Colorado Health Observation Regional Data Service (CHORDS)
Governance Plan
Version 1.4

I. Introduction

The Colorado Health Observation Regional Data Service (CHORDS) began in 2011. It is a regional collaborative partnership among Colorado health providers, public health departments, the Colorado Regional Health Information Organization (CORHIO), and the University of Colorado Anschutz Medical Campus to share health data. CHORDS collects, analyzes and presents data from participating partners' electronic health records (EHRs) to monitor population health, target areas for intervention and conduct research and evaluation activities.

CHORDS uses a distributed data approach in which data partners maintain physical and operational control over their electronic data stored in virtual data warehouses. CHORDS aggregates data so that findings are never attributed to specific data partners nor individual patients.

CHORDS is committed to patient privacy and believes that the protection of privacy, confidentiality, and data security is essential to the existence and success of public health monitoring and research. All data are securely exchanged through a web-based, password protected portal accessible only by approved users. All CHORDS queries remove patient identifiers before information is shared and adhere to privacy-preserving guidelines. Data partners and data users are responsible for proper stewardship of CHORDS data in their possession.

This document establishes governance policies and guidelines based on the guiding principles of good will, trust and appropriate stewardship of data among all participating partners. It addresses activities, duties and expectations regarding data curation, query processing, permitted data uses, participating partners' rights and responsibilities and other topics. This document identifies a governing structure to implement guidelines, oversee CHORDS' development, operationalize changes and engage stakeholders.

This governance plan will align policies, guidelines and processes for public health monitoring, evaluation and research activities to reduce costs of parallel structures and operations, reuse knowledge and support a learning health system.

These policies and guidelines are flexible. They will be revised as needed to meet changing circumstances and reviewed at least annually. When necessary for specific research activities, supplemental materials will be developed.

Revision History

Revision Number	Date	Comment
Version 1.0	11/16/2016	First draft to committee
Version 1.1	12/2/2016	Project development work group created, revised public health presentation and publication guidelines and appendix documents.

Version 1.2	12/19/2016	Revisions based on Governance Committee member comments.
Version 1.3	2/15/2017	Added CHORDS Network graphic; incorporated revisions from Governance Committee members; moved Revision History.
Version 1.4	08/17/2017	Added details on research governance throughout.

A. Guiding Principles

CHORDS is committed to public health surveillance and research that improves the health of Coloradans.

CHORDS is a collaborative project that relies on the good will and trust of participating partners. Essential to this trust is appropriate stewardship of private and confidential health information. Each partner respects and honors the autonomy of their site and others and recognizes the benefits and responsibilities of using this information to serve public health and research.

The following principles reflect the core values that guide the governance process and are reflected in the governance plan.

I. The Governance Board will:

- Establish, document and conduct transparent decision-making processes.
- Facilitate high-quality monitoring, evaluation, research and health improvement activities.
- Identify and address new governance issues as necessary.
- Review, revise and approve governance policies on an annual basis or more frequently as needed.

II. Data Partners will:

- Retain autonomy in sharing their data.
- Provide efficient stewardship of site and network data by leveraging resources.
- Engender collaboration and leadership within CHORDS.
- Strengthen and ensure compliance with site-specific, local, state and federal policies and regulations.

III. Data Users will:

- Adhere to responsibilities for data access and use.
- Use data to enhance evidence-based health care and public health practice.
- Foster innovative research and monitoring methods to address and improve health.
- Assess population measures and discover generalizable knowledge for the public domain.

- Encompass diverse perspectives: patient-centered, population-based, provider and health care delivery systems.
- Maintain participating data partner confidentiality by never attributing data to a site nor comparing data from one partner site with data from another.
- Rapidly disseminate findings into the public domain.
- Be vigilant in protecting patient confidentiality and privacy.
- Engage stakeholders regarding data, findings and population metric decision-making.
- Provide aggregate data and public recognition to data partners when requested.

B. Scope

CHORDS was initially established by partner organizations affiliated with the Colorado Clinical and Translational Science Institute (CCTSI). It has grown and developed using funding from several federal, state and foundation public health surveillance and implementation grants and contracts.

In 2015, the Colorado Health Foundation awarded the Colorado Health Institute (CHI) a two-year, $1.9 million grant to support activities including updating CHORDS technology, expanding CHORDS data partners and users, developing a formal governance structure and establishing a sustainability plan so CHORDS can continue to develop in the future. CHI serves as the convener for the CHORDS Network (see figure below).

CHORDS is currently supporting public health activities in the Metro Denver region. Initial priority areas of focus for public health include cardiovascular disease, diabetes mellitus, mental health and obesity. Our intention is to expand beyond the Metro Denver region as we gain greater interest from other data partners and LPHAs.

CHORDS can also support research (e.g., observational, clinical, or health services studies) across multiple institutions. Research and other potential uses for CHORDS will be incorporated into a sustainability plan.

CHORDS is committed to student education at numerous degree levels. The Network is exploring options for supporting appropriate student projects that balance available time and resources and meet the sustainability mission.

Specific aims for CHORDS include:
- Advance trustworthy, efficient, responsive and regulation-compliant data sharing in support of collaborative, regional public health initiatives and research, across patient populations.
- Expand network scale and utility toward an array of public health and research activities.
- Engage a broader community of stakeholders to maximize overall utility.
- Define a community-engaged, sustainability plan for CHORDS beyond the next five years.

CHORDS Network (to include data partners, data users and overall governance structure)

II. Organizational Structure

CHORDS governance consists of a Governing Board and Executive Committee, Participants, Advisory Councils and Work Groups. Participation in these bodies is voluntary. These bodies and the organizational structure may evolve as the Network identifies a sustainable home for CHORDS.

CHORDS is not a legal entity; therefore, each organizational partner within the Network must first adhere to the policies and procedures of that organization as well as CHORDS data sharing agreements.

A. Governing Board
The Governing Board ("Board") has overall responsibility for CHORDS, providing leadership and accountability in the following areas:

- Strategic direction;
- Financial coordination and sustainability, including guiding and overseeing grant applications and budgets and identifying new resources;
- Initiatives and projects, including research;
- Recruiting new data partners; and
- Formalizing relationships among data partners, including executing data agreements.

The Board consists of data partners, data users and key stakeholders including community members and technology partners. All data partners will be invited to designate one Board member. Board members with multiple roles in CHORDS may represent only one role while serving on the Board.

The Board will have a chairperson and will designate five Members to serve on the CHORDS Executive Committee. Board members may also serve as chairpersons for CHORDS Advisory Councils and Work Groups.

B. Executive Committee
The Executive Committee represents the CHORDS Network among external stakeholders and identifies growth opportunities for Board consideration. The Executive Committee is also charged with guiding and informing the Board's strategic direction and preparing items for Board consideration and review. The Board may charge the Executive Committee with decision-making authority in certain circumstances to allow the Board to be more nimble and efficient and reduce response time for time-sensitive issues.

The five-member Executive Committee provides balanced representation of CHORDS data partners, data users, and stakeholders.

Members of the Executive Committee are responsible for convening the Advisory Councils and Work Groups. The Executive Committee will make recommendations to the Board regarding activities and responsibilities tasked to Work Groups.

C. Participants
> 1. Data partners
> Data partners are organizations that provide health observation data to be queried by CHORDS data users or services to catalog, curate, manage or improve such data. Data partners include health care and mental health providers; service providers, such as the Colorado Department of Public Health and Environment (CDPHE), for geocoding of health observation data; and the Colorado Community Managed Care Network, which serves as an intermediary managing data for Community Health Centers. A current list of data partners is available on the CHORDS website.

2. Data users
Data users request partner data for public health monitoring and evaluation, as well as for research studies. Data users must have legal agreements with data partners and complete the CHORDS User Access form. Current data users are described below.

a. Public health agencies
State and local public health agency officials use CHORDS to monitor population health, target and evaluate interventions and collaborate with communities to support health.

b. Researchers
Researchers, including students, from a variety of fields such as medicine, public health, behavioral health, pharmacy, sociology and anthropology may seek permission to use CHORDS data to answer clinical and health services research questions. The use of data for research purposes will require vetting of ideas, different legal, regulatory and institutional review agreements than data used for public health activities.

The Board may further expand the types of users in the future.

D. Advisory Councils
CHORDS has two Advisory Councils comprised of key stakeholders that provide the Board with input and feedback on CHORDS' direction and activities. The Councils meet as needed but at least twice a year and are chaired by a Board member.

1. Community Council
The Community Council will include representatives from community organizations, patient and consumer advocacy groups, and funders. The Council will provide input on current public health activities that use data from CHORDS and identify priorities for future uses of data. The Council will also review information output (e.g., maps, tables and other public materials) prepared using data from CHORDS and provide recommendations for improving dissemination and engaging other stakeholders.

2. Research Council
The Research Council (when convened), will include representatives from the University of Colorado Anschutz Medical Campus, funders, data partners and community-based research organizations. The Research Council will assist researchers in identifying topic areas of interest, feasibility of research domain and data access, applying for research grants, selecting research projects and aiding researchers in all phases of ongoing projects. Each data partner will independently decide whether to participate in any given research study.

E. Work Groups
Three Work Groups made up of CHORDS data partners and data users address operational issues. The Board may decide to create additional Work Groups at any time as needed. Work

Groups are convened and staffed by an Executive Committee member. Work Groups will meet at least quarterly.

1. Data Work Group
   The Data Work Group is responsible for identifying requirements and standards for data curation, exchange and use. The Data Work Group oversees the process for defining the data model, proposing modifications including new tables and variables or changes to existing ones and scheduling their implementation. This group also plans query interface changes; and is responsible for activities related to data definition and harmonization, and data quality assurance. These activities may require handling of confidential data not for public release. The Data Work Group may convene Communities of Practice among data partners around specific data related needs.

2. Network Operations Work Group
   The Network Operations Work Group is responsible for day-to-day oversight of CHORDS operations, including installing, testing, maintaining and developing CHORDS data sharing software. This group is also responsible for implementing all network access and security privileges for data partners and data users, including scheduling of technical activities and identifying necessary resources; and prompt reporting of any data privacy or security incident to the Board, affected organizations, and appropriate regulatory authorities.

3. Project Development Work Group
   The Project Development Work Group is responsible for fostering high-quality monitoring, evaluation, quality improvement and research activities through CHORDS. This may include assisting users in crafting their questions, assessing project feasibility and advising the Board or Executive Committee on prioritizing opportunities to demonstrate the value of the CHORDS Network. The Project Development Work Group develops and reviews new and existing projects and advises on public health and research uses; and manages a spectrum of project planning and timelines. It will work with existing data users and engage new ones. The Project Development Work Group may convene Communities of Practice among data users around specific projects and opportunities. As the need arises, separate subgroups may be convened for public health users and researchers.

III. Decision-Making Strategies and Policies

A. CHORDS Network
   Members of the Board are responsible for decisions that impact the CHORDS Network. Recommendations to the Board may come from CHORDS councils and work groups. The Board recognizes that each organizational partner is bound by the policies and procedures of that organization, as well as CHORDS data sharing agreements.

B. Decision-making Procedures
   The CHORDS Governing Board, councils and work groups will follow CHORDS decision-making procedures. Consensus-based decisions are preferred; however, decision-making processes will vary depending upon the issue. The chairperson of the Board, councils or work group will decide,

with input from the members, whether to seek consensus or use a voting process and how to carry out either process.

When a consensus-based approach does not result in a decision, or when voting is preferred, a simple voting process will be used so activities can move forward. Formal votes require a quorum (more than 50 percent of members). Members may abstain or recuse themselves from any vote. Each member is entitled to one vote; data partner and data user organizations are limited to one vote based on their designated role on the Board, council or work group. A motion carries when a simple majority (51 percent of those present) is reached. Votes may be taken verbally or using web-based voting software. Individual votes may be public or anonymous. All decisions will be documented and include a rationale.

C. Policy Enforcement
The Board is responsible for upholding CHORDS guiding principles and enforcing CHORDS policies and guidelines. The Board may delegate certain oversight activities to the Executive Committee.

Emphasis will be placed on prevention or early detection of noncompliance with Network policies, with prompt and confidential communication with data partners or data users who are not following policies or are at risk of compromising established policies.

D. Dispute Resolution
Efforts to resolve disputes between data partners, data users and other members of the CHORDS Network will begin at the most decentralized level. If resolution is not achieved, the Executive Committee will be alerted so it can advise and assist in the process. If the dispute still cannot be resolved, the Board will make a final decision.

E. Stakeholder Role in Policy Development and Decision-making
Members of the Community Council and/or Research Council are important resources to consider issues or identify concerns. Members should be included in deliberations wherever possible, as their time and interest allow.

F. Projects or Requests for Collaboration
Suggestions for projects or requests for collaboration may arise from multiple sources inside and outside of the CHORDS Network. The Board is responsible for identifying, and securing when appropriate, funding and resources as needed to carry out projects.

1. Public Health
Public health data users may request data to support regional collaboration between authorized CHORDS data users and other public health agencies. One public health data user may initiate a request on behalf of other users. Public health agencies may also request data that will be used by community partners to inform shared activities.

2. Research
Researchers who wish to use CHORDS data should contact the Research Project Manager or a member of the Project Development Work Group. After evaluation by the PDWG, each research request will be brought to the Governing Board by the Project Development Work Group. Members of the Governing Board will vote on all research requests. Each research

project needs to comply with research regulatory requirements such as executing data agreements, obtaining IRB approval or non-human subjects research designation, and any other institutional approvals required by individual data partners. When necessary for grant writing or assessing feasibility of a project, preparatory to research requests can be completed using CHORDS data after project approval by the Board (or Executive Committee if so delegated).

G. Participation of Principal Investigators (PIs) and Data Partners in Scientific Proposals and Projects
In order to support research projects using CHORDS, each data partner will identify a Site PI who will represent his or her site in the Project Development Work Group. Each Site PI is responsible for internal site review of potential projects and for final decisions regarding participation in each project. Site PIs for data partners who decide to participate in a specific research project will work closely with that project's designated PI.

Each data partner reserves the right to participate in or to opt out of any research project at any time. Reasons why a data partner decides to opt out should generally be disclosed to the project's PI or a member of the Executive Committee.

IV. Data Infrastructure and Governance
Data partners use different electronic health record (EHR) vendors and different clinical, administrative, and patient data-access applications to support clinical care and to collect patient data. Institutional differences in configurations, workflows, operational impact on real-time systems, and codes prevent sharing data directly from existing systems, even among partners using the same EHR product.

The approach chosen by CHORDS to resolve these barriers to multi-institutional data sharing is to require that data partners adhere to a common data model, which provides unambiguous definitions for structuring each data element and assigning codes to data values.

CHORDS adapted its data model, the Virtual Data Warehouse (VDW), from the Health Care Systems Research Network (formerly HMO Research Network). Data governance principles have been adapted for CHORDS from prior scientific networks including the Patient Outcomes Research To Advance Learning (PORTAL) Network and the Scalable Partnering Network (SPAN) for Comparative Effectiveness Research.

A. Distributed Data Approach
CHORDS uses a distributed data approach in which data partners maintain physical and operational control over their electronic data. Data partners and data users are responsible for proper stewardship of CHORDS data in their possession.

Governance principles for distributed data include:
- Data partners retain full physical and operational control over the content and availability of their patient-level and local organizations' data.
- Datasets containing patient identifiers are not shared outside of the data partner's local institution for public health purposes. Sharing data that are not fully de-identified (data that

are part of a limited data set) for public health purposes, while allowable under the HIPAA Privacy Rule, would require more detailed data sharing agreements that follow data partner sites' regulatory requirements.

- Sharing patient-level records for public health purposes may require further IRB clarification and institutional approval, as well as appropriate data sharing agreements. (See Appendix A.).
- Each site has the right to opt in or opt out of requests to share health data, including counts and/or patient-level records for research purposes. Researchers must also obtain IRB approval or a non-human subject research designation, as well as appropriate data sharing agreements for each participating site.
- For research studies, researchers must agree that they will limit their use of the data to the purpose(s) stated in the research plan approved by the Governance Board. They must also identify a data storage and destruction plan for any data obtained from CHORDS. This plan must be included in the application submitted for IRB approval. Misuse of CHORDS data will result in 1) termination of access to CHORDS data, 2) report of misuse to the study's IRB and 3) any additional consequences as outlined in study-specific DUAs.
- For research studies, all research personnel who will work with CHORDS datasets must be identified in the study's IRB application and must have successfully completed IRB and HIPAA trainings as required by their home institution.
- Participation in projects requires adequate resources for the work requested.
- Data partners have the right to determine if they will participate in a specific data request.
- Data partners must have a minimum level of participation in CHORDS requests to be considered active contributors. Data partners who do not maintain active participation in responding to requests will be evaluated as described in section III.D.

The CHORDS PopMedNet (PMN) application is used to query network data (make a "request") that meet specified clinical or demographic criteria. Results come from data partners who supply data from their institutional databases stored in their local instances of the VDW (format described below). Each data partner establishes its internal processes for determining whether the Data Mart Client (DMC) Administrator executes the request on the local VDW instance and whether to release the data back to the requestor.

A Data Mart consists of: (a) a database of the data partner's clinical information that has been transformed and loaded into the VDW (see Section IV.B), and (b) a user interface, the DMC program, used by the data partner's DMC Administrator for gatekeeping decisions.

B. Virtual Data Warehouse
Each participating CHORDS data partner institution will maintain a Data Mart that can be queried according to governance specifications (see Appendix B). The structure and content of each Data Mart is based on the VDW.

The full scope of the VDW is explained in Appendix C. At the core of the VDW are standardized tables and definitions. A data dictionary specifies the common format for each table and metadata about its data elements, e.g., variable name, variable label, code values. Data partner programmers have mapped and transformed data elements from their local EHR systems to the common VDW standards.

All tables in a data partner's Data Mart are linked by a locally-generated, anonymized patient identifier. No direct identifiers of individuals can be returned in a request. The crosswalk between patient and randomly generated identifier **is never shared**.

