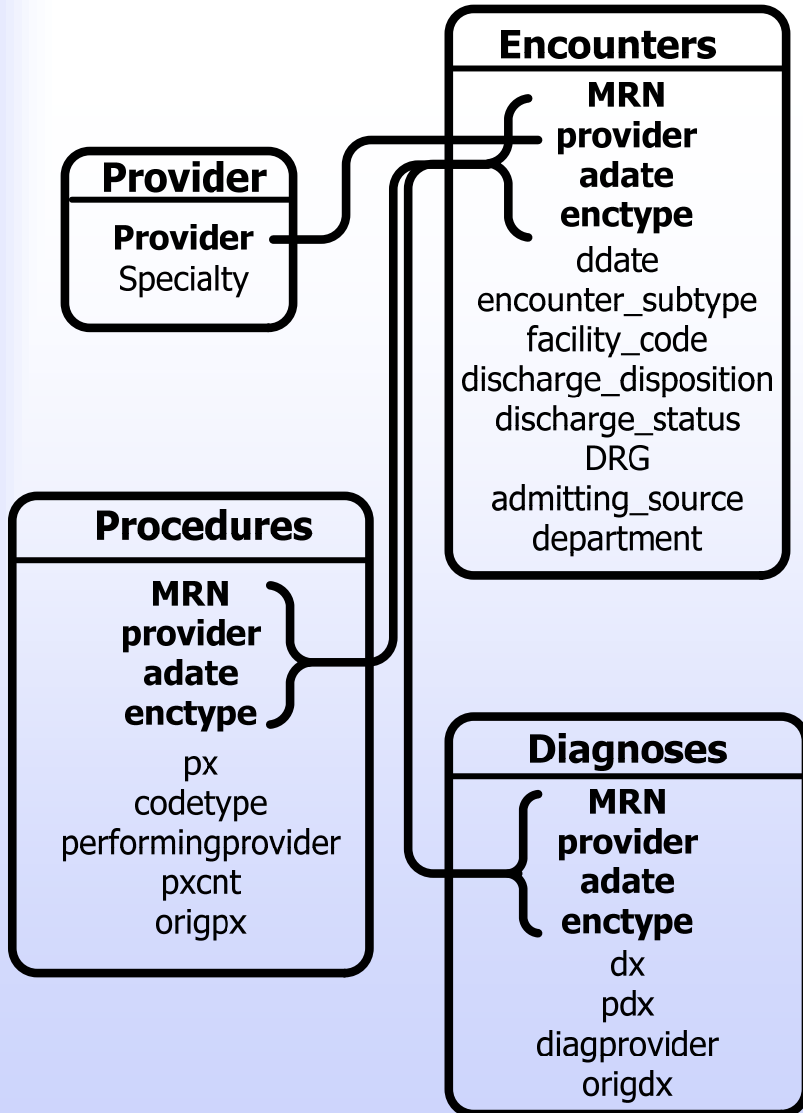
education vars...

income vars...

race vars...

