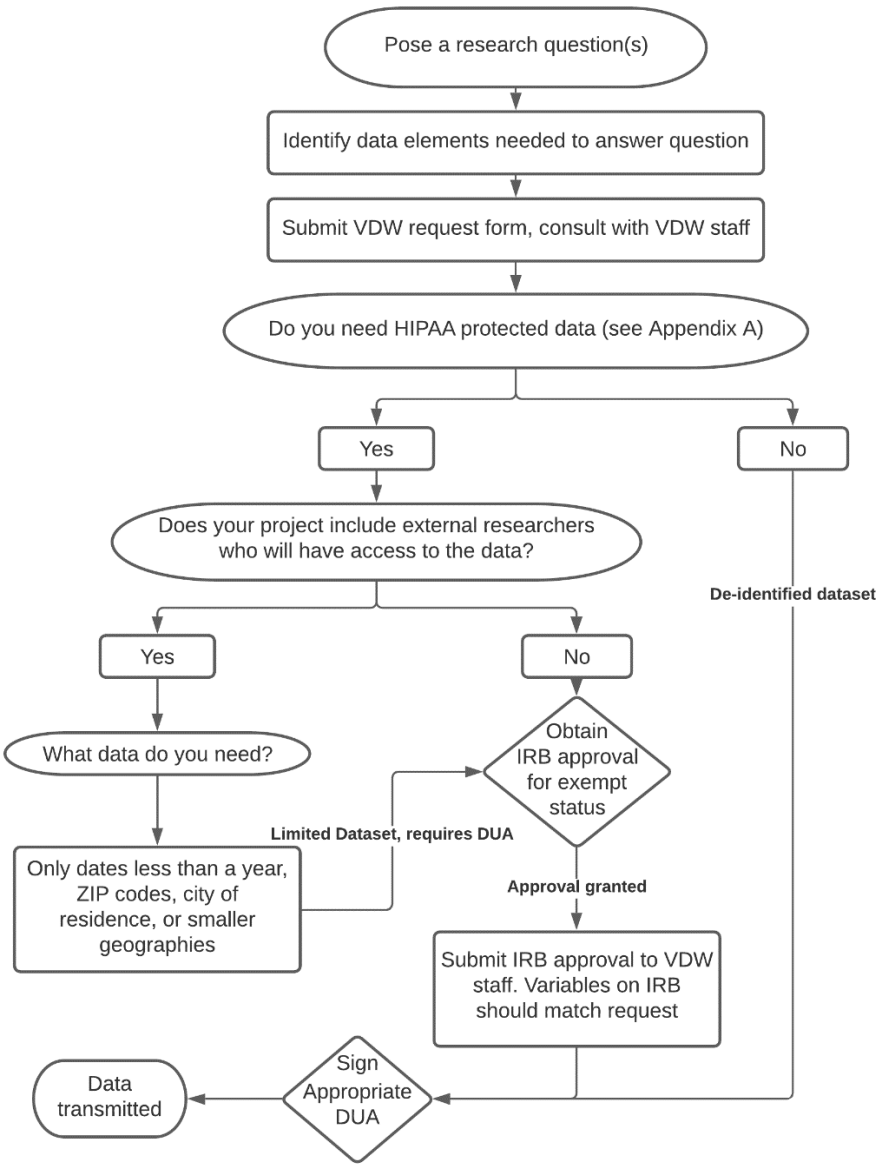
VDW USER REQUEST GUIDE

1. PURPOSE

The purpose of this document is to provide an overview of data available, request process, cost, and requirements of data use for the Virtual Data Warehouse (VDW) for research at Hennepin Healthcare and Hennepin Healthcare Research Institute (HHRI). VDW staff reserve the right to make adjustments to information as described in the guide to match the particularities of a given project. For questions, please email VDW@hhrinstitute.org

2. REQUEST PROCESS OVERVIEW



3. ABOUT THE VDW

Hennepin Healthcare (HHS) and HHRI implemented a common data model following the Health Care Systems Research Network (HCSRN) VDW specifications. Hennepin Healthcare developed the VDW in order to facilitate internal research and collaborative research studies with external organizations. Even though Hennepin Healthcare is not a formal member of the HCSRN network, Hennepin Healthcare can collaborate with any organization that can provide standardized data.