Governance principles regarding CHORDS common data model include:

- The Board, with support from the Data Work Group, will identify and prioritize changes to the VDW.
- Data partners will implement all tables and attributes identified in the data dictionary and populate with data as many as is feasible.
- Data partners will implement VDW changes to the best of their ability, balancing time and budget constraints.
- Data partners are responsible for maintaining and updating data extracts with data refreshes at least quarterly.
- Data partners, with support from the Data Work Group, are responsible for data quality assessment and resolution of data quality issues.

C. Data Requests and Data Exchange
CHORDS data requests and data exchange are managed by the PMN™ application, a software which is used by multiple national networks such as the FDA Mini-Sentinel Initiative and the PCORI-funded Patient Centered Outcomes for Research Network (PCORnet). PMN™ provides security, authentication and auditing required to ensure that only approved data requests are made. All CHORDS data partners will install the PMN™ client for responding to requests.

The CHORDS network consists of a single query portal for making federated queries that is connected to multiple Data Marts through an instance of PopMedNet client installed at each data partner for responding to federated queries. The query portal, hosted at the University of Colorado Anschutz Medical Campus, includes an authorization and authentication (unique logon and password for users) infrastructure and is maintained by the Network Operations Work Group. Authorized data users may access the query portal to pose queries to the network and retrieve results.

Governance principles for all CHORDS data queries and data exchange include:
- All data requests must identify the requesting organization, responsible individual and contact information and intended use(s).
- Data partners will identify a technical contact (if separate from the DMC Administrator) who is responsible for ensuring timely approval and execution of requests.
- Approving, executing and releasing data should be completed within two weeks of receiving a request. Data users may contact data partners to request an expedited process, if necessary.
- Data partners may decline to respond to requests.
- Data partners will review requests prior to execution.
- Data partners will review results prior to release.
- Data queries and exchange will adhere to the CHORDS policy of not publically releasing aggregated cell sizes less than or equal to 10.

Some principles governing data requests and data exchange vary based on the data user.

1. Public Health

- Data requests and data exchange will never include identifiable patient or site/data partner identifiers.
- Data requests will be sent to all data partners, who may choose to run them.
- Data users should limit the frequency of their requests to minimize the burden placed on CHORDS data partners. The Network Operations Work Group will establish a suggested maximum frequency.
- Aggregate results must be stored in a secure manner and may be retained as long as needed.
- When sharing results, query metadata (e.g. date of data request, filter settings, number of responses received, number of sites declining request, etc.) will not be separated from query results.
- Data users will adhere to recommendations about insufficient data. While included in the aggregate results (if >10), cells not meeting sufficient data criteria should not be used externally.
- Designated data users at LPHAs are responsible for ensuring that other LPHA staff use CHORDS results responsibly and accurately.

2. Researchers
- Data requests and data exchange may include protected health information (PHI). Any use of PHI for research must be approved by COMIRB or other IRB and agreed to by all participating sites by these sites entering into the appropriate data sharing agreement. Data requests and data exchange will never include site/data partner identifiers unless explicitly requested by the researcher, agreed to by data partners and stipulated in the data sharing agreement.
- Data requests will be sent only to data partners who have agreed to participate in a given study.
- All CHORDS results must be stored in a secure manner. Data storage and data destruction plans must be provided by the researcher and receive IRB and CHORDS approval.
- When sharing results, query metadata will not be separated from query results.
- For studies using aggregated data, researchers will adhere to recommendations about insufficient data. While included in the aggregate results, cells not meeting sufficient data criteria should not be used externally.
- Researchers are responsible for ensuring that all dissemination of research using CHORDS data is done responsibly, accurately and with appropriate attribution and acknowledgement.
- For research studies, researchers must agree that they will limit their use of the data to the purpose(s) stated in the research plan approved by the Governance Board. Researchers also agree to not re-distribute data to individuals not listed as study personnel in their IRB applications. See the Governance Principles section above for additional details on consequences of data misuse. Any amendments to the use of the data will be reviewed by the Governance Board and must be approved prior to implementation.

D. Ensuring Data Consistency and Quality

CHORDS relies on consistent, quality data to produce valid, reliable information. Data partners agree to ensure the quality of VDW data and the accuracy of the ETL program that extracts data from source systems, transforms data to conform to the VDW specifications, and loads data into the partner's VDW data mart.

Governance principles regarding data consistency and quality include:
- Data quality assessment and improvement is a shared responsibility. The Data Partner has the lead role, supported by the Data Work Group. The data partner will take primary responsibility for executing data quality assessment programs and for investigating data quality issues that are detected at the site and network levels. The Data Work Group may develop and share data quality tools within the PMN™ client.
- Each data partner is responsible for refreshing data on a quarterly basis and for staying aligned with CHORDS updates.
- Each data partner is responsible for completing site specific QA checks after each data refresh or change to the VDW.
- When a significant data quality issue is discovered, the data partner will take the lead in informing the CHORDS Network Operations Work Group, identifying the source and developing solutions.

E. Data Access
Access to data exchange software through CHORDS is limited to specific roles in PMN™. These roles are described in Appendix D.

Requesting data access for public health purposes involves the following steps:
- The public health director, or designee, works with the Board to execute data agreements with data partners as needed.
- Upon executing data use agreements, the public health director, or designee, identifies a staff member to serve as the agency's authorized CHORDS user. The director, or designee, includes the data users' contact information in a signed CHORDS User Access Form (see Appendix E), which is sent to the chairs of the Network Operations Work Group for verification and to create a PMN™ account for the public health user.
- The public health user works with the Network Operations Work Group to set up a PMN™ tutorial and to review CHORDS policies and guidelines.
- Public health users may submit queries after data agreements and User Access Forms are signed and necessary training is received. Users may only use CHORDS data for purposes that align with public health purposes outlined in the data agreements.

Requesting data access for research purposes involves the following steps:
- The research Principal Investigator (PI) contacts the Research Project Manager (RPM) or another member of the Project Development Work Group (PDWG) with a proposed research question. The research PI completes the CHORDS Research Request information form.
- The RPM and the PDWG will perform an initial assessment of the request based on fit, feasibility, and funding opportunities.
- If the PDWG supports the project moving forward, the project will be brought to the Governance Board for review and final approval. Governing Board approval is required for all - requests before any prep-to-research (PTR) or other data requests are initiated.
- After Governing Board approval, PTR data requests may be completed.
- The PI will apply for IRB approval/non-human subject research designation.

- The PI will work with the Governing Board and data partners to execute the necessary data agreements.
- The PI will complete the required CHORDS access forms and tutorials (see above under public health).
- PIs initiate queries when data agreements are signed, after receiving necessary training, and in partnership with a representative of the PDWG or CHORDS RPM.
- Researchers must limit their use of the data to the purpose(s) stated in the research plan approved by the Governance Board.

CHORDS strives to identify viable, feasible projects before submitting any data requests to our data partners. Several stages of screening may be required in selecting appropriate and feasible research projects.

Note on Preparatory to Research (PTR) data requests:

Preparatory to research (PTR) queries of CHORDS do not require IRB review, although individual institutions may require other forms of pre-review. HIPAA requirements that govern the access to PHI for PTR activities are operationalized at the institutional level. As with any data request, data partners may choose to opt out of any PTR request.

In the case of a CHORDS PTR query where simple counts are returned to the requestor, no PHI is shared. The Board views it as consistent with HIPAA to not require any specific action for PTR queries. Researchers interested in executing a PTR query may contact the PDWG.

If a data partner's local institution requires a different process for PTR queries, the researcher must work with that data partner and the CHORDS RPM in order to ensure that the site's needs are met before completing a PTR request. Please note that the Data Mart Administrator has the opportunity to review every request prior to executing it and sharing results.

Research requests must be vetted by the RPM, the PDWG, and approved by the Governing Board before any PTR data requests are completed.

F. Data Use Limitations
Data partners may use their own source data stored in the VDW for other purposes, including research, as long as they comply with applicable state and federal laws and regulations, including HIPAA and the Common Rule, and undergo local review processes.

Data users may only use data obtained from CHORDS for purposes identified in the data agreements. Data may not be reused, disclosed, altered, or sold for any purposes other than those defined in the agreements.

V. Security, Privacy and Confidentiality
All CHORDS network members are charged with responsible stewardship of patient data by maintaining, and strengthening when possible, the privacy and confidentiality of patient data. The CHORDS network believes that the protection of privacy, confidentiality, and data security is essential to the existence and success of public health monitoring and research.

The HIPAA Privacy and Security Rules establish minimum federal standards for protecting privacy and maintaining confidentiality of Protected Health Information (PHI). In the context of CHORDS these responsibilities begin with the data partner organizations and extend, through formalized relationships based on federal law, between the data users, Executive Committee, Advisory Councils and Work Groups.

The Investigational Review Board (IRB) of record for CHORDS is the Colorado Multiple Investigational Review Board (COMIRB) and any additional IRBs as appropriate for the data partner organizations. COMIRB has reviewed the CHORDS infrastructure and declared its public health uses not human subject research.

Each research study using CHORDS must apply for its own IRB approval (e.g., describing study's procedures, data handling, and intended use of the data) which may reference CHORDS' prior review by COMIRB.

If CHORDS data are released, shared, and/or accessed in a way that is inconsistent with processes approved by the IRB of record, collaborating IRBs and executed data use agreements, the procedures in Appendix F, Data Incident Response Plan, will be followed. These procedures include timely and transparent communication regarding the disclosure with the data partners, appropriate Advisory Councils and the Board. CHORDS staff will also communicate any data incidents to the appropriate HIPAA officials at the University of Colorado Denver's Anschutz Medical Campus and collaborating organizations in accordance with University policy. CHORDS staff and the Board will cooperate with and assist the University's HIPAA Compliance officers in their Incident Response and Reporting Process.

A.   Security

Privacy and confidentiality of electronic data depend on control over access. Secure access consists of deterring unwanted access and authorizing only specific types of access to authenticated individuals.

Security functionality depends on PopMedNet™, which securely protects data for several national and federally funded efforts. PopMedNet™ design provides secure, compliant, auditable data transfer.

Public health users must complete the CHORDS onboarding process. This includes identifying users. (In the early stages of CHORDS expansion, each public health agency will have one approved user. The number of approved data users per site may change in the future.)

Researchers must also complete the CHORDS onboarding process. This includes identifying all study personnel who will work with CHORDS data and specifying those that will log into CHORDS to request and access datasets.

All users must complete a User Access Form before receiving account approval. Users must also complete CHORDS/PopMedNet™ training.

B.   Privacy & Confidentiality

The Privacy Rule permits assigning a code or other identification to a patient's health information that protects the person's identity. Data partners and data users may not use or disclose the code and may not disclose its method of identifying the information.

The Security Rule addresses the technical and non-technical safeguards that organizations must have to protect the privacy of individuals' PHI.

CHORDS takes a number of steps to limit risks to privacy.

- The only data to be released based on a public health query request must meet small-cell guidelines (cells containing 10 or fewer individuals are masked).
- Data are shared in aggregate or as a limited dataset to remove identifying information such as patient names, addresses and phone numbers.
- For research requiring access to patient-level data, only limited dataset information is available and will only be shared when the required IRB approval(s) and data sharing agreement(s) are in place. Attempts to re-identify patients are not permitted, and will be treated as incidents requiring investigation with the Data Incident Response Plan previously described.
- Access to CHORDS is permitted only to users who have permission from the Board to access the data network and have signed data agreements. Each time a user accesses the network, he/she must identify themselves through user authentication (providing a username and password), and a record of access is kept.
- CHORDS has written documentation of its HIPAA compliance policies and procedures to ensure proper use and storage of data by all users.
- Data are never attributed to a site nor are data from one partner site compared with data from another unless sites have provided written consent and approval.

VI. Publication and Presentations Guidelines
Each CHORDS data partner has complete control over the use and confidentiality of its data. By providing data for public health uses, data partners are agreeing for their data to be used publicly, in accordance with data agreements and CHORDS principles and guidelines. The Project Development Work Group will establish a process for public health users to notify data partners when they are using CHORDS data with external stakeholders. *Section to be updated when process is in place.*

A presentation or publication (whether in popular press or an academic journal) including CHORDS data from multiple CHORDS data partners will generally have at least one co-author or acknowledged contributor from each site. Co-authorship implies that the data partner is responsible for the quality and integrity of its data and working with users on its interpretation and meets the requirements of co-authorship per ICMJE guidelines: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. Contributors review and provide feedback on manuscript drafts, but may be less involved in the planning, execution and analysis of the research project and/or data.

Review and Comment by Co-Authors and Contributors:
An author must submit a draft manuscript, abstract or presentation (online or in-person) based on CHORDS data to each data partner supplying data for the study (or a designated committee) for review and comment. Each data partner (or the designated body) will review the manuscript within 30 days or less. CHORDS research or public health project managers must also review these drafts before submission or publication. Authors are responsible for allowing sufficient and reasonable time for review by all parties.

Blinding Source of Data in Publications and External Communications:

Joint Studies: A publication or other external communication resulting from a joint study may not include identifying information of participating data partners without the permission of that site. Such publications or other external communications do not need to be presented in aggregated form, but a data partner may object to the presentation of the non-aggregated information if the data partner believes the information could be used to identify the site.

Single-Member Studies: Except as related to joint studies as specified above, a publication or other external communication related to use of data may not include the name, location and other geographic data, ranking, or confidential information of a site, without consent of that site.

Acknowledgments: Each data partner and data user agrees to acknowledge CHORDS and use of CHORDS data in any work based in whole or part on any data received through CHORDS, and to acknowledge the organizations that provided funding for CHORDS. An author takes full and final responsibility for his or her analysis of CHORDS data and the presentation of any analyses or conclusions in any publication or presentation.

VII. Conflicts of Interest

Project data will only be accessible to CHORDS investigators and approved nonaffiliated investigators whose home institutions maintain and enforce Conflict of Interest (COI) policies for staff investigators. These policies must address employees and their immediate family members. For research activities, investigators must also comply with COMIRB's COI policies.

CHORDS relies on the investigator's home institution to maintain an appropriate written, enforced policy on COI that complies with Federal Regulation 42 CFR 50 Subpart F: Responsibility of Applicants for Promoting Objectivity in Research for Which PHS Funding Is Sought. Participating institutions are expected to have COI policies that meet these minimum standards. Further, it is expected that CHORDS investigators and approved nonaffiliated investigators who have access to project data, abide by the policies of their home institution. These must include, at a minimum:
- Processes to determine COI;
- Requirements to disclose financial interests (including those of immediate family members) that might pose COI or perceived COI;
- Requirements to disclose COI that might affect the research process or study participants, including situations in which the investigator may have a real or perceived undue influence over the research process;
- Remedies to manage, reduce or eliminate the COI or the appearance of COI
- Enforcement mechanisms that impose sanctions when appropriate.

VIII. Scientific Misconduct

The CHORDS Governing Board must be informed of any scientific misconduct by the site PI from which the misconduct originated. These instances are expected to be addressed by the investigator's home institution and issues and resolution communicated to the Board and data partners. The Board may decide to take action based on the reported scientific misconduct activity.

IX. Glossary
- **Adapter –** a group of pre-defined requests on a specific topic. For example, CHORDS has a Mental Health adapter that contains diagnosis-specific requests (e.g., depression, schizophrenia or anxiety).

- **CHORDS Portal** – the web application component of the PopMedNet system. There is one CHORDS Portal, hosted by the University of Colorado Anschutz Medical Campus. Investigators log in to the CHORDS Portal to submit their data queries. The Data Mart Client at each site accesses these requests, and returns data to the Portal. The Portal aggregates data across sites; investigators log in to the Portal to access their result files.

- **Data Governance** – strategies that define the structure, format and purpose for collecting data.

- **Data Mart Client (DMC) –** the software application component of PopMedNet that is installed by each data partner. The DMC connects the site's VDW to the CHORDS network. The DMC polls the central CHORDS Portal for any waiting requests, and is used to execute those requests against the site's Virtual Data Warehouse. The DMC is then used to return results to the central CHORDS Portal for aggregation and sharing with the data user.

- **Data Partner** – an organization that establishes a connection between its virtual data warehouse and the CHORDS network using PopMedNet.

- **Data Standard*** – a predetermined set of structural and semantic data requirements for each VDW.

- **Data Stewardship** – using data accessed through CHORDS according to the established governance plan.

- **Data Use Agreement** – a contract concerning the transfer of non-public data. A data use agreement (DUA) outlines the terms and conditions of the data transfer, including limitations on its use and required procedures for keeping data secure.

- **De-identified Data Set**** – data that have been stripped of all direct patient identifiers. De-identified data sets are not considered protected health information. According to HIPAA, covered entities may use or disclose health information that is de-identified without restriction under the Privacy Rule. Covered entities seeking to release this health information must determine that the information has been de-identified using either statistical verification of de-identification or by removing certain pieces of information from each record as specified in the Rule. The Privacy Rule allows a covered entity to de-identify data by removing all 18 elements (see below) that could be used to identify the individual or the individual's relatives, employers or household members. The covered entity also must have no actual knowledge that the remaining information could be used alone or in combination with other information to identify the individual who is the subject of the information.

  Under this method, the identifiers that must be removed are the following:
  1. Names.
  2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three

digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census:
   a. The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people.
   b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers.
5. Facsimile numbers.
6. Electronic mail addresses.
7. Social security numbers.
8. Medical record numbers.
9. Health plan beneficiary numbers.
10. Account numbers.
11. Certificate/license numbers.
12. Vehicle identifiers and serial numbers, including license plate numbers.
13. Device identifiers and serial numbers.
14. Web universal resource locators (URLs).
15. Internet protocol (IP) address numbers.
16. Biometric identifiers, including fingerprints and voiceprints.
17. Full-face photographic images and any comparable images.
18. Any other unique identifying number, characteristic or code, unless otherwise permitted by the Privacy Rule for reidentification.

- **Limited Data Set**** – protected health information that excludes certain direct patient identifiers. According to the HIPAA Privacy Rule, a limited data set (LDS) may be used for research, public health, or health care operations when the data set recipient enters into a Data Use Agreement (DUA) with the site (data owner) providing the data set.

