The HMO Research Network Virtual Data Warehouse: A Model for Responsible and Efficient Multi-Center Health Data Use

**Context**

- Electronic health records (EHRs) are increasingly common in the United States.
- Financial incentives are available through Medicare to encourage health plan, hospital, and physicians to install EHRs that perform “meaningful use” function.
- Using EHRs for research and surveillance is compelling—but not easy especially in multi-site projects.

- Health records:
  - Are grouped
  - Include information in many formats.
  - Very wary in content, format, and structure from site to site

- The HMO Research Network Virtual Data Warehouse (VDW) is an example of a tool to facilitate responsible and meaningful use of electronic health record information across organizations.

**The HMO Research Network**

A consortium of 16 integrated health care delivery systems with affiliated research centers committed to closing the loop between research and clinical care. The HMORN fosters public health research collaborations in real-world settings.

- Benefits from multidisciplinary expertise, robust population diversity, and large statistical power.
- To make our work more efficient, we developed tables of data standards and software.

**The Virtual Data Warehouse**

The HMORN Virtual Data Warehouse (VDW) facilitates multi-site research while protecting patient privacy and proprietary health practice information. Originally developed by the HMO Cancer Research Network, the VDW now supports studies of cancer, drug safety, cardiovascular disease, mental health, and more. Participating centers convert their local data to VDW format.

- Source data that vary substantially within and across sites.
- Underlying data are collected for treatment, payment, and operations—not for research.

**How It Works**

- HMORN Governing Board provides overall policy direction about content, sources, and access.
- HMORN Assets Stewardship Committee addresses data development and quality priorities.
- VDW Operations Committee (VOC) manages cross-site development activities, with technical and scientific input.
- VDW Workgroups for specific data areas define, maintain, and interpret data file specifications, propose new variable specifications, and identify site-specific issues with data standards.
- VDW Implementation Group (SIG) members include Site Data Managers who extract information from local systems, work with the VOC and Workgroups, verify data file specifications, and share best practices.

**Availiability**

Each institution's VDW data remain at their site until a study-specific need arises. The required data are extracted after ethical, contractual, and IRB requirements are met.

The HMORN VDW is a common data model, not a central database.

**Challenges**

- Underlying data are collected for treatment, payment, and operations—not for research.
- Source data that vary substantially within and across sites.
- It takes time to:
  - Agree on the need for a new variable or data area
  - Develop clear specifications to guide implementers and end-users
  - Implement new variables at each site
  - Verify and document the implementations
- Consult with users throughout.
- Health plans continually change their information systems, often requiring adaptation or reimplementation at all implementing sites.
- Sharing data beyond project collaborators is complicated for technical, regulatory, and political reasons.

**How Researchers Use It**

1. Work with collaborators to develop the study protocol.
2. Obtain relevant regulatory, contractual and ethical approvals.
3. A SAS analytic program is developed at the lead study site.
4. The project sites run the program against their local VDW and return project-specific datasets (often aggregated) to the lead site for data pooling. This process may be iterative, depending on the availability of data within the VDW.

**The VDW is a cornerstone of collaborative research, protecting privacy and fostering standardization.**

**Adapting standards as health plan data evolve an example**

Kaiser Permanente (KP) introduced new data collection procedures for ethnicity and race in April 2009 and created 268 ethnic categories and 9 race categories. When patients report that their ethnicity is “Hispanic,” this is entered in the “race” variable in the electronic medical record because Hispanic is no longer allowed as an “ethnicity” response. This change inspired the Demographics Workgroup to develop plans for incorporating the new KP ethnicity and race data into the VDW format. It will likely lead to changes in the VDW definitions of ethnicity to allow for more accurate responses than the narrow set specified by the U.S. census ethnicity standard.

**Funding and Acknowledgement**

NIH National Cancer Institute U19CA079689; NIH National Center for Research Resources UL1RR025014; Group Health Cooperative.

Special thanks to: Karen Johnson, PhD; Sarah Greene, MPH; Roy Pardee, JD; Gene Hart, MS; Kate Siggerud, PhD; Jane Iseli, MPH; Mark C. Hendrick, PhD of Kaiser Permanente Northwest; and Daniel Ig, MPH, of Kaiser Permanente Northern California.

**Benefits**

- Improved data efficiency, accuracy, and completeness
- Analytical precision plus patient and institutional protections
- More generalizable results

**The VDW is an example of a distributed or federated data-sharing model based on electronic clinical claims, and administrative health care data. It is applicable for multi-site health services and population health research. With planning and ongoing funding, it yields data across multiple sites and over time. Benefits for multi-site analysis include:**

- **Data efficiency, accuracy, and completeness**
- **Analytic precision plus patient and institutional protections**
- More generalizable results