4. ELIGIBILE INVESTIGATORS

Hennepin Healthcare (HHS) and Hennepin Healthcare Research Institute (HHRI) faculty, trainees, or staff who have an hcmed.org, hhri.org, berman.org or cdrg.org or email address are eligible to submit a data request. External (non-Hennepin Healthcare) collaborators may participate in research projects using the VDW but the project must include a Hennepin Healthcare co-investigator who will initiate the request. See Section 6. Data Access and Documentation for more information on how HH data are shared with collaborators.

5. DATA AVAILABLE

The VDW contains information from HHS EPIC medical chart and billing data. The VDW is based on a common data model developed by the Health Care Systems Research Network (HCSRN)[1]. The HCSRN common data model can be adapted to other data models, including the PCORnet and OMOP common data models. The common data model defines standard tables and variables so investigators, both internal and external, have a basic understanding of the data and can easily share programming code across sites for collaborative studies.

The Hennepin Healthcare VDW includes twelve domains. They include: Death, Demographics, Diagnosis, Encounters, Facility, Lab Results, Languages, Med Orders, Procedures, Providers, Social History, Census Location (addresses) and Vital Signs. See the "VDW Master Data Dictionary" on the Hennepin VDW website for a description of each table, the variables contained in each table, variable formats, and variable code tables.

5.1 POPULATION

The VDW includes all patients who have been seen at a HHS location. This includes Hennepin Healthcare (service area 2) and Minnesota Visiting Nurse Agency (MVNA) (service area 21). The VDW excludes patients who have opted-out of research (research and clinical trials) as well as ambulance only and affiliate site encounters. The VDW includes over 770,000 patients and 11 million encounters.

5.2 DATA AVAILABILITY AND REFRESH

Data are available from February 1, 2007 to the most current week. The VDW is refreshed daily at 12:00am.

[1] HSCRN Virtual Data Warehouse (Health Care Systems Research Network, 2019)

5.3 SENSITIVE VARIABLES AND IDENTIFIERS

The VDW data tables include personally identifying information (PII) as well as personal health information (PHI) (e.g., Patient id, dates, and patient ZIP code) in order to accommodate a variety of research requests. Note that HIPAA protections include the *minimum necessary standard*, i.e., that PII/PHI should not be disclosed unless it is necessary to conduct research. Hennepin Healthcare investigators requiring identifiable data for secondary research will receive VDW variables in the original identifiable format; such data must be accessed/stored on a Hennepin Healthcare server in compliance with HIPAA security standards. For secondary research involving identifiable data, IRB review is required (see section 7. IRB REVIEW).

For secondary research involving only de-identified data, the VDW uses the HIPAA Safe Harbor method of de-identification [2] to create the dataset. IRB review is not required for secondary research involving only de-identified data.

For datasets containing identifiers, the VDW will establish a Data Use Agreement (DUA) and will require the investigator to provide documentation of IRB review prior to transmitting data.

The VDW provides de-identified datasets as defined in Appendix A.

If the investigator requires a unique key in order to link VDW tables together, staff will strip the VDW patient identifier (MRN) and encounter identifier (ENC_ID) and replace each with a de-identified pseudo-ID. VDW staff will retain the crosswalk until the end of the study in case the investigator needs the data rerun. Key crosswalks will be stored on secure servers only available to VDW staff and will not be available to investigators.

See Appendix A for a table listing all sensitive variables, the table(s) they are found in, the original variable format, and the alteration of the variable, if provided as a de-identified dataset.

6. TYPES OF REQUESTS AND COST

6.1 PREPARATORY RESEARCH REQUESTS

Requests for summary level statistics for the purpose of preparation for research are usually free. An investigator may submit a VDW request for preparatory research using our request form at https://is.gd/hhri_vdw and selecting the option “Preparatory research” in the question about requested services on the first page. If you have any questions or require assistance you can email VDW staff at VDW@hhrinstute.org

6.2 RESEARCH REQUESTS

Investigators are encouraged to contact the VDW early in the process of developing their research to consider all data requirements and whether the dataset will contain identifiers.