    An LDS can include dates, limited geographic information and a link field (e.g., an encrypted identifier), such as:
    - Dates (e.g., admission, discharge and service dates; dates of birth and death) and ages of research participants;
    - Certain general geographic information, including five or nine-digit zip codes and state, county, city and precinct; and
    - Links which may be used to identify individuals when the researcher maintains and holds confidential the key required for reidentification.

    An LDS must exclude all other protected health information identifiers, such as:
    - Names and street or postal addresses;
    - Telephone and fax numbers;

- E-mail and Internet Protocol (IP) addresses and web Universal Resource Locators (URL);
- Social Security, medical record, health plan beneficiary and other account numbers;
- Certificate and license numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Device identifiers and serial numbers;
- Biometric identifiers, including finger and voice prints; and
- Full-face photos and any other comparable images.

- **Process Standard*** – format, language and content of queries, data models and processes that affect CHORDS operations.

- **Protected Health Information** (PHI) - individually identifiable health information.

- **Virtual Data Warehouse (VDW)** – a database containing data extracted directly from a local EHR that is reconfigured using standard variable names and values. Each CHORDS data contributor establishes its own virtual data warehouse (VDW), allowing CHORDS to produce comparable data that can be easily merged across sites in order to conduct public health monitoring and other research.

*Definitions for these terms are drawn from Holmes JH, Elliott TE, Brown JS, et al. *J Am Med Inform Assoc.* 2014; 21:730–736.

**Definitions for these terms are from the Governance of the Patient Outcomes Research to Advance Learning (PORTAL) Network, Version 3 November 2015. Authors: Steiner JF, Nelson AF, Paolino AR, McGlynn E.

XII. Appendix

Appendix A: CHORDS Data Use Agreement Template

DATA USE AGREEMENT

This Data Use Agreement ("Agreement") is entered into by and between _____
(hereinafter, "Covered Entity") and **Boulder County Public Health** located at 3450 Broadway, Boulder
CO 80304; **Broomfield Health and Human Services** located at 100 Spader Way, Broomfield CO
80020; **Colorado Department of Public Health and Environment** located at 4300 Cherry Creek S Dr,
Denver 80246; **Denver Health and Hospital Authority** located at 660 Bannock Street, MC 1919,
Denver CO 80204; **Jefferson County Public Health** located at 645 Parfet St, Lakewood CO 80215; and
Tri-County Health Department located at 6162 S. Willow Dr Ste 100, Greenwood Village CO 80111
(hereinafter, "Data Recipient(s)"). Covered Entity and Data Recipients may herein be individually
referred to as "Party "or collectively as "Parties".

RECITALS:

WHEREAS, the Parties are committed to improving the health of populations and the communities
they serve; and

WHEREAS, any sharing of data is consistent with the definitions set forth in the Health Insurance
Portability and Accountability Act §164.512:

> *"Uses and disclosures for which an authorization or opportunity to agree or object is not required.*
>
> *(i) A public health authority that is authorized by law to collect or receive such information for
> the purpose of preventing or controlling disease, injury, or disability, including, but not limited
> to, the reporting of disease, injury, vital events such as birth or death, and **the conduct of
> public health surveillance**, public health investigations, and public health interventions; or,
> at the direction of a public health authority, to an official of a foreign government agency that
> is acting in collaboration with a public health authority".* And,

WHEREAS, Covered Entity is providing certain Protected Health Information ("PHI") to Data Recipients
in the form of a Limited Data Set for the purpose(s) identified in **Schedule 1**; and

WHEREAS, in connection with the provision of that PHI, pursuant to the Health Insurance Portability
and Accountability Act and regulations promulgated pursuant thereto (collectively "HIPAA"), Covered
Entity is required to obtain assurances from Data Recipients that Recipients will only use or disclose
PHI as permitted herein;

WHEREAS, the Parties enter into this Agreement as a condition to Covered Entity's furnishing the Limited Data Set to Recipients, and as a means of Recipients' providing assurances about use and disclosure; the provisions of this Agreement are intended to meet the Data Use Agreement requirements of HIPAA.

NOW, THEREFORE, in consideration of the mutual promises and considerations set forth below, the Parties agree as follows:

1. Definitions. Any capitalized terms used in this Agreement and not otherwise defined, shall have the meanings set forth in 45 CFR Parts 160 - 164 issued under the Health Insurance Portability and Accountability Act of 1996, P.L. 104-91 ("HIPAA"), as amended

2. Term. This Agreement shall commence on the Effective Date and continue until terminated in accordance with Section 4 below.

3. Data Recipient's Obligations. Data Recipients shall:

   a. With respect to the PHI, comply with all applicable federal and state laws and regulations relating to the maintenance of the PHI, the safeguarding of the confidentiality of the PHI, and the Use and Disclosure of the PHI.

   b. Use and disclose the PHI as is minimally necessary only for the purpose(s) identified in Schedule 1, or as otherwise required by law. Schedule 1 may be modified by the parties at any time pursuant to a writing executed by both parties. No uses or disclosures different from that permitted by the currently in-force Schedule 1 may be made until the new Schedule 1 has been signed by all parties.

   c. Use appropriate safeguards as required by law to prevent any use and disclosure of the PHI, other than for as provided for by this Agreement. Upon request by Covered Entity, Recipient(s) shall describe the safeguards being used to prevent unauthorized use or disclosure of the PHI.

   d. Immediately report to the Covered Entity any use or disclosure of the PHI other than as expressly allowed by this Agreement.

   e. Ensure that its employees and representatives comply with the terms and conditions of this Agreement, and ensure that its agents, Business Associates and subcontractors to whom

Recipient(s) provides the PHI agree to comply with the same restrictions and conditions that apply to Recipient(s) hereunder.

 f. Not identify or attempt to identify the information contained in the Limited Data Set, nor contact any of the individuals whose information is contained in the Limited Data Set.

 g. Not request the Covered Entity to use or disclose more PHI than the minimum amount necessary to allow Recipient(s) to perform functions pursuant to the purpose identified in Schedule 1.

4. Termination.

 a. Covered Entity may terminate this Agreement and any disclosures of PHI pursuant hereto, upon ten (10) days' notice to Recipient(s) if Recipient(s) violates or breaches any material term or condition of this agreement. Covered Entity may terminate this Agreement without cause upon 30 days' written notice.

 b. Upon termination, Recipient(s) shall promptly return or destroy the Limited Data Set received from Covered Entity in connection with the purposes identified on Schedule 1. If return or destruction of the Limited Data Set is not feasible, Recipient(s) shall continue the protections required under this Agreement consistent with the requirements of applicable HIPAA privacy standards. If Recipient(s) cease to do business or otherwise terminates its relationship with Covered Entity, Recipient(s) agree to promptly return or destroy all information contained in the Limited Data Set received from Covered Entity in a timely manner.

5. This Agreement in no way limits any obligations of Recipient(s) or its workforce members under any law or policy that is otherwise applicable to the Recipient, such as an obligation to comply with their institution's policies and procedures.

6. Any ambiguity in this Agreement relating to the use and disclosure of the Limited Data Set by Recipient(s) shall be resolved in favor of a meaning that further protects the privacy and security of the information.

IN WITNESS WHEREOF, the parties have executed this Agreement as follows:

COVERED ENTITY
Name

By_____
Name
Title

DATA RECIPIENT
Boulder County Public Health

By_____
Jeff Zayach
Executive Director

DATA RECIPIENT
Broomfield Health and Human Services

By_____
Jason Vahling
Public Health Director

DATA RECIPIENT
Colorado Department of Public Health and Environment

By_____
Name
Title

DATA RECIPIENT
Denver Health and Hospital Authority

By_____
Bill Burman, MD
Chief Executive Officer

DATA RECIPIENT
Jefferson County Public Health

By_____
Mark B. Johnson, MD
Executive Director

DATA RECIPIENT
Tri-County Health Department

By_____
Alyson Shupe
Informatics, Epidemiology and Health Planning Manager

Schedule 1

1. Effective Date:

2. Name of Covered Entity contact person/department releasing the Limited Data Set (LDS):

3. Name of Recipients of the LDS:
   - Boulder County Public Health, Strategic Initiatives Branch
   - Broomfield Health and Human Services
   - Colorado Department of Public Health and Environment, Center for Health and Environmental Data.
   - Denver Health and Hospital Authority's Denver Public Health
   - Jefferson County Public Health
   - Tri-County Health Department

4. Purpose of LDS disclosure: Public health surveillance activities for the following conditions:
   - Infectious or communicable diseases.
   - Chronic diseases, conditions or health behaviors.
   - Behavioral health conditions.

5. The recipient of the LDS listed in #2 are permitted to use and disclose the LDS for the following specific purpose(s):

   Covered Entity's LDS data will be combined with data from other regional collaborators to create a Consolidated Data Set (CDS). The CDS will be used to monitor population health, target areas for intervention and conduct evaluation activities. Analyses conducted on the CDS data will be shared on a regional basis with the goal of improving health and the quality and availability of health care and health resources in neighborhoods or sub-county regions. Reports developed from limited and consolidated data sets may be disclosed at the county, council district, neighborhood level or census tract in accordance with the purpose described above.

6. Description of Limited Data Set disclosed for purposes of public health surveillance.

   Elements of the LDS will include public health surveillance data related to the conditions described above in Section 4 for the year 2000 and forward. Data elements (based on the Virtual Data Warehouse [VDW] are:

   - Enrollment data: health plan membership enrollment with indicators of insurance types, benefits, and effective dates of coverage.

- Demographic information including: date of birth, race, ethnicity, gender
- Census information (to be calculated at Covered Entity before LDS is transferred) including: geocode, block, tract, state, zip code, and county.
- Vital sign information including: measure date, height, weight, blood pressure
- Social history information including: contact dates, tobacco use and second hand smoke exposure status (includes length of time used and types of use, i.e., cigarettes, cigars, chew, etc.)
- Encounter characteristics including: date, type, random unique ID
- Pharmacy dispensing and claims including: date of dispensing, National Drug Code or GPI code (to standardize across sites), therapeutic class, days supplied, and amount dispensed
- Diagnosis information including: dates, code, and type
- Laboratory information including: procedure codes, collection dates, values, and measurement units, including LOINC codes
- Death - Contains death date and cause of death information
- Tumor – Incident cancers.

Covered Entity will generate, assign and store a unique identifier for patients, providers and encounters that will be used to join health information across tables.

Patient Protected Health Information that must be <u>excluded</u> from the LDS is as follows:
   a. Names
   b. Postal address information, other than town or city, state, and zip code
   c. Telephone numbers
   d. Fax numbers
   e. Electronic mail addresses
   f. Social Security numbers
   g. Medical record numbers
   h. Health plan beneficiary numbers
   i. Account numbers
   j. Certificate/license numbers
   k. Vehicle identifiers and serial numbers, including license plate numbers
   l. Device identifiers and serial numbers
   m. Web Universal Resource Locators (URLs)
   n. Internet Protocol (IP) address numbers
   o. Biometric identifiers, including finger and voice prints, and
   p. Full face photographic images and any comparable images.

Appendix B: CHORDS Data Exchange Requirements

Objective: Identify existing technical capabilities and prioritize new technical capabilities for the client and the portal.

Note: In this table, "data user" refers to either an Investigator or an Enhanced Investigator. Investigator and Enhanced Investigator are two different roles that can be assigned within PopMedNet (PMN). In these scenarios, it may be helpful to think of Investigators as LPHA data users (who must submit queries to all participating sites), and to think of Enhanced Investigators as researcher data users (who can select sites to submit queries to). A DataMart Client (DMC) Administrator is a data partner who reviews and responds to queries from data users and manages their own DataMart. The Network Administrator manages the network, including creating network entities, managing access controls and approving or creating users based on CHORDS Governance policies and guidelines.

	Scenario	Can it be programmed?	Does it exist? (Current Setting)	Priority level (1, 2, 3)	Final Recommendation or Discussion Requested
1.	As a data user, I want to query the prevalence of the following health issues:			1	
2.	• BMI	Yes	Yes (available)	1a	First priority Assess need for changing min. cell size for stable BMI estimates once we begin running queries
3.	• Mental health	Yes	No	1b	First priority – starting with depression prevalence
4.	• Cardiovascular disease	Yes	No	1c	Second priority – we have some previous work done on CVD
5.	• Diabetes mellitus	Yes	No	1c	Third priority
6.	As a data user, I want to customize my query using parameters for custom time range, demographic and geographic characteristics.	Yes – can stipulate time range (down to date), age, race, gender, ethnicity, county level geographic.	Yes (available)	1	Decision: Move parameters over from PMN v4. More discussion needed around age parameters.
7.	As a data user, I want to customize my query using parameters for encounter type (for example, outpatient vs. inpatient encounters).	Yes	No	1	Recommend
8.	As a data user, I want to receive a page with instructions/metadata/query information and background with my results.	Yes	No	1	Recommend – provide requirements to Bryant

9.	As a data user, I want a text warning in my results when cell sizes contain small numbers of patients (text reading "≤10" is returned in cell) in data aggregated from all sites.	Yes	No	1	Decision: OK to follow CMS guidelines
10.	As a DMC Admin, I want a text warning in my results when cell sizes contain small numbers of patients (text reading "≤10" is returned in cell) in data from my site	Yes	No (current: yellow highlight. Controlled at adapter level, can be set to anything we decide.)	1	Recommend
11.	As a data user, I want a text warning before opening results or in the header of results document when cell sizes contain small numbers of patients (text reading "≤10" is returned in cell) in data aggregated from all sites.	Yes	No	1	Recommend
12.	As a DMC Admin, I want a text warning before opening results or in the header of results document when cell sizes contain small numbers of patients (text reading "≤10" is returned in cell) in data from my site.	Yes	No (current: yellow highlight of cells. Controlled at adapter level, can be set to anything we decide.)	1	Recommend
13.	As a PMN admin, I want the ability to turn on and off small cell size suppression as needed for pilot testing purposes	Yes	Yes	1	
14.	As an Investigator, I want to select certain sites to submit my query to.	Yes	Yes	1	Recommend not allowing.
15.	As an Enhanced Investigator, I want to select certain sites to submit my query to.	Yes	Yes	1	Recommend
16.	As a data user, I want text reading "0" to be returned in a cell when a query results in a cell value of 0.	Yes	No	1	Recommend

17.	As a DMC Admin, I want text reading "0" to be returned in a cell when a query results in a cell value of 0.	Yes	Yes (on – needs to be programmed at adapter level.)	1	Recommend
18.	As a DMC Admin, I want to receive a message when a query I have run against my VDW fails at my site.	Yes	Yes (on – message depends on error type and includes dialog box for VDW connect failure and scheme issues, empty result set only for error in join in database code.)	1	Recommend Decided on need for training re: chain of communication involving query failure. (DMC Admin to Portal Admin to Investigators)
19.	As a data user, I want to receive a message when a query cannot be completed at one or more sites / fails to return data (text in results indicates "blank/failed query" and is easily distinguished from a successful query returning 0 patients).	Yes	Yes	1	Question for group: Where do users want this information? In results, or on PMN dashboard?
20.	As an Investigator, my query will fail if only one site returns data.	Yes	Yes (on)	1	Recommend
21.	As a data user, I want to change my query routing to different sites (send my query to different sites than I originally identified) after initial submission.	Yes	Yes (on)	1	Recommend be turned off.
22.	As a data user, I want to enter one or more ICD-9/10 codes and query prevalence for those codes.	Yes	No	2	Recommend- Starting point is to build the knowledge storage tables, followed by building this query builder tool.

#	User Story				
23.	As a data user, I want to select a value that is attributed to a group of diagnosis codes/procedures/drugs in CHORDS' "knowledge storage" tables.	Yes	No	2	Recommend
24.	As a data user, I want to search for terms/fields that are available in CHORDS' "knowledge storage" tables.	Yes	No	2	Recommend
25.	As an Enhanced Investigator, I want to write a custom research program (submit SQL/SAS/R code)	Yes	No (would be easy. SQL and SAS distribution model exist from LP)	2	Discussion: Recommend but need to establish parameters, agreements around this. Do we want to include an R adapter in this conversation? JW brings up SPSS, Stata as programs to consider also.
26.	As a data user, I want to query for condition prevalence based on criteria other than ICD-9/-10 codes (for example, query for depression prevalence using PHQ9 scores).	Yes	No	2	For future review and discussion
27.	As a DMC Admin/Investigator, I want a warning when my results mix is mostly from one site (ex. 95% of patients are from one site after aggregating data from all sites).	Yes	No	3	Future discussion: This is an issue of both business privacy and of correctly interpreting results. Because there could be many different scenarios and thresholds, we've decided to put this on hold until we run queries and establish how frequently this is an issue. Also need to train sites/DMC Admins on this.
28.	As a DMC Admin, I want thresholds for denying results return due to small cell size	Yes	Yes (On)	x	
29.	As a data user, I want to receive notifications when query is completed	Yes	Yes (on - email and portal)	x	
30.	As a data user, I want to receive notifications when a DMC Administrator places my query on hold	Yes	Yes (on - email and portal)	x	

31.	As a data user, I want to receive notifications when a DMC Administrator denies my query.	Yes	Yes (on - email and portal)	x	Reason for denial must be provided by DMC when denied
32.	As an Enhanced Investigator, I want to receive metadata showing the names of sites that responded to a query.	Yes – visible in Portal	Yes (on)	X	Recommend
33.	As a DMC Admin, I want to run QA queries against my site's own datamart	Yes	Yes (available)	x	
34.	As a Network Administrator, I want queries available to researchers for submission to be automatically filtered based on the user submitting the query	Yes	Yes (on - happens through Project permissions)	x	
35.	As an Enhanced Investigator, I want to view results at any time while query is being completed after submission	Yes	Yes (on)	x	

In addition to the data exchange requirement scenarios above, the following security settings can be configured within application settings:

ConfiguredPasswordExpiryMonths	6
FailedLoginAttemptsBeforeLockingOut	5
KeepResponseDocumentsDays	365
PasswordExpirationNagDaysPrior	7
PasswordExpirationNagPeriodDays	1
PasswordMaxLength	14
PasswordMinLength	8
PasswordRequireMixedCaseCharacter	TRUE
PasswordRequireNumericCharacter	TRUE

PasswordRequireSpecialCharacter	TRUE
UnapprovedRequestReminderPeriodDays	1
UnapprovedResultsReminderNotificationPeriodDays	1
UnapprovedResultsReminderPeriodDays	1
UnexaminedResultsReminderPeriodDays	1
UnrespondedRequestReminderPeriodDays	1

Appendix C: CHORDS VDW V2.0 Data Model Manual

A manual describing the CHORDS data model was prepared for CHORDS data partners and data consumers to provide a detailed description of what data is stored in the CHORDS Virtual Data Warehouse (VDW).

