

THE VIRTUAL DATA WAREHOUSE

THE VDW

The Virtual Data Warehouse (VDW) is a series of data set standards and automated processes that allow SAS programs written at one Health Care Systems Research Network (HCSRN) site to be run against other VDW sites quickly and with a minimum of site-specific customization. It is 'virtual' in the sense that there is no centrally located store of data against which data from all sites can be touched in a single run. Distributed programs are manually run, results reviewed, and data returned by personnel at the sites who retain complete control over their local data. In general, VDW files contain all data captured in the health plan's clinics, as well as claims submitted by third party providers.

In addition to streamlining multi-site research, use of the VDW provides IHR analysts and programmers efficient access to data from a variety of sources and systems that have already been quality checked. There are currently 17 different HCSRN sites that have implemented one or more VDW files.

DATASETS COVER:

CONCEPT AREAS	
Enrollment	Contains a record for every period of enrollment for each person enrolled in the health care plan.
Utilization (e.g., encounters, procedures, diagnoses)	Document encounters between health care professionals and patients, both inpatient and outpatient. Data are sourced from any or all of: integrated electronic medical record systems (EMR), pre-EMR legacy electronic systems, or claims data. Each encounter can have any number of associated diagnoses and procedures.
Pharmacy	Document outpatient pharmacy fills. Can be sourced from either or both of internal electronic systems and claims data.
Tumors	Incident cancers, typically sourced from SEER data.
Demographics	Birthdate, gender, race, and ethnicity (as available) for everyone enrolled or ever treated at the health plan.
Lab Results	Contains laboratory test and results information for a limited set of tests. Sourced from EMR or legacy internal systems.
Vital Signs	Contains height, weight, BMI, and blood pressure on all visits captured in the EMR.
Death	Contains death date and cause of death information.
Census Demographics	Contains 2000 census information based on home address. In addition to a Geocode it contains information on education, income, housing, and race information based on a person's neighborhood based on home address.
Bone Mineral Density	Contains calculated BMD, t-score, scan date/time, location, and fracture risk scores.
Geographically Enriched Member Socio-Demographics	Contains race probabilities based on Bayesian Improved Surname and Geocoding methodology and non-PHI geographic descriptors.
Infusion	Contains data on drugs that were ordered, dispensed, and administered to patients in the infusion center, and treatment plan.
Patient Reported Outcomes	Contains data from patient self-administered questionnaires (BPI, PHQ-2, PHQ-8, PHQ-9, PHQ-T, and EVS).
Problem List	Contains status of patients' problem lists in the EHR.
Social History	Contains various behavioral measures such as use of tobacco, alcohol and illegal drugs, sexual behavior, and contraceptive use.