[2] De-identification Standard described (Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 2019)

To submit a research request, fill out the Data Request Form at https://is.gd/hhri_vdw.

INVESTIGATORS MAY REQUEST VDW DATA BASED ON ANY OF THE POPULATED VARIABLES IN THE VDW. REVIEW THE DATA DICTIONARY ON THE VDW WEBSITE FOR A LIST OF TABLES (AKA DOMAINS) AND VARIABLES THAT MAY BE INCLUDED. IF DATA IS NOT AVAILABLE IN THE VDW BUT IS AVAILABLE IN EPIC CLARITY, A HYBRID VDW-CLARITY EXTRACT MAY BE POSSIBLE.6.3 COST

Most projects cost the standard hourly rate of $120/hr. For federally funded projects, as well as other projects as required by the funder, charges are determined based on the VDW staff effort needed to provide the data necessary for the project. Analysis services will also be charged according to effort. Investigators may request a cost estimate when submitting a prep for research request as described in section 4.1. VDW staff will supply an estimate based on an expected hours required to complete the request. Analysis services, if requested, will be included in this estimate. For projects with a primary investigator from outside of HHS or HHRI, payment will be negotiated as part of the contract. Resident/education based projects and/or non-sponsored projects may request a fee waiver which can be made during the initial consultation with VDW staff.

7. IRB REVIEW

7.1 PROJECTS REQUIRING IDENTIFIABLE DATA

Secondary research involving identifiable data requires IRB review. Documentation of review (e.g., exemption letter) must be provided to the VDW prior to the transmittal of data. Visit the Human Research Protection Office (HRPO) website for additional information and access to the online submission system, Cayuse Human Ethics (HE)

See the diagram in Section 2 of this document and the list of HIPAA identifiers in Appendix A for more details.

7.2 PROJECTS REQUIRING NO IDENTIFIABLE DATA

Unless it is subject to FDA oversight, secondary research that involves only de-identified data (see Appendix A) does not meet the definition of *human* research and, therefore, does not require IRB oversight. Contact VDW staff to confirm that your data request meets the definition of de-identified data. If you have questions regarding IRB oversight, contact HRPO@institute.org.

7.3 EXTERNAL PROJECTS

Projects originating from other institutions may be covered by an IRB outside of HHS, but proof of IRB review will have to be submitted to the HHS IRB and accepted before transmittal.

8. LIMITED DATASETS

A Limited Data Set (LDS) may be established to disclose a limited set of PHI (in accordance with HIPAA) to an external researcher. Consult with VDW staff for additional information.

9. DATA ACCESS & DOCUMENTATION

9.1 DATA SPECIFICATIONS & COST ESTIMATION

Once a request is received, VDW staff will contact the requestor to confirm the details of the request. VDW staff will outline the required data pull, summarization or analysis required along with an estimate of costs. Requestors will have the opportunity to request modifications in the data specification/estimate. Once the specification/estimate are approved, work on the data pull can commence.

9.2 RELEASE OF DATA

The VDW Analyst will create the data file per the investigator's specifications. If the data will reside internally and are de-identified, the VDW Analyst will make data available by email, OneDrive or a similar method. If the data file contains identifiable data), then the VDW Analyst will share the file via a secure folder. Instructions for data retrieval will be emailed to the investigator. A Data Use Agreement (DUA) will be established prior to transmittal of identifiable data.

If the data are to be stored outside of Hennepin Healthcare server, the VDW Analyst will arrange data access through a secure file transfer protocol. External storage will require a signed Data Use Agreement (DUA) between Hennepin Healthcare and parties using, storing or transmitting the data.

The list of files that may accompany a data extract and their descriptions appear below.