How to Use this Document

Data partners should use this document in conjunction with the CHORDS Data Partner Implementation and Maintenance Manual to build and populate their VDW. Use the table descriptions and column descriptions to know what data should be mapped into each table and column. Use the column descriptions to identify what default values should be for each column and when a column can be null. Data quality checks are provided with each table to help data partners think through the data that has been loaded and whether it accurately represents their data.

VDW Summary

The CHORDS VDW V2.0 Data Model includes 17 tables and is based on leading distributed data models, sharing the most similarities with the Health Care Services Research Network (HCSRN, formally the HMORN) VDW Version 3. Data can be connected across tables by primary and foreign keys. The figure below summarizes the 17 CHORDS VDW V2.0 tables and the table relationships.

CHORDS Virtual Data Warehouse Data Model
Version Issued: January 2014
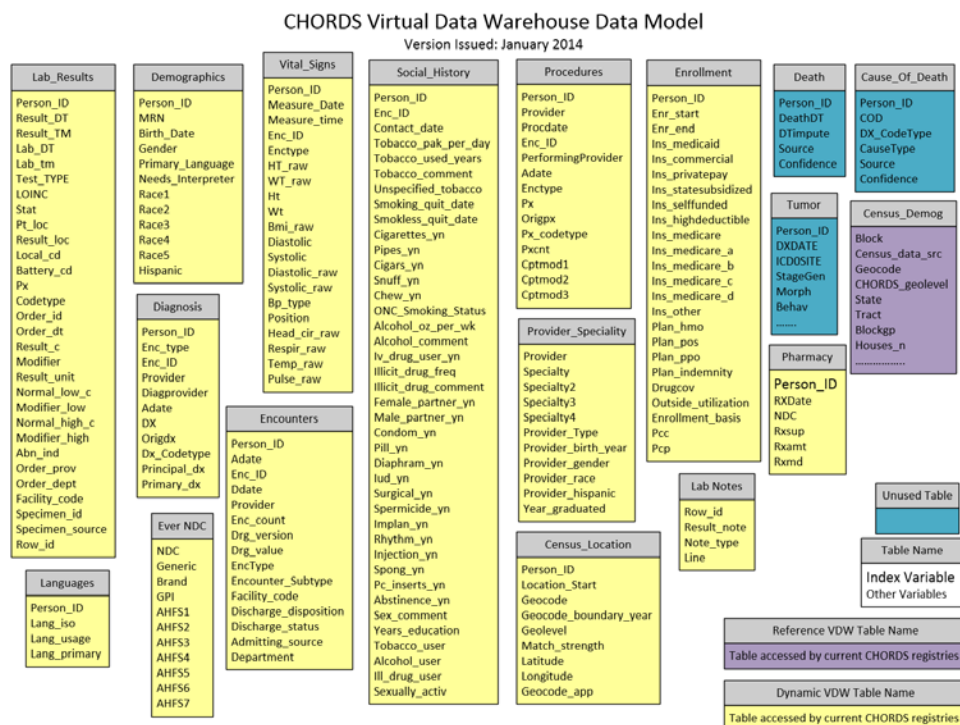
What Data to Load into VDW

The CHORDS VDW V2.0 should include all available patient, encounter, and health service and related information since 1/1/2011. There are no date limits if data partners wish to load historical data from before 1/1/2011; this recommended start date is provided if resources or storage needs are constrained. Please consult your organization's Institutional Review Board (IRB) as you may be limited by your individual data use agreements/IRB protocols on the amount of data you can provide.

Data loaded into the VDW should reflect the data stored in the source system. Data partners should limit their efforts to clean data or infer information that may not be captured electronically. There is no expectation that data partners extract health information captured on paper forms or use text mining tools to extract information stored in unstructured text fields.

VDW Table Definitions

The 17 VDW tables have been prioritized to allow data partners to develop and load data incrementally. Priority 1 tables will be used in the near future and must be populated to respond to queries that will be developed this year to estimate disease prevalence. Priority 2 tables will be used in the following year for more detailed queries such as the lab results table used to estimate diabetes control. There is not an immediate need for priority 3 and 4 tables, but CHORDS may revisit the decision to build them in future years.

Priority 1 Tables	Brief Table Description
Census_Demog	The CENSUS_DEMOG table holds population information from the American Community Survey at the census block level. **This is a static reference table that will be provided to each data partner through the PopMedNet client or can be sent through secure file transport. Data partners do not need to develop this table.**
Census_Location	The CENSUS_LOCATION table holds information about the patient's location based on their address stored as a geocode. To complete this table, and populate the related geo-demographic data, patient addresses must be geocoded. Detailed instructions about this will be discussed in the CENSUS_LOCATION section of the document. **Every PERSON_ID appearing in the demographics table should have a record in this table.**
Demographics	The DEMOGRAPHICS table contains basic person-descriptive data for the people found in VDW tables. It should also serve as a lookup dataset for person ids. **Every PERSON_ID appearing in any other VDW file should appear in DEMOGRAPHICS (even if nothing substantive is known about the person)**.
Diagnoses	The DIAGNOSES table lists all final diagnoses associated with the encounters in the ENCOUNTER table. Diagnoses from the problem list and chief complaint should not be included. The index variable ENC_ID uniquely identifies each encounter and is used to link diagnoses to other VDW tables.
Encounters	The ENCOUNTERS table should include records from encounters between patients and medical personnel indexed by ENC_ID and PERSON_ID. All available encounters with a medical provider should be included. Medical providers include: physicians, nurse practitioners, registered nurses, lab technicians, social workers, etc.—generally, people licensed to provide medical care and closely related services. Recurring visits to the same clinicians on the same day should be maintained as separate encounters if possible. Multiple encounters to the same provider on the same day are allowed if that is

	the truth in the source data and have unique ENC_ID values. Inpatient hospital stays lasting multiple days should be represented as one encounter. FACILITY_CODE is included in the ENCOUNTERS table for data partners to populate with any data partner-specific clinic or department identifier that can aid data validation and preserve data lineage.
Vital Signs	The VITAL_SIGNS table includes various physiological measures taken by health professionals during an encounter including body temperature, pulse rate, blood pressure, respirations, and anthropometry (height, weight, and BMI). There is no unique primary key for the VITAL_SIGNS table. VITAL_SIGNS does include an ENC_ID but multiple measures may be captured for any given encounter.
Priority 2 Table	
Ever NDC	The EVERNDC table is a dynamic reference table that provides additional information about fill records that are included in the pharmacy table.
Lab Results	Lab result information is stored in two tables, LAB_RESULTS and LAB_NOTES. The LAB_RESULTS table holds structured result information. The main table, LAB_RESULTS, has one record per lab result. If a test has no result for whatever reason (specimen not sufficient, patient did not show) then the test is not included in LAB_RESULTS.
Languages	The LANGUAGES table contains information on the languages that patients speak and write, where known. There is one record per person per known language.
Pharmacy	The PHARMACY table lists each record and should represent one dispensing transaction per patient per NDC per day. This table should only be populated for data partners with an outpatient pharmacy or records of outpatient fills. The file should be summed by dispensed amount if necessary to remove rollback transactions. The pharmacy table should never include any non-positive amounts or days supply.
Procedures	The PROCEDURES table lists all procedures performed. Each record is unique with regard to the combination of procedure code/original procedure code/performing provider/performing provider/procedure date and ENC_ID. Include denied claims if you consider the utilization to be valid. Includes all performed procedures from all settings including lab, radiology and immunization procedure codes. Lab, radiology, immunization and other procedure codes may exist without a linked record in the ENCOUNTER table. Require that the PROCEDURES table contain each type of procedure code when available: CPT, HCPCS, ICD9, Revenue Code. Exclude procedures that were ordered but not performed. Include procedures from professional and inpatient rounding services (outpatient providers who visit their patients in the hospital) in inpatient encounters. Identify the professional/inpatient rounding provider in the "performingprovider" variable and the procedure date in the "procdate" variable. Ensure that the codetype variable is correct for each procedure value (px).
Provider Specialty	The PROVIDER_SPECIALTY table includes information about providers' demographic characteristics, practice and specialty, and credentials.
Social History	The SOCIAL_HISTORY table includes various health behavior related measures including tobacco use, alcohol use, drug use, and family planning.
Priority 3 Table	
Enrollment	The ENROLLMENT table should contain one record per unique combination of a PERSON_ID, enrollment start date, and basis. The ENROLLMENT table documents periods of time during which all medical events are expected to be captured in the database. This

	concept is often insurance-based but other methods of defining enrollment are possible. The enrollment table can be difficult for a data partner not representing an insurance based population to populate. More details and examples are available in the enrollment table section.
Priority 4 Table	
Cause_of_Death	The CAUSE_OF_DEATH table contains one record per person per known cause of death. The unique primary key is the combination of MRN, COD, Dx_CodeType, and CauseType. All variables on the table are required to be non-null. All MRN in the CAUSE_OF_DEATH table should have a corresponding record in the DEATH table, but not all MRN in the DEATH table will necessarily have a record in the CAUSE_OF_DEATH table. When death sources provide conflicting information, data partner data managers should make local determinations as to which source to use, collating information when possible, and reflecting their confidence in the observation using the Confidence variable.
Death	The DEATH table contains one record per person in the VDW DEMOGRAPHICS table for whom you have some belief that the person may be dead. The unique primary key of the table is PERSON_ID. The PERSON_ID, Confidence, and Sources variables are the only variables required to be non-null for inclusion in the table. Although records with unknown death dates (DeathDt) have fewer uses than those with known dates, they should still be included in the file.
Lab Note	Lab result information is stored in two tables, LAB_RESULTS and LAB_NOTES. LAB_NOTES is a table that can store unstructured text information about lab tests.
Tumor	The TUMOR table should contain one record per unique combination of PERSON_ID and tumor related illness.

Information about VDW Key Variables

The VDW includes several primary and foreign keys to facilitate easy analysis. A primary key is a key in a relational database that is unique for each record. It is a unique identifier, such as a patient identifier or medical record number, provider identifier or encounter identifier. In the context of relational databases, a foreign key is a field (or collection of fields) in one table that uniquely identifies a row of another table. In simpler words, the foreign key is defined in a second table, but it refers to the primary key in the first table. Three keys exist in the current VDW model. Each of the three are primary keys for one table and foreign keys for other tables. Some VDW tables do not have a primary key. The primary key for the demographics table is PERSON_ID and is unique for each patient. The primary key for the encounters table is END_ID and is unique for each encounter. The primary key for the provider table is provider and is unique for each encounter. Because healthcare organizations maintain internal variables like medical record number, encounter identifiers and provider identifiers that would be considered protected health information, it is important to note that data partners should populate VDW key variables with random identifiers that can be shared within the constraints of a limited data set.

Keys

Name	Description	Definition

PERSON_ID	Person ID. 12 digit number, with the data partner code always as the first two characters (See Data partner Identifier below). Following 10 digits are randomized PERSON_ID for which each data partner will maintain a crosswalk.	CHAR(12) NOT NULL
ENC_ID	Arbitrary encounter level identifier used to link encounters across tables (ENCOUNTERS, PROCEDURES, DIAGNOSES). Each encounter should get a unique ENC_ID.	CHAR(32) NOT NULL
PROVIDER	Arbitrary provider identifier used to link providers to provider level information and across tables (ENCOUNTERS, PROCEDURES, DIAGNOSES). Identifier unique to a provider. Provider code for the provider who is most responsible for this encounter. Usually physician, nurse practitioner, physician assistant, optometrist, etc. For encounters with multiple providers and there isn't a clear one in charge, please choose one arbitrarily so the encounter can be linked to the diagnosis and procedure files If there is no provider code for an encounter, then specify the value for provider as "UNK."	CHAR(15) NOT NULL

Data partner Identifier

Each PERSON_ID should reflect the data partner code in the first two characters.

Status	Data partner	Data partner Code	Example PERSON_ID
Current	Kaiser Permanente Colorado	KP	KP1234567889
Current	Denver Health and Hospital Authority	DH	DH1234567889
Current	Children's Hospital Colorado	CH	CH1234567889
Current	High Plains Health Center	HP	HP1234567889
Current	Colorado Alliance for Health Equity and Practice	CA	CA1234567889
Current	Metro Community Providers Network	MC	MC1234567889
Current	Salud Family Health Centers	SA	SA1234567889
Current	Clinica Tepeyac	CT	CT1234567889
Current	Clinica Campesina	CC	CC1234567889
Current	Stout Street Clinic	SS	SS1234567889
Future	*Mental Health Centers of Denver*	*MH*	*MH1234567889*
Future	*Jefferson Center for Mental Health*	*JC*	*JC1234567889*
Future	*University of Colorado Health*	*UC*	*UC1234567889*
Future	*Centura Health*	*CE*	*CE1234567889*

Unstructured Data

Unstructured data, such as free form notes or comments, are not put into the VDW (data not parsed). The only free form data currently incorporated into the VDW currently are Lab Notes, and other areas of this data, including Chief Complaints and PRM messages replies need to be considered.

Missing, NULL or Unknown Data Vales

The CHORDS VDW allows for several flavors of NULL to represent a variety of scenarios. General guidance about how to map missing, null or unknown data values are as follows:

- If a data field is not present in the source system, populate with null
- If a data field is present in the source system (e.g. there is a place to capture and store information), but the source value is null or blank, then populate with NI for no information.
- If a data field is present in the source system, but the source system explicitly denotes an unknown value, then populate with UN for unknown or otherwise specified unknown value.
- If a data field is present in the source system but the data cannot be mapped to the VDW, the populate with value for Other, if allowed.

Table Descriptions

Priority 1 Tables

CENSUS_LOCATION

The CENSUS_LOCATION table holds patient geographic location information collected at healthcare encounters. Patient addresses should be geocoded, and FIPS codes down to the census tract level should be populated in the CENSUS_LOCATION table. Every individual in the DEMOGRAPHIC table is represented in the CENSUS_LOCATION table at least once. For data partners that store historical addresses, the CENSUS_LOCATION table is historical; it contains census tracts of past residence(s) with dates when that location was valid. The 'current' address for an individual is identified through their missing location end date (LOC_END). A new record or row in the CENSUS_LOCATION table is generated whenever a patient has an encounter with the system and a new address is reported.

Patient addresses that cannot be geocoded, or if the patient is homeless, should be attributed to the county if possible (e.g. Denver county geocode=08031). Individuals with no residence information should have a record in the CENSUS_LOCATION table with no FIPS code.

Inclusions: Every PERSON_ID in the VDW DEMOGRAPHICS table is represented in the CENSUS_LOCATION table.

Exclusions: None

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL

LOC_START	Beginning date address for individual is valid (if date is unknown, assign a date estimated to when electronic address data began collection.	DATE NOT NULL Recommended format MMDDYYYY
LOC_END	Ending date address for individual is valid (dates prior to 01/01/2010 should default to 01/01/2010. If location record is a patient's current location, leave null. Queries will use this criterion to identify a patient's current address.	DATE NULL Recommended format MMDDYYYY
GEOCODE	Numeric string of multiple census variables. The geocode can be up to 15 digits long with geographic information hierarchically documented through the string of digits: state(2) + county(3) + tract(6), blockgp(1) + block(3) = GeoCode(15) When an address can be geocoded to a tract, the geocode is 11 digits long. When an address cannot be geocoded to a tract but can be geocoded to a county, the geocode is 5 digits long. Geocodes with block group and block information should be truncated to 11 digits to meet the requirements of a limited dataset (state(2) + county(3) + tract(6)). When an individual is identified as homeless through a flag or homeless value in the address fields, this field should be left blank.	CHAR(15) NULL
GEOCODE_BOUNDARY _YEAR	Optional field indicating the census year for which geocode applies (year in which geocode is valid)	NUMERIC(8) NULL
GEOLEVEL	Indicates the specificity of the GEOCODE match B = Block G =Block Group T = Census Track C = County Z = Zip Code P = Post Office U = Unknown, unable to append This can be assessed using logic that considers the length of the GEOCODE value (2 characters for state; 5 characters for county; 11 characters for census tract)Data partners should have the values T or C. Data partners should not map to B or G.	CHAR(1) NULL
MATCH_STRENGTH	Optional field indicating the type of match achieved by the geocoding software.	CHAR(1) NULL

	0 = No coordinates 1 = Zip +0 centroid 2 = Zip +2 centroid 3 = Zip +4 centroid 4 = Shape Path Centroid 5 = Street address position 6 = Point zip centroid X = Street Intersection	
LATITUDE	The latitude of the location. Note: this field will not be accessed by PMN queries	NUMERIC(8) NULL
LONGITUDE	The longitude of the location. Note: this field will not be accessed by PMN queries	NUMERIC(8) NULL
GEOCODE_APP	Optional field indicating the application used to generate FIPS codes/geocode addresses. Examples include "ESRI", "SAS" "ARCGIS"	CHAR(50) NULL

Geocoding

All CHORDS efforts include place-based analysis to be facilitated by the CENSUS_LOCATION table. Selected data partners without an internal geocoder have several options for acquiring and storing a patient geocode and affiliated place based information in the VDW census table.

