VDW Tutorial for Programmers

Stolen from Roy Pardee
Mangled by Gene Hart
The VDW is…

- …a series of **dataset standards** and **automated processes** at each of 11 sites,
- that allow a SAS program written at one site to be run against all the others
  - quickly
  - and with a minimum of site-specific customization.
The VDW is *Distributed*

- You could also call it a *Distributed Data Warehouse*
  - the data are real enough
  - but each site maintains its own separate slice of the whole.
The Visual Aid

Advance Work

- Site-Specific Databases

  Site-specific programs, run against their own databases

VIRTUAL DATA WAREHOUSE

- Real databases set up identically at the 11 CRN sites

Each Project

- Program written using common data dictionary
- Output = Datasets from 11 different sites

Research Team
VDW Data Areas

- Tumor
- Enrollment
- Demographics
- Outpatient Pharmacy
- Procedures and Diagnoses
- Census
- Vital Signs
Compared to CHS SAS DW

- MRN is person identifier
- NDC is drug identifier
- Single dataset per data area – not annualized
- Decimal points in DX codes
- Some data cleaning done
- Some quality checks comparing across HMOs
Where is VDW Documentation

- On CRN website
- Programmer’s Guide
- CHS Website
The Standard Macros

- In the course of writing VDW programs, we have come up with a series of SAS macros to use as building blocks for new programs.
- Full Details are in the handout, but here is a quick overview…
<table>
<thead>
<tr>
<th>Current Macros</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%NDCLookup</td>
<td>Searches EverNDC for drugs whose generic/brand names include user-supplied search strings</td>
</tr>
<tr>
<td>%GetRxForPeople</td>
<td>Pulls all Rx fills for a user-supplied list of people.</td>
</tr>
<tr>
<td>%GetRxForDrugs</td>
<td>Pulls all Rx fills for a user-supplied list of drugs.</td>
</tr>
<tr>
<td>%GetRxForPeopleAndDrugs</td>
<td>Pulls all Rx fills for a user-supplied set of NDC codes and people</td>
</tr>
<tr>
<td>%GetPxForPeople</td>
<td>Pulls all Procedures (inpt/outpt) for a user-supplied list of people.</td>
</tr>
<tr>
<td>%GetDxForPeople</td>
<td>Pulls all Diagnoses (inpt/outpt) for a user-supplied list of people.</td>
</tr>
</tbody>
</table>
## Current Macros (Cont.)

<table>
<thead>
<tr>
<th>Macro Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%GetPxForPx</td>
<td>Pulls all of a user-specified list of procedures.</td>
</tr>
<tr>
<td>%GetDxForDx</td>
<td>Pulls all instances of a user-supplied list of diagnoses.</td>
</tr>
<tr>
<td>%GetDxForPeopleAndDX</td>
<td>Pulls all diagnoses for a user-supplied list of DX and people.</td>
</tr>
<tr>
<td>%BreastCancer-Definition01</td>
<td>Pulls the set of incident breast tumors, based on the Early Screening criteria.</td>
</tr>
<tr>
<td>%PullContinuous2</td>
<td>Reduces an input dataset of people to those continuously enrolled over a user-specified period.</td>
</tr>
<tr>
<td>%Del DDset</td>
<td>Replaces true identifiers with an arbitrary StudyID.</td>
</tr>
</tbody>
</table>
%GetRxForDrugs

(DrugLst /* The name of a dataset containing the NDCs of the
drugs whose fills you want. */
, StartDt /* The date on which you want to start collecting
   fills. */
, EndDt /* The date on which you want to stop collecting
   fills. */
, Outset /* The name of the output dataset containing the
   fills. */) ;
How to get VDW macros

filename crn_macs FTP
"CRN_VDW_MACROS.sas" HOST = 
"centerforhealthstudies.org" CD = 
"/CRNSAS" PASS = "$blue33volcano#"
USER = "CRNReader" ;
%include crn_macs ;
Standard Variables

%include "\\ctrhs-sas\\warehouse\\Sasdata\CRN_VDW\lib\StdVars.sas;"

Contains lines such as:

%let _TumorLib = "\\ctrhs-sas\\warehouse\\sasdata\\crn_vdw";
%let _TumorData = tumor;
Look at VDW Guide

- Data Structures (Appendix 3)
- Walk through sample program
How to develop a VDW program

1. Familiarize yourself w/the data structures
2. Formulate the question
3. Choose sites
4. Decide on desired outputs
5. Get IRB approval
6. Write your program
7. Beta test at selected sites
8. Distribute to remaining sites.
9. Use Secure File Transfer to collect your data
1: Familiarize yourself with the data structures

- The most current versions will be on the CRN Website:
  https://www.kpchr.org/crn2/

- They are also reproduced in Appendix 3 of the VDW Developer’s Guide (handout).
2: Formulate The Question

- Always important
- Probably more important for a VDW project—the potential time waste is multiplied by the number of people involved.
3: Choose your Sites

<table>
<thead>
<tr>
<th></th>
<th>GHC</th>
<th>HPHC</th>
<th>HPRF</th>
<th>HFHS</th>
<th>Fallo n</th>
<th>KPNC</th>
<th>KPNW</th>
<th>KPSC</th>
<th>KPCo</th>
<th>KPH</th>
<th>KPG</th>
</tr>
</thead>
</table>

* Not yet available.
† Data available on a 10% sample
4: Decide on outputs

- What do you want back from the sites?
  - SAS datasets (recommended!)
  - Log files (recommended!)
  - Listing files
  - ODS Output (HTML, RTF, etc.)
5: Confirm IRB approval

- While this is the Investigators’ ultimate responsibility, don’t forget to make sure that your project has IRB approval to touch data at all your chosen sites.
- Allow for delays in herding the various IRB committees into a consensus.
6: Develop and Debug your program

- Program defensively
- Make things easy on the HMOs
- Use the variables defined in StdVars.sas to make code portable.
- Make things easy on yourself
  - Use already written macros and code snippets
7: Send ‘beta’ program to your test HMOs

- Expect additional bugs to appear as each additional site runs your program.
- Do one HMO at a time until program runs without errors.
- The ‘least virtual’ sites (e.g., those with standing datasets) usually give the best turnaround.
8: Send tested program to remaining sites

- Here again, expect this to be an iterative process.
- *Hopefully* the iterations are only caused by yet-undetected programming errors.
- Consider mocking up multi-site versions of your output, to cut down on the “Oh. That’s not what I meant.” factor.
9: Transfer data back to lead site using Secure File Transfer

- As part of the CRN Web Infrastructure, KPNW has put together an excellent Web-based application for moving files from one site to another: https://secure2.kpchr.org/crnfiletransfer
- Send mail to Gary Ansell for access: Gary.Ansell@kpchr.org
Getting Help: Sources

- Your Local Site Data Manager.
- The VDW-USERS listserv:
  - [http://lists.centerforhealthstudies.org/read](http://lists.centerforhealthstudies.org/read)
- Gene and Roy:
  - hart.je@ghc.org
  - pardee.r@ghc.org