Document	**Description**
[YYYY.MM.DD]_[filename].csv	Data file(s)
[YYYY.MM.DD]_[filename]_data_dictionary	Data dictionary including variables names and definitions
[YYYY.MM.DD]_[filename]_request	Document will include investigator and project information, the inclusion/exclusion criteria, the de-identification method used, and the final file layout.
[YYYY.MM.DD]_[filename]_DUA	Copy of signed data use agreement (as appropriate)
[name]	IRB documentation (provided by investigator)

5. TECHNICAL SUPPORT AVAILABLE

At any point in the process, an HH investigator or member of the study team may contact VDW staff for assistance. VDW staff can assist with study feasibility, sample size calculations, data loading issues, and data interpretation to name a few.

Email VDW@hhrinstute.org with your questions.

6. DATA RETENTION

Following the NIH guidelines for records retention, the VDW data should be retained for a period of at least three years from the date the study is closed-out with the IRB for internal investigators and at least three years from the end date of the Data Use Agreement for external investigators or projects not requiring IRB oversight. After the retention period has expired, the investigator is responsible for destroying the data, including all raw data and any analytical files, as well as notifying the VDW that the data has been destroyed.

7. PUBLICATION OF FINDINGS

Any publications or presentations based all or in part from data sourced from the VDW are required to include the following citation.

"This research was supported by the National Institutes of Health's National Center for Advancing Translational Sciences, grant UL1TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health's National Center for Advancing Translational Sciences."

Please report any publications using VDW data to VDW staff. Citing your work helps keep the VDW funded and available for studies like yours. You can email a short synopsis and citation of your work to VDW@hhrinstitute.org. Please Include a short description that includes key points, the research question, the findings, and meaning.

APPENDIX A: VDW SENSITIVE VARIABLES, TABLES, AND FORMATS			
Variable Name	Table(s)	Original Format	De-identified Format
Medical Record Number	Demographics, Language, Encounter, Diagnosis, Procedure, Med Order, Lab Results, Vital Status, Social History	Pat_id (number)	Randomly generate a patient id (NOT based on Pat_id or MRN)
Patient birth date	Demographics	MM/DD/YYYY	Birth Year, if 89 or >, then present as 89
Encounter id	Encounter, Diagnosis, Procedure, Vitals, Social History, Med Orders	Enc_id (number)	Randomly generate an encounter id (not based on the true encounter id)
Admit date	Encounter, Diagnosis, Procedure	MM/DD/YYYY	YYYY OR Days to event
Discharge date	Encounter	MM/DD/YYYY	YYYY OR Days to event
Procedure perform date	Procedure	MM/DD/YYYY	YYYY OR Days to event
Order date	Med Order, Lab Results	MM/DD/YYYY	YYYY OR Days to event
Med start date	Med Order	MM/DD/YYYY	YYYY OR Days to event
Med end date	Med Order	MM/DD/YYYY	YYYY OR Days to event
Specimen collection date	Lab Results	MM/DD/YYYY	YYYY OR Days to event
Test result date	Lab Results	MM/DD/YYYY	YYYY OR Days to event
Measure date	Vital Status	MM/DD/YYYY	YYYY OR Days to event
Contact date	Social History	MM/DD/YYYY	YYYY OR Days to event
Birth Control Comment	Social History	Free text	Not supplied
Alcohol Ounces Week	Social History	Free text	Not supplied
Alcohol Comment	Social History	Free text	Not supplied
Illicit Drug Use Freq	Social History	Free text	Not supplied
Illicit Drug Use Comment	Social History	Free text	Not supplied
Tobacco Use Years	Social History	Free text	Not supplied
Tobacco packs day	Social History	Free text	Not supplied
Tobacco smoked/smokeless start date	Social History	MM/DD/YYYY	YYYY OR Days to event

Tobacco smoked/smokeless quit date	Social History	MM/DD/YYYY	YYYY OR Days to event
Tobacco Comment	Social History	Free text	Not supplied
Death Date	Death	MM/DD/YYYY	YYYY OR Days to event
ZIP code	Census Location	Free text	3-digit ZIP code
Address1	Census Location	Free text	Not supplied
Address2	Census Location	Free text	Not supplied
City	Census Location	Free text	Not supplied