- Option 1: Utilize a web-based geocoding service
- Option 2: Utilize internal geocoding resource (e.g. Centrus or ESRI product)

- Option 3: Partner with CDPHE for geocoding services

Registry Contributing Site Geocoding Options
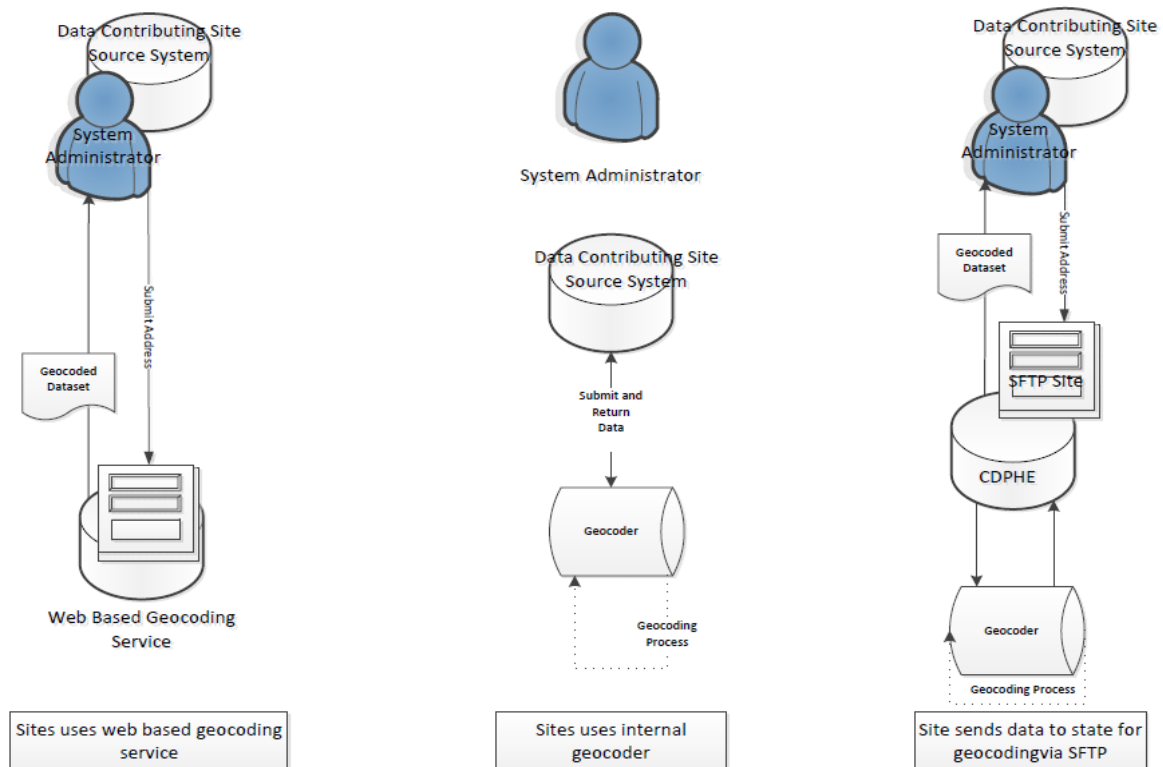
Process for Accessing Support for Geocoding Services

1) Contact Chris Wells at CDPHE to draft Business Associates Agreement (BAA) between data partner and CDPHE for geocoding services. BAAs usually originate at the covered entity (data partner) and the process should be initiated by the data partner.
2) Once the BAA is signed and executed, contact CDPHE (Chris Wells) to gain access to SFTP file sharing. Data partners will complete an access request form. CDPHE will create credentials and share the login information with data partner. This process takes approximately 1 to 2 weeks.

Process for Geocoding

1) Data partner pulls address file including with PERSON_ID, ENC_ID, and a complete address (includes city, state, and zip code) in CSV file. See Address File Format section below for exact field.
   a) Quarterly geocoding request should occur in February, May, August, and November.
   b) Whenever possible, data partner should email CDPHE contact to schedule the week that the file will be sent and geocoding will occur.
2) Data partner uploads address file to CDPHE SFTP in CHORDS data partner folder. To ensure security, each data partner will only have access to a partner specific folder.
3) CDPHE pulls down address file, geocodes addresses and creates a file in standard format including PERSON_ID, ENC_ID, and VDW CENSUS_LOCATION variables.

a) The CDPHE geocoding process will use the MapMarker USA geocoding software (utilizing updated TomTom and U.S. Census TIGER Centrus data) to examine each address string and determine whether it matches directly with an address string found in the data. In this process MapMarker USA defines a match and location code for each address that can be interpreted to determine if there is a direct match and accurately assigned location. CDPHE will interpret these values and assign the valid information to the output table for each address. In this case a rooftop latitude/longitude coordinate (6 decimal places each) and census tract ID (11 character string) is assigned to each matched address and non-matches are flagged.

b) If there is no direct match to an address string, an additional step is made to accurately assign a census tract ID (11 character string) or county census ID (5 character string) to each address using the "Fallback to ZIP centroid" process in MapMarker USA. No latitude or longitude will be assigned to any address in this additional step. In this process MapMarker USA defines a match and location code for each address that can be interpreted to determine if the assigned ZIP code lies completely inside an individual census tract boundary. CDPHE will interpret these values and assign the valid information to the output table for each address. Depending upon if this is true or not, a census tract ID (11 character string) or county ID (5 digits) is assigned to each non-matched address. Addresses for which the assigned ZIP centroid is not found are assigned a census tract ID or county ID.

c) Through some post processing, MapMarker USA will assign a county ID to P.O. Boxes and addresses with a value of 'Homeless' will not be assigned a Latitude and Longitude.

d) Data partners should allow 2 weeks for the geocoding process at CDPHE.

4) CDPHE uploads return file to SFTP data partner folder and will notify data partner when file is ready for pick up.

   a) If you have not heard from CDPHE after 14 days, please contact Chris Wells.

5) Data partner retrieves return file and loads geocoded variables into VDW as a part of VDW update.

Address File Format

Data partners should prepare a file for geocoding that includes the following data elements. BAAs are specific to data elements that are shared between organizations. No additional information or fields should be included in this file.

Variable name	Req/Opt	Description
PERSON_ID	Req	See Keys section.
ENC_ID	Req	See Keys section.
ADDRESS 1	Req	Line 1 Address
ADDRESS 2	Req	Line 2 Address (commonly used to denote apt. number, suite number, etc.)
CITY	Opt	City
STATE	Opt	State 2 letter code: CO for Colorado
ZIP	Req	Five digit zip code

Return File Format

CDPHE will prepare a standard return file for data partners.

Variable name	Req/Opt	Description

PERSON_ID	Req	See Keys section.
ENC_ID	Req	See Keys section.
ADDRESS 1	Req	Line 1 Address
ADDRESS 2	Req	Line 2 Address
CITY	Opt	City
STATE	Opt	State 2 letter code: CO for Colorado
ZIP	Req	Five digit zip code
VDW CENSUS_LOCATION	Req	Eleven digit Numeric string containing State (2), County (3), and/or Census Tract ID (6) *This variable should be used to populate the GEOCODE field in the Census Location table.*
Latitude	Req	Number with six decimal places. The latitude of the location for those addresses that directly matched.
Longitude	Req	Number with Six decimal places. The longitude of the location for those addresses that directly matched.

Census_Location Data Quality Checks

- Does every PERSON_ID have at least one record in the CENSUS_LOCATION table?
- Do a majority of geocodes represent the counties in and around your healthcare organization?
- Does every PERSON_ID have at least one CENSUS_LOCATION record with a missing LOC_END?
- Is PERSON_ID or LOC_START ever missing?
- What percent of PERSON_ID's have no geocode? Does this align with the frequency of homeless patients?
- Do some PERSON_IDs have one record while others have many records?
- Do any PERSON_IDs have an unreasonable number of CENSUS_LOCATION records?

DEMOGRAPHICS

The DEMOGRAPHICS table contains basic person-descriptive for individuals found in all other VDW tables and serves the traditional 'person table' role. The DEMOGRAPHICS table should reflect a patient's most recent demographic information. If a self-reported demographic attribute, such as race or primary language changes, this should be updated in the DEMOGRAPHICS table. Thus, the DEMOGRAPHICS table should be rerun when the VDW is updated to reflect any changes in patient demographic information.

One issue that can arise in the DEMOGRAPHICS table is duplicate patients. Duplicate patients should be removed/collapsed from the DEMOGRAPHICS table when they are identified in the healthcare organization. If an organization has a known problem with duplicates and no standard process to remove them, please note this on the data dictionary.

Inclusions: Every PERSON_ID appearing in any other VDW file should appear in DEMOGRAPHICS

Exclusions: No individuals are excluded from the DEMOGRAPHICS table

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
MRN	An arbitrary identifier unique to an individual within a data partner. Data partners may choose to store the patient MRN in the VDW Demographics table and use this value to link people across files within a data partner. May or may not contain the official local indigenous person identifier (e.g., "medical record number"). **Regardless of whether it contains the official local identifier, this variable should never leave the data partner and should never be queried with the PMN tool.**	CHAR(9) NOT NULL
BIRTH_DATE	Date of Birth	DATE NOT NULL Recommended format MMDDYYYY
GENDER	Gender or sex of the person M = Male F = Female U = Unknown O = Other (Transsexual, Transgendered, or anything else that does not fit into one of the prior categories) Default value: U.	CHAR(1) NOT NULL DEFAULT 'U'
PRIMARY_LANGUAGE	Primary language spoken at last contact. As defined in the ISO-639-2 specification, or 'UNK', for unknown. This variable should only be populated if assessment of primary language is captured within the EHR. Partial list: ENG = English SPA = Spanish	CHAR(3) NOT NULL DEFAULT 'UNK'
NEEDS_INTERPRETER	Needs interpreter to communicate w/an English-only speaker? Y = Yes N = No U = Unknown This variable should only be populated if assessment of interpreter needs is captured within the EHR. If a person speaks English, Data partners should not infer that the person does not need an interpreter and vice	CHAR(1) NOT NULL DEFAULT 'U'

	versa for non-English primary languages. Default value: U.	
RACE1	Race(s) of the person HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported UN = Unknown or Not Reported Default value: UN	CHAR(2) NOT NULL DEFAULT 'UN'
RACE2	Race(s) of the person HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported UN = Unknown or Not Reported Default value: UN	CHAR (2) NOT NULL DEFAULT 'UN'
RACE3	Race(s) of the person HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported UN = Unknown or Not Reported Default value: UN	CHAR (2) NOT NULL DEFAULT 'UN'
RACE4	Race(s) of the person HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported	CHAR (2) NOT NULL DEFAULT 'UN'

	UN = Unknown or Not Reported	
	Default value: UN	
RACE5	Race(s) of the person	CHAR (2) NOT NULL DEFAULT 'UN'
	HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported UN = Unknown or Not Reported	
	Default value: UN	
HISPANIC	Hispanic origin (ethnicity). Y = Yes N = No U = Unknown This variable should only be populated if values for race are captured within the EHR. If ethnicity and race are captured in the same field within the EHR, values of Hispanic for race may be mapped to Hispanic='Y'. Default value: U.	CHAR (1) NOT NULL DEFAULT 'U'

Demographics Data Quality Checks

- Is PERSON_ID or BIRTH_DATE ever null?
- Does the distribution of gender reflect your patient population? Does the distribution of race and ethnicity reflect your patient population?
- When current age is calculated using the BIRTH_DATE, does the mean age reflect your patient population?
- Are any BIRTH_DATEs in the future?
- Does the number of patients more than 100 years of age reflect your patient population?
- Do the most common primary languages represent your patient population?
- Does the proportion of patients with an unknown race, Hispanic, gender, and primary languages represent what you know about your workflow and EHR system?
- Are any BIRTH_DATEs unreasonable?
  - SQL will sometimes fill NULL with 1900-01-01

DIAGNOSES

This table lists all diagnoses associated with the encounters in the ENCOUNTERS table. A record is a diagnosis code/original diagnosis code/diag provider combination unique to the index variable ENC_ID combination. The index variable ENC_ID uniquely identifies each encounter and is used to link the ENCOUNTERS file to the both the DIAGNOSES and PROCEDURES files. Changes in diagnosis coding from ICD-9 to ICD-10 in October 2015 are expected. Data partners should not include the problem list or chief complaint as a source of diagnoses. Also, do not include admitting diagnosis for inpatient stays.

Inclusions: Include all diagnoses associated with each encounter, primary and otherwise. Only final diagnoses should be included.

Exclusions: Problem list, chief complaint and admitted diagnoses.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
ENCTYPE	Encounter Type. See ENCOUNTER table below.	CHAR(2) NOT NULL
ENC_ID	See Keys section above	CHAR(32) NOT NULL
PROVIDER	See Keys section above Provider code for the provider who is most responsible for this encounter. Usually physician, nurse practitioner, physician assistant, optometrist, etc. For encounters with multiple providers, choose a single provider so the encounter can be linked to the diagnosis and procedure files. If there is no provider code for an encounter, then specify the value for provider as "UNK". Many data partners reserve a dummy provider ID assigned when the provider of services is unknown. Do not map that to a provider value in VDW; instead replace this with "UNK".	CHAR(15) NOT NULL
DIAGPROVIDER	Diagnosing Provider. Provider that made the diagnosis. If this provider is unknown, set equal to the provider variable.	CHAR(15) NOT NULL
ADATE	Encounter date or admit date for an outpatient, inpatient or institutional stay. If the encounter date or admit date is unknown from a claim, then use the first date of a claim.	DATE NOT NULL Recommended format MMDDYYYY
DX	Standard Diagnosis codes	CHAR(10) NOT NULL
ORIGDX	Original Diagnosis code from Source Data	CHAR(10) NOT NULL
DX_CODETYPE	ICD Coding Version 07 = ICD-7-CM (including ICD-7) 08 = ICD-8-CM (including ICD-8) 09 = ICD-9-CM (including ICD-9) 10 = ICD-10-CM (including ICD-10)	CHAR(2) NOT NULL DEFAULT 'UN'

	11 = ICD-11-CM (including ICD-11) OT= Other UN= Unknown A transition from ICD-9 to ICD-10 in October 2015 is expected here.	
PRINCIPAL_DX	Principal Diagnosis Flag P = Principal Dx N = Not Principal Dx X = Principal Dx Status not classifiable Assigned after discharge after review by the medical record department, the **principal** diagnosis is main reason why the patient was admitted to the hospital for care. This is the diagnosis on which the DRG is based. Note that the principal diagnosis is very different from the admitting diagnosis which is assigned at the beginning of the stay. For example, if a patient was admitted to a hospital with an admitting diagnosis of chest pain which was later diagnosed as a heart attack during the stay, the principal diagnosis would be heart attack. Specify principal diagnosis as defined by the data partner's institutional source data. It is expected that there should be one and only one principal diagnosis per hospitalization.	CHAR(1) NOT NULL DEFAULT 'X'
PRIMARY_DX	Primary Diagnosis Flag P = Primary Dx S = Secondary Dx X = Primary Dx Status not classifiable Primary diagnosis is the illness or injury that was the most serious/severe/life-threatening and/or resource intensive. For an outpatient encounter, it is expected that there should be one and only one primary diagnosis. For an inpatient stay, there can be multiple primary diagnoses, one for each provider claim during the stay.	CHAR(1) NOT NULL DEFAULT 'X'

Diagnoses Data Quality Checks

- Is PERSON_ID, ADATE, DX, or ENC_ID ever null?

- Are any values for ADATE in the future?
- Are there diagnoses for every day that there are encounters?
- Do encounters have secondary diagnoses without primary diagnoses?
- Is the most common diagnosis code what you would expect?
- Do all diagnosis codes have a DX_CODETYPE
- Are any diagnosis dates in the future?
- ENC_ID cross check with ENCOUNTERS table to ensure encounter details exist

ENCOUNTERS

The ENCOUNTERS table contains information about all health care encounters of any kind (email, ambulatory, phone call) that are captured and tracked within a healthcare organization's electronic data for every patient in the DEMOGRAPHIC table. Encounters are categorized through the encounter type and subtype fields (ENCTYPE and ENCOUNTER_SUBTYPE). For a payer, encounters at other systems (i.e. emergency department visits at Children's hospital for a KPCO patient) may be included in the ENCOUNTERS table. At a healthcare organization, encounters or healthcare services at other healthcare providers outside of the Denver Health enterprise are not included. Each encounter has a unique encounter identifier (ENC_ID). Every individual in the encounter table is in the demographics table. For a payer, an individual may exist in the DEMOGRAPHICS table without having an encounter because they had active insurance coverage but received no healthcare services.

For ambulatory visit, telephone, e-mail, lab only, radiology only and "other" encounters, a unique encounter record, is defined as each patient and provider medical contact documented in the source data (exclude scheduling appointments and other administrative tasks). However, if the source data contain separate records with a lab or radiology tech, then code these encounters as "Lab Only" or "Radiology Only" encounter types respectively. Classify "Vaccine Only" encounters as "Other Encounter". Visits to the pharmacy to pick-up medications are not encounters and therefore are excluded.

Tips for Populating Encounter:

- The ENCOUNTERS table is rich and complex, just as medical care is rich and complex. We are asking data partners to match the source EHR data systems when the source data matches this specifications document. However, no amount of detail provided here can ensure a seamless match between the EHR and VDW. Each project should decide how to handle the differential ascertainment of encounters across the data partners participating in the specific study and document their logic and document these decisions.
- A single inpatient stay, non-acute institutional stay, such as a skilled nursing facility, rehab, nursing home, overnight hospice or dialysis stay, or emergency visit has one record in the ENCOUNTERS table.
- For ambulatory visits, telephone, e-mail, lab only, radiology only and "other" encounters, a unique encounter record is defined as each patient and provider medical contact documented in the source data (exclude scheduling appointments and other administrative tasks).