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

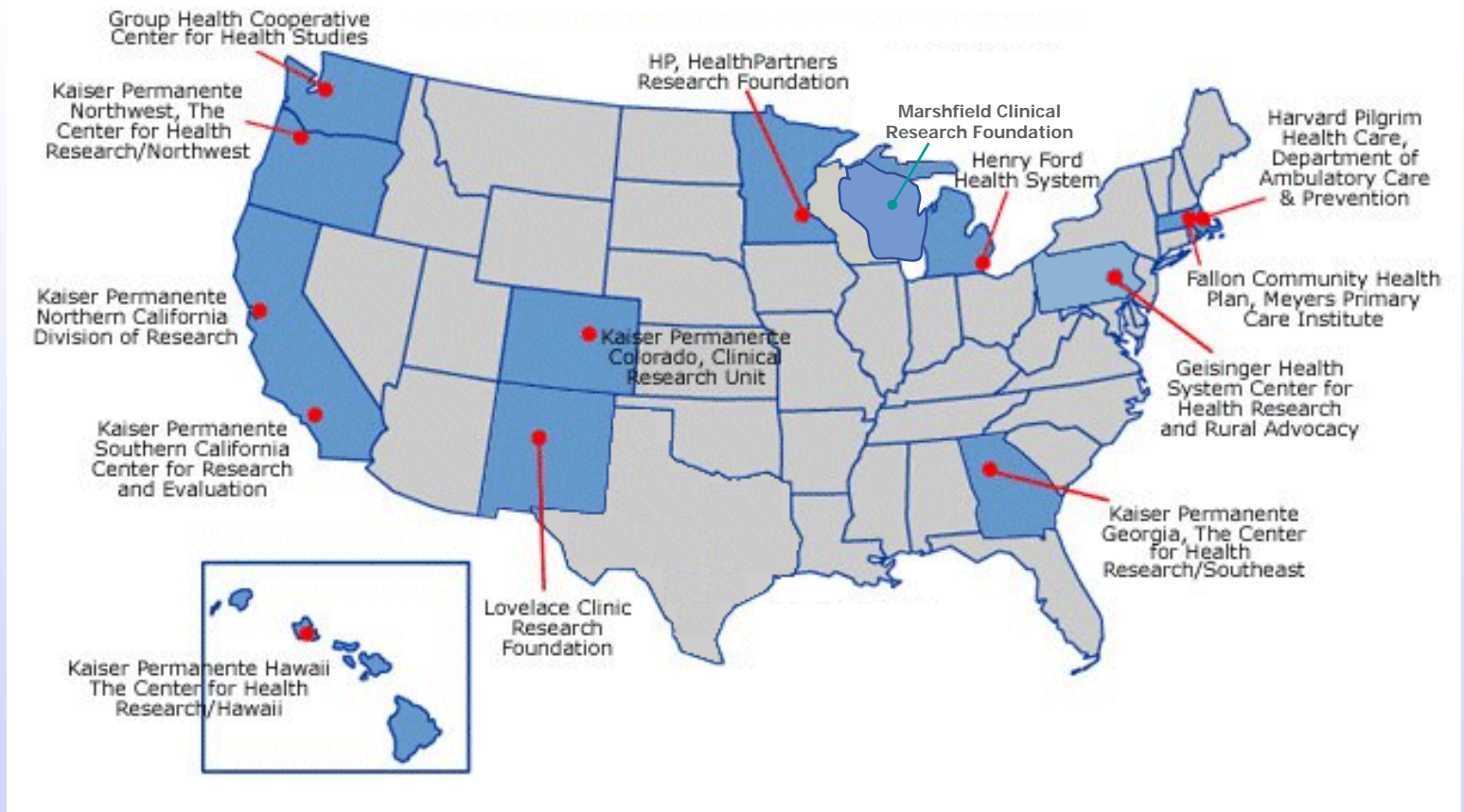
Encounters



- ❖ In- and Out-patient encounters
- ❖ Tons of macros for pulling people, procedures & diagnoses.

Sites Participating in VDW

Cancer Research Network Sites



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

6/26	Tyler learns about the project
6/27	Tyler learns about modification to request. Tyler writes program and sends to sites
6/28	Karen (@HFHS) returns results. Investigator sees results and modifies specifications
6/29	Tyler learns of spec changes, modifies program, resends to all sites. Karen returns new results.
6/30	Jenn (@KPCO) returns results. Tyler assembles 3 sites results and send to investigator.

Summarized data project- A.I.

Mar 2004	Formal study group forms
Mar 2004 – Jun 2005	Develop oncologist survey; sketches for identifying recurrence w/out abstraction, etc.
July 2005	Final table shells/NDC list developed.
Sep 2005	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• Combined output produced—provokes requests for tweaks & additional data.• Candidate program revised & re-sent to Site 2 (1 day turnaround).
Nov 2005	<ul style="list-style-type: none">• Biostat consult provokes further revision requests.
Dec 2005	<ul style="list-style-type: none">• 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.

Jan 2006	<ul style="list-style-type: none">• Data Issue: no ER marker data at Site 6.• Data Issue: no DxDate at Site 7.• Programming issue: Misleading program comment resulted in 2 sites missing fills from 2004—need to re-run at sites 4 & 5• Usable data returned by Site 8.
Feb 2006	<ul style="list-style-type: none">• Re-runs completed at Sites 4 & 5.
Mar 2006	<ul style="list-style-type: none">• Site 6 cures ER marker issue and returns usable data.• Final combined output distributed to group w/results from sites 1, 2, 3, 4, 5, 6 & 8

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

8/16	Roy learns of the project—shell tables provided.
8/28	Single-site output generated for feedback
8/29	Program sent to Site 2 for testing
9/6	Data sent to Site 4 for testing
9/7	Data back from Site 2; Program sent to Sites 3 and 5

Hybrid Project – BOW2

Site 1	8/29	No problems
Site 2	9/7	Testing site
Site 3	10/24	IRB delay (also: required that only site-aggregated results would be disseminated)
Site 4	11/20	Data availability delay (no tumor registry)
Site 5	12/19	Data availability delay

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