- The ENCOUNTERS table only contains data where a medical provider interacts with a patient. Medical providers include: physicians, nurse practitioners, registered nurses, lab technicians, social workers, etc.—generally, people licensed to provide medical care and closely related services.
- Ambulatory Visits (a subset of the ENCOUNTERS table) are limited to outpatient encounters where the provider is licensed to prescribe medical services. However, if the source data contains separate records with a lab or radiology tech, then code these encounters as "Lab Only" or "Radiology Only" encounter types respectively.
- Recurring visits to the same clinicians on the same day should be maintained as separate encounters, if possible.
- Classify "Vaccine Only" encounters as "Other Encounter."
- Telephone calls should be included only if the call was between the patient and a provider who is licensed to prescribe medical services. Hence, a call to schedule or cancel an appointment should be excluded.
- Visits to the pharmacy to pick-up medications are not encounters and, therefore, are excluded.
- Classify same-day inpatient discharges as inpatient if the patient is admitted to the hospital for an inpatient stay.
- Rules about transfers within hospitals: Treat as one stay in the ENCOUNTERS table if a patient is transferred from one acute inpatient station to another acute inpatient station within the same hospital. Treat as separate encounters if a patient is transferred from an acute inpatient station to a long term care station (such as a skilled nursing facility, rehab or other non-acute inpatient care).
- The PROVIDER variable is most useful for outpatient encounters. Inpatient stays should have a single provider for the entire stay, even if multiple providers performed procedures during the stay. If possible, use the admitting physician as the provider for all care during the stay. (Note that the PROCEDURE and DIAGNOSES tables have fields to signify the provider that actually performed a given procedure/made a given diagnosis where that is known.)
- Classify admissions to residential alcohol and chemical dependency programs as "Rehab".
- The index variable ENC_ID uniquely identifies each encounter and is used to link the ENCOUNTERS table to the both the DIAGNOSES and PROCEDURES tables. Multiple encounters to the same provider on the same day are allowed if that is the truth in the source data and have unique ENC_ID values.
- All variables are required to simplify analysis. Data partners should set variables to missing or unknown if the variable is unavailable.
- HEDIS definitions can be useful to classify encounters into encounter type and subtype values at your data partner.

Inclusions: The ENCOUNTERS table only contains data where a medical provider interacts with a patient

Exclusions: Exclude no-show appointments, calls to schedule appointments and medication pick-up only.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
ADATE	Encounter date or admit date for an outpatient, inpatient or institutional stay. If the encounter date or admit date is unknown from a claim, then use the first date of a claim.	DATE NOT NULL Recommended format MMDDYYYY
ENC_ID	See Keys section above	CHAR(36) NOT NULL
DDATE	Discharge date. Use for hospital and overnight encounters. **This field may be left null for outpatient encounters.**	DATE NULL Recommended format MMDDYYYY
PROVIDER	See Keys section above	CHAR(15) NOT NULL
ENC_COUNT	Encounter Count: Number of visits for this encounter. Usually 1. The reason for this variable is to track the number of visits from a claim when the visit dates for each encounter can't be identified. For example, the dialysis visit dates are not identified in the claims data at some data partners but the number of dialysis visits is recorded.	NUMERIC(3) NOT NULL
DRG_VERSION	Identifies the version of the Diagnostic Related Group. Expected for hospital encounters and some institutional stays but include for all encounters if in the source data. **A** = DRG values using the old coding system **B** = DRG values using the current coding system The DRG coding system officially changed 10/1/2007, but the variable should indicate the code version used, regardless of the actual date of conversion at the individual data partner which may be different from 10/1/2007. Must use with DRG_VALUE.	CHAR(1)
DRG_VALUE	Identifies the value of the Diagnostic Related Group. Used for hospital encounters. Use leading zero's for codes less than 100. Must use with DRG_VERSION to ensure appropriate classification of values.	CHAR(4)
ENCTYPE	Encounter Type. Valid Encounter Subtype values are in brackets "[]."See Encounter type and subtype matrix below. **IP** = Acute Inpatient Hospital Stay: Inpatient stays, same-day hospital discharges, hospital transfers	CHAR(2) NOT NULL

	where the patient was admitted into the hospital. Includes acute inpatient psych and detox hospital stays. [ENCOUNTER_SUBTYPE = AI] **ED** = Emergency Department Encounter: Excludes urgent care visits. [ENCOUNTER_SUBTYPE = HA,OC] **AV** = Ambulatory Visit: Outpatient clinics, same day surgeries, observation beds, urgent care visits, and other same-day ambulatory hospital encounters. Excludes emergency department encounters). [ENCOUNTER_SUBTYPE = OC, OB, SD, HA, UC, RH, DI, OT] **TE** = Telephone Encounters: [ENCOUNTER_SUBTYPE = OT, HH] **EM** = E-mail Encounters: [ENCOUNTER_SUBTYPE = OT, HH] **IS**=Non-Acute Institutional Stays: Hospice, SNF, rehab, nursing home, residential, overnight non-hospital dialysis and other non-hospital stays. [ENCOUNTER_SUBTYPE = HS, SN, NH, RH,DI, OT] **OE**=Other Encounters (not overnight): Hospice visits, home health visits, SNF visits, other non-hospital visits. [ENCOUNTER_SUBTYPE = HS, HH, SN, RH, DI, OT] **LO**=Lab Only Encounter: Optional. Lab encounters that cannot be matched to another encounter. Include to link variables from ENCOUNTER table to the PROCEDURES table. [ENCOUNTER_SUBTYPE = OC,OT] **RO**=Radiology Only Encounter: Optional. Radiology encounter that cannot be matched to another encounter. Include to link variables from utilization file to procedure file. [ENCOUNTER_SUBTYPE = OC,OT]	
ENCOUNTER_SUBTYPE	AI = Acute Inpatient Stay. Excludes observation bed. HA = Hospital Ambulatory. Outpatient care at a hospital excluding same day surgery and observation beds. OC = Outpatient Clinic Visit OB = Observation Bed	CHAR(2) NOT NULL

	SD = Same Day Surgery UC = Urgent Care DI = Dialysis OT = Other Non-hospital HH = Home Health HS = Hospice SN = Skilled Nursing Facility NH = Nursing Home (includes ICF) RH = Rehab	
FACILITY_CODE	Facility code that identifies hospital or clinic. Data partners should enter data partner specific data partner or clinic identifiers to aid data validation. If no such variable exists, Specify the FACILITY_CODE as "UNK" if the facility code is missing.	CHAR(12) NOT NULL DEFAULT 'UNK'
DISCHARGE_DISPOSITION	A = Alive E = Expired U = Unknown	CHAR(1) NOT NULL DEFAULT 'U'
DISCHARGE_STATUS	Discharge Status. AF = Adult Foster Home AL = Assisted Living Facility AM = Against Medical Advice AW = Absent without leave EX = Expired HH = Home Health HS = Hospice HO = Home / Self Care IP = Other Acute Inpatient Hospital NH = Nursing Home (Includes ICF) OT = Other RS = Residential Facility RH = Rehabilitation Facility SH = Still In Hospital SN = Skilled Nursing Facility UN = Unknown	CHAR(2) NOT NULL DEFAULT 'UN'
ADMITTING_SOURCE	Admitting Source. AV = Ambulatory Visit ED = Emergency Department AF = Adult Foster Home AL = Assisted Living Facility HH = Home Health HS = Hospice HO = Home / Self Care IP = Other Acute Inpatient Hospital NH = Nursing Home (Includes ICF) OT = Other	CHAR(2) DEFAULT 'UN'

	RS = Residential Facility RH = Rehabilitation Facility SN = Skilled Nursing Facility UN = Unknown Use for hospital and overnight encounters	
DEPARTMENT	Department Code. Data Partners should crosswalk information about encounter to department categories. ACUP = Acupuncture ALGY = Allergy AMBU = Ambulance Services ANES = Anesthesiology AUD = Audiology CARD = Cardiology CASR = Cast Room CHEM = Chemical and Alcohol Dependency CHIR = Chiropractic CMHL = Community Health CRIT = Critical Care Medicine CRMG = Care Management DENT = Dental DERM = Dermatology DIAL = Dialysis DME = Durable Medical Equipment EDUC = Education ENDO = Endocrinology ENT = Otolaryngology ER = Emergency Room FP = Family Practice GEN = Genetics GER = Gerontology/Geriatrics GI = Gastro-Intestinal Medicine HAP = Health Appraisals HEP = Hepatology HH = Home Health HOSP = Hospital Care HSPC = Hospice ICF = Intermediate Care Facility IM = Internal Medicine IMUN = Immunology IND = Industrial Medicine INF = Infectious Disease INFU = Infusion Center IR = Injection Room LAB = Laboratory MH = Mental Health	CHAR(4) DEFAULT 'UNK'

	NATU = Naturopathy
	NEPH = Nephrology
	NEUR = Neurology
	NEWB = Newborn
	NRSG = Neurosurgery
	NUCL = Nuclear Medicine
	NUT = Nutrition
	OBGN = Obstetrics/Gynecology
	OCTH = Occupational Therapy
	ONC = Oncology
	OPTH = Ophthalmology
	OPTO = Optometry
	ORTH = Orthopedics
	OST = Osteopathy
	PAL = Palliative Care
	PATH = Pathology
	PC = Primary Care
	PEDS = Pediatrics
	PERI = Perinatology
	PHYS = Physiatry
	POD = Podiatry
	PSRG = Plastic Surgery
	PT = Physical Therapy
	PULM = Pulmonary Medicine
	RAD = Radiology
	RADT = Radiation Therapy
	RECT = Recreational Therapy
	REHB = Rehabilitation
	RESP = Respiratory Therapy
	RHEU = Rheumatology
	RN = Registered Nurse
	SNF = Skilled Nursing Facility
	SPOR = Sports Medicine
	SPTH = Speech Therapy
	SURG = General Surgery
	TRAN = Transplant
	URG = Urgent Care
	URO = Urology
	OTH = Other
	UNK = Unknown

Mapping encounter types and subtypes can be a challenging task. Data partners should consider what services are provided by their healthcare organization and how to appropriately represent those encounters using both variables. The allowed combinations of encounter type and subtype are represented by this matrix.

Encounter Type and Subtype Matrix

ENCTYPE	ENCOUNTER_SUBTYPE												
	AI	OB	OC	SD	UC	HA	HS	HH	SN	NH	RH	DI	OT
IP	A												
ED			A			A							
AV		A	A	A	A	A					A	A	A
TE								A					A
EM								A					A
IS							A		A	A	A	A	A
OE							A	A	A		A	A	A
LO			A										A

Encounters Data Quality Checks

- Is PERSON_ID, PROVIDER or ENCOUNTER_ID ever null?
- Do the services reflect what your facility provides to patients in terms of encounter type and subtype?
- Is ADATE missing?
- Is DDATE prior to ADATE?
- Are there encounters every day of the week (weekends include if weekend healthcare services are provided)
- Does the mean number of encounters per patient represent a logical pattern?
- Is the average length of stay for IP encounters logical?
- Is ADATE in the future?
- Does the date fall within the correct range (first date of historical data to today)?
- Are any ADATEs or DDATEs unreasonable?
  - SQL will sometimes fill NULL with 1900-01-01

VITAL_SIGNS

The VITAL_SIGNS table includes various physiological measures taken by health professionals during the clinic visit. The traditional clinical vital signs are body temperature, pulse rate, blood pressure, and respirations. Because of HMORN investigator interest, the VDW VITAL_SIGNS table also includes anthropometry (height, weight, BMI [when pre-calculated], and head circumference).

There is no unique primary key for the VITAL_SIGNS table. The PERSON_ID and MEASURE_DATE are the only variables required to be non-null for inclusion in the table.

Inclusions: Include all encounters, even when all of the vitals elements are missing.

Exclusions: None

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL

MEASURE_DATE	Encounter or measure date	DATE NOT NULL Recommended format MMDDYYYY
MEASURE_TIME	Encounter or measure time (if available)	TIME NOT NULL
ENC_ID	See Keys section above. This field is meant to provide more information about the setting of a vitals measurement where available. It should only be populated if there is an encounter in the utilization file that can be definitively matched to the vital record.	CHAR(36) NOT NULL
ENCTYPE	Encounter Type. See ENCOUNTERS table. Useful for interpreting measures collected in different clinical contexts will have different interpretations (e.g., BP collected in ED trauma situation may not reflect typical BP).	CHAR(2) NOT NULL
HT_RAW	Optional. Height in inches. Non-standard values may be suitable for the raw versions of these variables.	CHAR(10)
WT_RAW	Optional. Weight in pounds. Non-standard values may be suitable for the raw versions of these variables.	CHAR(10)
HT	Height in inches. Null if height is missing or not expressed as a single number (e.g., a range is given, or just a description like "medium"). Only populate if height was measured on the date given in MEASURE_DATE. Do not impute or carry over from other dates.	NUMERIC(7, 3)
WT	Weight in pounds. Null if weight is missing or not expressed as a single number (e.g., a range is given, or just a description like	NUMERIC(8, 4)

	"normal"). Only populate if weight was measured on the date given in MEASURE_DATE. Do not impute or carry over from other dates.	
BMI_RAW	Body Mass Index. Data partners can include BMI directly from EHR as their option. The main purpose of this variable is to facilitate capture of body mass information directly as taken from legacy electronic data (i.e., as BMI range or code). See Notes below.	CHAR(5)
DIASTOLIC	Diastolic blood pressure in mmHg. Integers between 0 and 999.	NUMERIC(3)
SYSTOLIC	Systolic blood pressure in mmHg. Integers between 0 and 999.	NUMERIC(3)
DIASTOLIC_RAW	The original diastolic blood pressure from the EMR. Non-standard values may be suitable for the Raw versions of these variables. See Notes below.	CHAR(7)
SYSTOLIC_RAW	The original systolic blood pressure from the EMR. Non-standard values may be suitable for the Raw versions of these variables. See Notes below.	CHAR(7)
BP_TYPE	Optional. Type of blood pressure taken. R = Rooming O = Orthostatic M = Multiple E = Extended	CHAR(1)
POSITION	Optional. Position for Orthostatic Blood Pressures. 1 = Sitting 2 = Standing 3 = Supine Null = Unknown	CHAR(1)
HEAD_CIR_RAW	Head Circumference (cm)	CHAR(6)

RESPIR_RAW	Respiration in breaths per minute	CHAR(6)
TEMP_RAW	Temperature	CHAR(6)
PULSE_RAW	Heartbeats per minute	CHAR(6)

Note 1: Body Mass Index (BMI) - Raw

BMI is a measure of relative body weight, derived (calculated from) height and weight. The recommended approach is to calculate BMI at the time it is actually needed, and write it out to another table rather than computing it in advance and storing it in the VITAL_SIGNS table. The main purpose of including a BMI variable (BMI_RAW) in the VITAL_SIGNS table is to accommodate handling of values that may be stored in data partner electronic data as categorical values (e.g., corresponding to a range of BMI or to a diagnosis).

Note 2: Blood Pressure Measures - Raw

DIASTOLIC_RAW/SYSTOLIC_RAW — These variables can be used for data partners that have blood pressures reported in non-standard ways such as when blood pressure is reported as falling within a range of values, or through a diagnosis. It can be numeric, text or a combination of both. Each data partner will have different valid values, which will be documented in the Data partner Specific Notes.

While it is the common practice for BP measures to be expressed in units of millimeters of mercury (mmHg) the values in SYSTOLIC_RAW and DIASTOLIC_RAW may be in any units. Please see the data partner-implementation pages for details on what each data partner uses for units. It is *not* safe to assume mmHg.

Vital_Signs Data Quality Checks

- Is PERSON_ID ever null?
- Are all WTs below 700 pounds?
- Are all HTs below 90 inches?
- Are the mean and median HT and WT within a reasonable range?
- Is SYSTOLIC below 300?
- Are any HTs and WTs less than 0 or negative?
- Is MEASURE_DATE in the future?

Priority 2 Tables

LAB_RESULTS

The VDW Lab Data Structure consists of two tables, but only the LAB_RESULTS table is included as a Priority 2 table. The LAB_RESULTS table holds structured data about the results of laboratory tests and has one record per inpatient or outpatient lab result. If a test has no result for whatever reason (e.g. specimen not sufficient, specimen contaminated, or patient did not show) then the test is not included as a record in the LAB_RESULTS table.

Inclusions: Lab tests are included in this table if they were complete and had a valid result.

Exclusions: Lab tests where the patient did not show up, duplicates, contaminated specimens, or inconclusive results are not included.

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
RESULT_DT	Date of lab result	DATE NOT NULL Recommended format MMDDYYYY
RESULT_TM	Time of lab results (if captured)	Optional. This variable may not be available at all data partners. This field may have all NULL values if this information is not captured and stored electronically. Recommended time format HH:MM
LAB_DT	Date specimen was collected	DATE NOT NULL Recommended format MMDDYYYY
LAB_TM	Time specimen was collected (if captured)	Optional. This field may have all NULL values if this information is not captured and stored electronically. This variable may not be available at all data partners. Recommended format HH:MM
TEST_TYPE	Test description of lab test type. These may vary by data partner but should be the best description of lab test.	CHAR(20) NOT NULL
LOINC	LOINC Code (if populated in source system). Do not fill LOINC codes with leading zeros; format is left-justified. LOINC codes are from 3-	CHAR(10) Note: Do not map LOINC codes manually; complete

	7 characters long. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. The hyphen, as well as all the numbers, is part of the LOINC and must be included. More information about LOINC codes is available here: http://loinc.org/.	this field if the LOINC code is available in the source system. This variable may not be available at all data partners.
STAT	Immediacy of lab test. The intent of this value is to determine whether the test was obtained as part of routine care or as an emergent/urgent diagnostic test (designated as Stat or Expedite). Some sites have Expedite as a value and some sites do not. The default value should be Routine unless the results for that test indicate Expedite or Stat. E = Expedite S = Stat R = Routine U = Unknown or missing	CHAR(1) NOT NULL DEFAULT 'R' Note: Do not infer from encounter information. Populate this field if captured and stored in the source system. This variable may not be available at all data partners.
PT_LOC	Location of the patient when the lab specimen was obtained. Patient location is intended to provide information about the location of the patient when the lab specimen was obtained. That is, was the specimen obtained in conjunction with a patient hospitalization, an emergency department visit, an outpatient medical office visit, was the patient at home, or was information not available to determine where the patient was when the specimen was obtained? I = Inpatient O = Outpatient E = Emergency Department H = Home U = Unknown or missing	CHAR(1) NOT NULL DEFAULT 'U' This field provides information about the location of the patient when the lab specimen was obtained. That is, was the specimen obtained in conjunction with a patient hospitalization, an emergency department visit, or an outpatient medical office visit? Populate if information is available to determine where the patient was when the specimen was obtained?
RESULT_LOC	Location of lab test result. Result location is used to identify whether the test was completed in a laboratory (e.g., hospital lab, outpatient clinic lab, emergency department lab) or if the test was completed using a point of care device at a location other	CHAR(1) NOT NULL DEFAULT 'L' The variable will have the default value of 'L' unless the result is Point of Care. This variable may not be available at all data

	than a certified laboratory (e.g., newborn nursery, finger stick in provider office, home). L = Lab P = Point of Care	partners. Some data partners do not include point of care information in their laboratory databases. However, this variable is captured at some data partners and in some situations (e.g., INR testing at an anticoagulation clinic may be incorporated into the lab databases at some data partners).
LOCAL_CD	Local lab test identifier specific to each data partner.	CHAR(25) NOT NULL Most healthcare organizations use home-grown lab test codes (non-LOINC codes) to identify and group lab tests which should be populated in this field. This field (and BATTERY_CD) may be used to develop mappings and retain data lineage.
BATTERY_CD	Local lab test identifier specific to each data partner.	CHAR(50) NOT NULL Most healthcare organizations use home-grown lab test names (non-LOINC codes) to identify and group lab tests which should be populated in this field. For example, 'HBA1C' or 'HIV_RNA'. This field may be used to develop mappings and retain data lineage.
PX	Procedure code related to administration of lab test. This information may be extracted from billing or orders and is most likely a CPT, HCPCS, ICD-9 or ICD-10 code.	CHAR(6) NOT NULL If not procedure code information is available, this field may default to a local lab test code defined by the data partner.

CODETYPE	Procedure code type for lab test procedure code field (most likely CPT or Local). C = CPT (format: 4 #####) I = ICD-9 (format: ##.##) T = ICD-10 (format: H = HCPCS (format: A####) L = Local Home-Grown O = Other ICD-10-PCS – The procedure classification system developed by the Centers for Medicare & Medicaid Services (CMS) for use in the U.S. for inpatient hospital settings only. The new procedure coding system uses 7 alpha or numeric digits while the ICD-9-CM coding system uses 3 or 4 numeric digits.	CHAR(1) NOT NULL DEFAULT 'L' If procedure code type is not stored electronically, mapping of codes is permitted. See **Lab Result Procedure Code Type Mapping example code below.**
ORDER_ID	This variable will be used to link lab orders to lab results and should be populated for the unique lab results order identifier which is unique to each data partner.	CHAR(35) NOT NULL If this variable is no already available, data partners may generate a unique ORDER_ID.
ORDER_DT	Date of lab test order, if lab test order information is available.	DATE This variable may not be available at all data partners. Recommended format MMDDYYYY
RESULT_C	Standardized result for qualitative and quantitative lab test results. Examples of valid qualitative values include: POSITIVE, NEGATIVE, UNDETERMINED, BORDERLINE For quantitative results, remove symbols >, <, >=, <=. For Example, if the result is <200, populate RESULT_C =200 and use the modifier value of 'LT'.	CHAR(8) NOT NULL If no valid result value can be populated in RESULT_C, no record should be created in the LAB_RESULTS table. If the input data contains a comment in the same field as the result value then RESULT_C should contain only the result value and the comment should be stored in the LAB_NOTES table.

MODIFIER	Modifies the values in RESULT_C. For qualitative results such as 'POSITIVE', modifier='TX'. If the result is an exact value (e.g. 200) then RESULT_C=200 and modifier='EQ'. TX = Text (for qualitative result values) EQ = Equal LT = Less than LE = Less than or equal to GT = Greater than GE = Greater than or equal to	CHAR(2) NOT NULL
RESULT_UNIT	Units in which result is reported. Example values include %, IU/L, MEQ/L, MG/DL, THERAPEUTIC.	CHAR(11) Many result units exist and more information may be found on the international system of units.
NORMAL_LOW_C	Lowest value of normal range for this test as short character. Some lab tests will have a range; some will have only low or high restrictions (e.g. greater than 6ppm or less than 9%). To accept digits, decimal, sign or text. The symbols >, <, >=, <= should be removed.	CHAR(8) For example, the normal range for neutrophils is 2.5-7.5 x 10^9 per liter. To represent this range, NORMAL_LOW_C = '2.5' MODIFIER_LOW = 'LT' NORMAL_HIGH_C = '7.5' MODIFIER_HIGH = 'GT'
MODIFIER_LOW	Modifies the values in NORMAL_LOW_C. Value is determined by considering the existence of or lack of the symbols >, <, >=, <= or the existence of text and mapping to allowed values below. TX = Text EQ = Equal (only used for a range when upper and lower boundaries are both supplied) LT = Less than (this value will probably not be used) LE = Less than or equal to (this value will probably not be used) GT = Greater than	CHAR(2) For example, The normal range for neutrophils is 2.5-7.5 x 10^9 per liter. To represent this range, NORMAL_LOW_C = '2.5' MODIFIER_LOW = 'LT' NORMAL_HIGH_C = '7.5' MODIFIER_HIGH = 'GT'

	GE = Greater than or equal to null is appropriate when NORMAL_LOW_C is null	
NORMAL_HIGH_C	Highest value of normal range for this test as short character. To accept digits, decimal, sign or text. The symbols >, <, >=, <= should be removed.	CHAR(8) For example, The normal range for neutrophils is 2.5-7.5 x 10^9 per liter. To represent this range, NORMAL_LOW_C = '2.5' MODIFIER_LOW = 'LT' NORMAL_HIGH_C = '7.5' MODIFIER_HIGH = 'GT'
MODIFIER_HIGH	Modifies the values in NORMAL_HIGH_C. Value is determined by considering the existence of or lack of the symbols >, <, >=, <= or the existence of text and mapping to allowed values below. TX = Text EQ = Equal LT = Less than LE = Less than or equal to GT = Greater than GE = Greater than or equal to UN = Unknown (null values should be set to UN)	CHAR(2) NULL For example, The normal range for neutrophils is 2.5-7.5 x 10^9 per liter. To represent this range, NORMAL_LOW_C = '2.5' MODIFIER_LOW = 'LT' NORMAL_HIGH_C = '7.5' MODIFIER_HIGH = 'GT'
ABN_IND	Abnormal result indicator AB = Abnormal C = Critical AH = Abnormal High CH = Critical High AL = Abnormal Low CL = Critical Low IN = Inconclusive UN = Unknown (null values should be set to UN) NL = Normal	CHAR(2) NOT NULL DEFAULT='UN' This variable may not be available at all data partners. If the ABN_IND is missing do not impute a value. Missing ABN_IND values should be set to 'UN'.
ORDER_PROV	Data partner defined unique provider identifier for the provider ordering the lab test. Same coding scheme as RXMD in PHARMACY. Unique to each Data partner.	CHAR(15)

ORDER_DEPT	Department of ordering provider as bucketed by scheme used in the utilization file. Same coding scheme as DEPARTMENT in ENCOUNTERS.	CHAR(4) It is STRONGLY recommend that a data partner keep its own local department number in this file in addition to the ORDER_DEPT so bucketing of department by different schemes is possible.
FACILITY_CODE	Local facility code that identifies hospital or clinic. Same coding scheme as FACILITY_CODE in ENCOUNTERS. Unique to each data partner.	CHAR(12)
SPECIMEN_ID	Used to connect multiple lab result records from the same specimen. For example, the ID for the blood sample to create a close estimate to the absolute neutrophil count	CHAR(35) A possible way to create SPECIMEN_ID is by concatenating lab accession number and date. At most data partners lab accession number is reused, thus concatenating the date will make a unique SPECIMEN_ID for that TEST_TYPE.
SPECIMEN_SOURCE	Sample of method that resulted in a specimen to be tested. Some tests can be performed on either, and both values are listed in the specimen source field. BLOOD = Blood SERUM = Serum PLASMA = Plasma SR_PLS = Serum/Plasma PPP = Platelet Poor Plasma CSF= Cerebral Spinal Fluid URINE = Urine STOOL = Stool NSWAB = Nasal Swab (Any specimen source in your data listed as "nose" should be mapped to "nasal swab.")	CHAR(6) Populated SPECIMEN_SOURCE may require data partner to map from many to one or more SPECIMEN_SOURCE information to allowed values.

	NWASH = Nasal Wash NPH = Nasopharyngeal Swab NPWASH = Nasopharyngeal Wash THRT = Throat, Oropharyngeal Swab SPUTUM = Sputum BAL = Bronchoalveolar Lavage (BAL) Specimen BALBX = BAL Biopsy OTHER = Other NS = Not Specified BLANK = Missing Value (assumptions can probably be made depending on TEST_TYPE)	
ROW_ID	Data partner identifier that links a LAB_RESULTS record to corresponding LAB_NOTES record(s). NOTE: Most records in LAB RESULTS will have a missing ROW_ID since most records do not need to link to the LAB_NOTES table. Any record in LAB_RESULTS that has a value other than missing will have a unique value. That is, any ROW_ID value other than missing will appear only once in the LAB_RESULTS data set.	CHAR(8)

Lab Result Procedure Code Type Mapping

```
Case when len(px) >= 7  then '10'
      when px like '[0-9][0-9][0-9][0-9][0-9]' then 'C4'
      when px like '[0-9][0-9][0-9][0-9][A-Z]' then 'C4'
      when px like '[0-9][0-9].[0-9]%' then '09'
      when px like '[0-9][0-9][0-9].[0-9]%' then '09'
      when px like '[A-Z][0-9][0-9].[0-9]%' then '09'
                      when px substring(dw_px,1,1) like '[A-Z]%'
          and px substring(dw_px,2,4) like '[0-9][0-9][0-9][0-9]'
          and len(dw_px) = 5  then 'HC'
                      Else 'OT'
```

Lab Results Data Quality Checks

- Is PERSON_ID or ROW_ID ever missing?
- Are any values for RESULT_DT, LAB_DT or ORDER_DT in the future?

LANGUAGES

The LANGUAGES table contains information on the languages that patients speak and write, when this information is collected and stored by a data partner. There is one record per person from the demographics table per known language.

Inclusions: All individuals in the DEMOGRAPHICS table.

Exclusions: None.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
LANG_ISO	A code signifying a particular language based on ISO-639-2 specification, or 'UNK,' for unknown. Partial list: ENG = English SPA = Spanish UNK = Unknown	CHAR(3) NOT NULL DEFAULT 'UNK'
LANG_USAGE	How this person uses this language S = Spoken (or signed) W = Written B = Both spoken and written U = Unknown	CHAR(1) DEFAULT 'U'
LANG_PRIMARY	For spoken languages, is this the person's primary spoken language? Y = Yes N = No U = Unknown	CHAR(1) DEFAULT 'U'

Languages Data Quality Checks

- Is PERSON_ID ever missing?
- Are the most common primary languages (likely English and Spanish) representative of your patient population?
- Does the proportion of patients where language is unknown match what you would expect based on your knowledge of your EHR and workflow?

PHARMACY

The PHARMACY table is populated with dispenses from an outpatient pharmacy setting. This information can be extracted from pharmacy records or from claims. Records of substance administration from an inpatient setting are not populated in the pharmacy table. Dispenses may include initial dispenses or refills. CHORDS data partners that do not have a pharmacy or dispense medications may leave this table empty.

Pharmacy records should be unique on the combination of PERSON_ID, dispensing date (RXDATE), drug (NDC), and prescribing provider (RXMD). The file should be summed by dispensed amount if necessary to remove rollback transactions or combine multiple dispensing to the same person for the same drug by the same provider on the same day.

For any single record **both** RXSUP and RXAMT may not be negative or null values. If RXSUP or RXAMT are greater than or equal to 1, then it's a valid record. Examples:

Valid: RXSUP = 0 and RXAMT = 1

Valid: RXSUP = -1 and RXAMT = 1

Not Valid: RXSUP = 0 and RXAMT = 0

Not Valid: RXSUP = -1 and RXAMT = 0

Compound Drugs are represented differently in the data across sites. If the drug you are looking for may be a compound drug, please check with individual sites to how they are handled in their data.

Inclusions: Dispenses or medication fills.

Exclusions: Provider orders for medications and prescriptions are excluded. Information from the medication history is excluded from this table.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
RXDATE	Date of dispensing	DATE NOT NULL Recommended format MMDDYYYY
NDC	National Drug Code for dispense. More information about NDC here. Please expunge any place holders (ex. '-' or extra digit)	CHAR(11) NOT NULL
RXSUP	Days supply for the dispense See note above about negative and zero values	NUMERIC(4) NOT NULL
RXAMT	Amount dispensed. Number of units (pills, tablets) dispensed. Net amount per day per NDC. Data partners should decide how to represent compound drugs. See note above about negative and zero values	NUMERIC(16, 10) NOT NULL
RXMD	Optional. Provider who prescribed medication (can be an MD, NP, PA, etc.). RXMD will be used to join to other tables (PROVIDER and ENCOUNTER). Data partners should use same coding scheme as PROVIDERID in other tables.	CHAR(15)

Pharmacy Data Quality Checks

- Is PERSON_ID or RXDATE ever missing?
- Are any values for RXDATE in the future?
- Are any values for RXSUP and RXAMT both negative numbers or both zero?

EVERNDC

Each data partner will create its own EVERNDC (National Drug Code) table and update it when it updates its PHARMACY table. Each data partner-specific EVERNDC table will only contain NDCs used at that data partner during the data period.

For multi-data partner projects, a local list of NDCs from the lead data partner rarely captures all NDCs used at all participating HMORN data partners. A two-step approach that uses the EVERNDC table will help compile a complete list of all NDCs for the studied drugs. The lead analysts of specific projects will first write a distributed program to query the data partner-specific EVERNDC tables. They may use generic names, brand names, AHFS codes (if available), or GPI codes (if available) to identify the NDCs in use at other data partners. They will then collate the NDCs identified from all participating data partners to compile the final list of NDCs for their study. Compound Drugs are represented differently in the data across data partners. If the drug you are looking for may be a compound drug, please check with individual data partners to how they are handled in their data.

Number of dispensing and of members who used each drug during a specific period can be obtained by querying the VDW PHARMACY table using the NDCs identified from the two-step approach.

Inclusions: Every NDC code represented in the PHARMACY table should have one record in the EVERNDC table.

Exclusions: Available medications with no dispensing records in the PHARMACY table.

Fields

Name	Description	Definition
NDC	National Drug Code This is the HIPAA/CMS/NCPDP standard 5-4-2 configuration (no dashes).	CHAR(11) NOT NULL
GENERIC	Generic Name It is highly recommended that each data partner extracts only the generic name (e.g., "fluoxetine", and not "fluoxetine tab" or "fluoxetine tablet")	CHAR(105) NOT NULL
BRAND	Brand Name	CHAR(105) Note: Do not map brand names. Populate them only if they are present in the source system.

GPI	The Generic Product Identifier (GPI) is a 14-character hierarchical classification system that identifies drugs from their primary therapeutic use down to the unique interchangeable product regardless of manufacturer or package size. The code consists of seven subsets, each providing increasingly more specific information about a drug available with a prescription in the United States. The first six characters of the GPI define the therapeutic class code, the next two pairs the drug name, and the last four define route, dosage or strength. Generic Product Identifier, NOT Generic Price Indicator available in FDB	CHAR(14) Note: Do not map GPI codes. Populate them only if they are present.
AHFS1	AHFS Code Please add two zeroes at the end of your codes if they are in the 6-digit format.	CHAR(8)
AHFS2	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)
AHFS3	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)
AHFS4	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)
AHFS5	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)
AHFS6	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)
AHFS7	AHFS Code Leave this blank if there are no other AHFS codes	CHAR(8)

EverNDC Data Quality Checks

- Are any NDC codes present in the PHARMACY table that are not represented in the EVERNDC table?

PROCEDURES

This table lists all procedures performed from all inpatient and outpatient settings including lab, radiology and immunization procedure codes. A record is a procedure code/original procedure code/performing provider/procedure date combination unique to an index variable ENC_ID combination. Lab, radiology, immunization and other procedure codes may exist without a linked record in the ENCOUNTERS file or may be represented as a lab only or radiology only visit. Some encounters may have no associated procedures if procedural billing is not in place at data partner. Procedures are often extracted from billing or healthcare activity data. Require that the procedure file contain each type of procedure code when available: CPT, HCPCS, ICD9, Revenue Code. Procedures are attributed to the professional/inpatient rounding provider in the PERFORMINGPROVIDER variable and the procedure date in the PROCDATE variable. It is important to ensure that the CODETYPE variable is correct for each procedure value (PX).

Inclusions: Include all inpatient and ambulatory procedures that were performed.

Exclusions: Exclude procedures that were ordered but not performed. Exclude procedures where no date can be attributed.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
PROVIDER	See Keys section above Provider code for the provider who is most responsible for this encounter. Usually physician, nurse practitioner, physician assistant, optometrist, etc. For encounters with multiple providers, choose a single provider so the encounter can be linked to the diagnosis and procedure files. If there is no provider code for an encounter, then specify the value for provider as "UNK." Many data partners reserve a dummy provider ID assigned when the provider of services is unknown. Do not map that to a provider value in VDW; instead replace this with "UNK."	CHAR(15) NOT NULL
PROCDATE	Actual date when the procedure was done. If procedure date is unknown, populate with encounter date. If no date can be associated with procedure, exclude from table.	DATE NOT NULL Recommended format MMDDYYYY
ENC_ID	See Keys section above	CHAR(32) NOT NULL
PERFORMINGPROVIDER	Provider that performed the procedure. If this provider is unknown, set equal to the PROVIDER listed in ENCOUNTERS table based on ENC_ID.	CHAR(15) NOT NULL

ADATE	Encounter date or admit date for the encounter. Pull directly from ENCOUNTERS table.	DATE NOT NULL Recommended format MMDDYYYY
ENCTYPE	Encounter Type. See ENCOUNTERS table above. Pull directly from ENCOUNTERS table.	CHAR(2) NOT NULL
PX	Procedure Code. Convert local codes to standard codes if possible. Examples of procedure code formats: ##.## or ##.# ICD-9, $$$$$$ ICD-10, ##### or ####A CPT-4, A#### for HCPCS, ### or #### for Revenue Codes (match source data) Where: # = Numeric Digit, A = Alphabet Letter $ = Alpha or Numeric	CHAR(10) NOT NULL
ORIGPX	Original procedure code from source data. This is an optional field that can be used for mapping to standard codes.	CHAR(10) NOT NULL
PX_CODETYPE	Code type flag. **09** = ICD-9-CM **10** = ICD-10-CM **11** = ICD-11-CM **C4** = CPT-4 **H4** = HCPCS-4 **RV** = Revenue code **LO** = Local Home Grown **OT** = Other	CHAR(2) NOT NULL See SQL code to match procedure code types below
PXCNT	Number of times the procedure was performed during the encounter. The usual value is '1' unless the procedure was performed multiple times during the encounter.	NUMERIC(7) NOT NULL DEFAULT '1'
CPTMOD1	CPT Modifier Code 1 as found in the source data	CHAR(2)
CPTMOD2	CPT Modifier Code 2 as found in the source data	CHAR(2)
CPTMOD3	CPT Modifier Code 3 as found in the source data	CHAR(2)

Procedure Code Type Mapping

```
Case when len(px) >= 7  then '10'
      when px like '[0-9][0-9][0-9][0-9][0-9]' then 'C4'
      when px like '[0-9][0-9][0-9][0-9][A-Z]' then 'C4'
      when px like '[0-9][0-9].[0-9]%' then '09'
      when px like '[0-9][0-9][0-9].[0-9]%' then '09'
      when px like '[A-Z][0-9][0-9].[0-9]%' then '09'
                         when px substring(dw_px,1,1) like '[A-Z]%'
           and px substring(dw_px,2,4) like '[0-9][0-9][0-9][0-9]'
           and len(dw_px) = 5  then 'HC'
                         Else 'OT'
```

Procedures Data Quality Checks

- Is PERSON_ID, ADATE or ENC_ID ever null?
- Are any values for ADATE or PROCDATE in the future?
- Do all procedure codes have a PX_CODETYPE?

PROVIDER_SPECIALITY

One record per provider code present in the ENCOUNTERS, LAB_RESULT, PHARMACY, PROCEDURE, and DIAGNOSES tables.

Inclusions: Include provider information for any providers reflected in any other VDW tables.

Exclusions: No Exclusions.

Fields

Name	Description	Definition
PROVIDER	See Keys section above	CHAR(15) NOT NULL
SPECIALTY	Provider specialty code (see Specialty Code list below). If a provider has multiple specialties, populate main specialty in this field and use additional fields to represent other specialties. Most data partners will only have one specialty.	CHAR(3) NOT NULL DEFAULT 'UNK'
SPECIALTY2	Optional specialty code if provider has more than 1 specialty	CHAR(3)
SPECIALTY3	Optional specialty code if provider has more than 1 specialty	CHAR(3)
SPECIALTY4	Optional specialty code if provider has more than 1 specialty	CHAR(3)
PROVIDER_TYPE	Type of provider. See provider type codes below.	CHAR(3) NOT NULL DEFAULT '999'
PROVIDER_BIRTH_YEAR	Birth year of the provider	NUMERIC(4)

		Recommended format YYYY
PROVIDER_GENDER	Gender or sex of the provider M = Male F = Female U = Unknown O = Other (Transsexual, Transgendered, or anything else that does not fit into one of the prior categories)	CHAR(1) NOT NULL DEFAULT 'U'
PROVIDER_RACE	Race(s) of the provider HP = Native Hawaiian or Other Pacific Islander IN = American Indian/Alaska Native AS = Asian BA = Black or African American WH = White MU = More than one race, particular races unknown or not reported UN = Unknown or Not Reported	CHAR(2) NOT NULL DEFAULT 'UN'
PROVIDER_HISPANIC	Hispanic origin (ethnicity) of the provider. Y = Yes N = No U = Unknown	CHAR(1) NOT NULL DEFAULT 'U' This variable should only be populated if values for race are captured within the EHR. If ethnicity and race are captured in the same field within the EHR, values of Hispanic for race may be mapped to Hispanic = 'Y'.
YEAR_GRADUATED	Year provider graduated from medical/nursing/technical school	NUMERIC(4) Recommended format YYYY

Provider Type Codes

001 = Acupuncturist   003 = Audiologist
002 = Anesthesiology Assistant   004 = Audiology Assistant

005 = Cardiac Rehab Therapist
006 = Case Manager
007 = Certified Nurse Specialist
008 = Certified Registered Nurse Anesthetist
009 = Chaplain
010 = Chemical Dependency Counselor, Adult
011 = Chemical Dependency Counselor, Child/Adolescent
012 = Chiropractor
013 = Cigna
014 = Clinical Nurse Specialist
015 = Coumadin Nurse
016 = CPAP Specialist
017 = Dental Assistant
018 = Dentist
019 = Diabetic Nurse
020 = Dietetic Technician
021 = Dietitian
022 = EEG Technician
023 = ER Technician
024 = Exercise Physiologist
025 = Fellow
026 = Healing Touch
027 = Health Educator
028 = Home Health Aide
029 = HP Physician
030 = Hygienist
031 = Independent Lab/X-Ray
032 = Infection Control Practitioner
033 = Infusion Therapist
034 = Interpreter
035 = Lab Assistant
036 = Lab Technician
037 = Lactation Consultant
038 = Licensed Family Therapist
039 = Licensed Social Worker
040 = Massage Therapist
041 = Medical Assistant
042 = Midlevel
043 = Midwife
044 = Music Therapist
045 = Not Applicable
046 = Nurse

047 = Nurse Clinician
048 = Nurse Practitioner
049 = Occupational Therapist
050 = Ophthalmic Assistant
051 = Ophthalmic Technician
052 = Optician
053 = Optometrist
054 = Orthopaedic Physician Assistant
055 = Orthopaedic
056 = Osteopath
057 = Personal Care Attendant
058 = Pharmacy Technician
059 = Pharmacist
060 = Physical Therapist
061 = Physician
062 = Physician Assistant
063 = Podiatrist
064 = Psychology Technician
065 = Psychologist
066 = Psychotherapist
067 = Psychotherapist, Adult
068 = Psychotherapist, Child/Adolescent
069 = Physical Therapist Assistant
070 = Radiation Therapist
071 = Radiology Technician
072 = Recreational Therapist
073 = Registered Nurse
074 = Rehabilitation Assistant
075 = Rehabilitation Therapist
076 = Reiki
077 = Resident
078 = Resource
079 = Respiratory Therapist
080 = Social Worker
081 = Speech Pathologist Certified
082 = Student
083 = Surgical Technician
084 = Surgeon Assistant
085 = Therapeutic Aquatic Specialist
086 = Ultrasound Technician
888 = Other
999 = Unknown

Provider Specialty Codes

SPECIALTY	Description	SPECIALTY	Description
ADO	Adolescent Medicine	NUM	Nuclear Medicine
AER	Aerospace Medicine	NUR	Nurse
ALC	Chemical Dependency	NUT	Nutrition
ALL	Allergy	OBO	Gynecologic Oncology
ANC	Ancillary Services	OBS	Obstetrics - Gynecology
ANE	Anesthesiology	OCM	Occupational Health
ATH	Sports Medicine	ONC	Oncology
AUD	Audiology	ONS	Surgical Oncology
BON	Bone And Mineral	OPH	Ophthalmology
CAR	Cardiology	OPL	Optical
CAV	Cardiovascular Surgery	OPT	Optometry
CHR	Chiropractor	ORA	Oral Surgery
CLC	Clinical Cardiac Electrophysiology	ORD	Orthodontia
COL	Colon & Rectal Surgery	ORT	Orthopedics
COM	Complimentary & Alternative Medicine	OTO	Otolaryngology
CON	Continuing Care	PAI	Pain Management
CRI	Critical Care	PAT	Pathology
DEN	Dentistry	PED	Pediatrics
DER	Dermatology	PES	Pediatric Subspecialty
DOR	D.O.R.	PEY	Perinatology
EDU	Medical Education	PHA	Pharmacy
EME	Emergency Medicine	PHT	Physical Therapy
EMI	EMI	PHY	Physiatry
END	Endocrinology	PLA	Plastic Surgery
ENT	Otolaryngology	POD	Podiatry
FAM	Family Medicine	PRE	Preventive Medicine
FLX	Flexible	PRO	Prosthodontia
GAS	Gastroenterology	PSY	Psychiatry
GEN	Genetics	PUB	Public Health
GER	Gerontology	PUL	Pulmonary Medicine
HAN	Hand Surgery	RAD	Radiology
HOM	Home Health	REH	Rehabilitation Medicine
HOS	Hospital	RES	Respiratory Therapy
HYM	Hyperbaric Medicine	RHE	Rheumatology
HYP	Hypertension	ROP	Radiation Oncology
IMG	General Internal Medicine	SLC	Sleep Center
INF	Infectious Disease	SOC	Social Services

LAB	Laboratory	SPP	Speech Pathology
MEN	Mental Health	SUR	Surgery
MGM	Care Management	TEE	Teen Clinic
MID	Midlevel	TOX	Medical Toxicology
MUL	Multispecialty	TRN	Transportation/Non-Emergency
NEH	Nephrology	TRS	Transplant Surgery
NEO	Neonatology	UNK	Unknown
NES	Neurosurgery	URG	Urgent Care
NEU	Neurology	URO	Urology
NOB	No Boards	VAS	Vascular Surgery

Provider Specialty Data Quality Checks

- Are all provider IDs from LAB_RESULTS, ENCOUNTER, PROCEDURES, PHARMACY and DIAGNOSES present in the PROVIDER_SPECIALTY table.
- Are any PROVIDER_BIRTH_YEAR or YEAR_GRADUATED values in the future?

SOCIAL_HISTORY

The SOCIAL_HISTORY table includes various behavioral measures taken by health professionals, either during the clinic visit or often over the telephone or via questionnaires. These measures may include, but are not limited to, the use of tobacco, alcohol and illegal drugs, as well as sexual behavior and contraceptive use. Data partners should populate the SOCIAL_HISTORY table only with what information is captured and stored electronically in structured fields. CHORDS does not expect data partners to extract information from scanned forms or unstructured text for structured fields. As a result, much of this table may remain empty for some data partners.

Many fields in SOCIAL_HISTORY include a response option 'X' for Not Asked. Data partners should use this field only if this data can be collected in the source system and some clinical settings where screening doesn't occur can be clearly identified. If screening does not occur system wide, these variables can be marked "Unknown."

The responses captured for tobacco use vary widely across EHR platform. Data partners should map tobacco use as best as possible to the allowed values. Tobacco use relates to both smoking and smokeless tobacco and additional variables are used to differentiate between the type of tobacco used. In some instances, it is possible to differentiate smoking tobacco from smokeless tobacco use. While these individual variables may offer greater detail, they may also conflict with the overall "TOBACCO_USE" data. Map and use tobacco variables with caution.

Regarding substance use, data partners may interpret the terms 'illicit drugs' and 'IV drugs' to the best of their ability. Substance use screening may vary widely across providers. Regarding sexual health and family planning, some systems collect sex partner as one or two variables. Thus some data partners may be able to capture if a patient reports sex partners of both gender. Some EHR systems collect one or multiple contraceptive methods. Data partners are encouraged to represent as much information as possible through the use of multiple fields.

There is no unique primary key for the SOCIAL_HISTORY table. The PERSON_ID, ENC_ID and CONTACT_DATE are the only variables required to be non-null for inclusion in the table.

Inclusions: Include encounter where information about a patient's tobacco use, sexual health, or substance use is collected and stored.

Exclusions: Exclude encounter where information about a patient's tobacco use, sexual health, or substance use is not collected and stored.

Fields

Name	Description	Definition
ENC_ID	See Keys section above	CHAR(32) NOT NULL
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
CONTACT_DATE	Date of visit or encounter where assessment of behavior was collected.	DATE NOT NULL Recommended format MMDDYYYY
TOBACCO_USER	Tobacco use status as of the visit date. Responses are hierarchical based on the level of risk and are not mutually exclusive. For example, a current smoker may also be exposed to second hand smoke. That person would receive a 'Y'. The acceptable values are included below in hierarchical order. Y = Yes/Current (*everyday tobacco use*) I = Infrequent (*current someday tobacco use*) Q = Quit/Former P = Passive/Environmental/Second-hand (*never smoked*) N = Never (*never smoked or used any kind of tobacco and not SHS exposed*) X = Not Asked U = Unknown, Missing (*no information about tobacco use collected*) Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
ONC_SMOKING_STATUS	This is an additional tobacco use variable that captures tobacco use based on a national standard for tobacco use categories. 1 = Current every day smoker 2 = Current some day smoker 3 = Former smoker 4 = Never smoker 5 = Smoker, current status unknown 6 = Unknown if ever smoked	CHAR(1) DEFAULT '6'
ALCOHOL_USER	Is the patient an alcohol drinker? Y = Yes/Current (*everyday use*)	CHAR(1) NOT NULL DEFAULT 'U'

	Q = Quit/Former *(former user)* N = Never *(never used alcohol)* X = Not Asked U = Unknown, Missing *(no information about alcohol use collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	
ILL_DRUG_USER	Is the patient an illicit drug user? Y = Yes/Current *(everyday use)* Q = Quit/Former *(former user)* N = Never *(never used illicit drugs)* X = Not Asked U = Unknown, Missing *(no information about illicit drug use collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
TOBACCO_PAK_PER_DY	Number of packs currently or historically smoked per day. Text variable allows non-numeric responses. Should be populated only for patients where TOBACCO_USER = Y, I, or Q	CHAR(20) DEFAULT NULL
TOBACCO_USED_YEARS	Number of packs currently or historically smoked per day. Text variable allows non-numeric responses. Should be populated only for patients where TOBACCO_USER = Y, I, or Q	CHAR(20) DEFAULT NULL
SMOKING_QUIT_DATE	Date of quit smoking, Should be populated only for patients where TOBACCO_USER = Y, I, or Q	DATETIME DEFAULT NULL Recommended format MMDDYYYY
SMOKELESS_QUIT_DATE	Date of quit smokeless Should be populated only for patients where TOBACCO_USER = Y, I, or Q	DATETIME DEFAULT NULL Recommended format MMDDYYYY
TOBACCO_COMMENT	Comment about Tobacco Use.	CHAR(255) DEFAULT NULL
CIGARETTES_YN	Patient reports cigarette use	CHAR(1) NOT NULL DEFAULT 'U'

	Y = Yes patient reports using this type of tobacco N = No, patient reports tobacco use but not this type X = Not Asked U = Unknown, missing because patient does not use tobacco or the tobacco type used was not recorded.	Complimentary variable to TOBACCO_USER to assess type of tobacco product used. Y and N responses valid for TOBACCO_USER = Y, I, or Q.
PIPES_YN	Patient reports tobacco pipe use Y = Yes patient reports using this type of tobacco N = No, patient reports tobacco use but not this type X = Not Asked U = Unknown, missing because patient does not use tobacco or the tobacco type used was not recorded.	CHAR(1) NOT NULL DEFAULT 'U' Complimentary variable to TOBACCO_USER to assess type of tobacco product used. Y and N responses valid for TOBACCO_USER = Y, I, or Q.
CIGARS_YN	Patient reports tobacco cigar use Y = Yes patient reports using this type of tobacco N = No, patient reports tobacco use but not this type X = Not Asked U = Unknown, missing because patient does not use tobacco or the tobacco type used was not recorded.	CHAR(1) NOT NULL DEFAULT 'U' Complimentary variable to TOBACCO_USER to assess type of tobacco product used. Y and N responses valid for TOBACCO_USER = Y, I, or Q.
SNUFF_YN	Patient reports tobacco snuff use Y = Yes patient reports using this type of tobacco N = No, patient reports tobacco use but not this type X = Not Asked U = Unknown, missing because patient does not use tobacco or the tobacco type used was not recorded.	CHAR(1) NOT NULL DEFAULT 'U' Complimentary variable to TOBACCO_USER to assess type of tobacco product used. Y and N responses valid for TOBACCO_USER = Y, I, or Q.
CHEW_YN	Patient reports tobacco chew use	CHAR(1) NOT NULL

	Y = Yes patient reports using this type of tobacco N = No, patient reports tobacco use but not this type X = Not Asked U = Unknown, missing because patient does not use tobacco or the tobacco type used was not recorded.	DEFAULT 'U' Complimentary variable to TOBACCO_USER to assess type of tobacco product used. Y and N responses valid for TOBACCO_USER = Y, I, or Q.
UNSPECIFIED_TOBACCO	Tobacco type used is not specified, or is of a type not specified in the TYPE_YN variable options listed (e.g. CIGARETTES_YN). Y = Yes, patient uses tobacco and type is unknown or is a type not listed by a TYPE_YN variable NULL = Patient uses a tobacco type listed above or patient does not use tobacco	CHAR(1) DEFAULT NULL
ALCOHOL_OZ_PER_WK	Number of ounces alcohol consumed per week Text variable allows non-numeric responses.	CHAR(15)
ALCOHOL_COMMENT	Comment about alcohol use	CHAR(255)
IV_DRUG_USER_YN	Is the patient an IV drug user? Y = Yes/Current (*everyday use*) Q = Quit/Former (*former user*) N = Never (*never used illicit drugs*) X = Not Asked U = Unknown, Missing *(no information about illicit drug use collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
ILLICIT_DRUG_FREQ	Frequency of illicit drug use. Text variable allows non-numeric responses.	CHAR(5)
ILLICIT_DRUG_COMMENT	Comment about illicit drug use	CHAR(255)
FEMALE_PARTNER_YN	Patient reports a female sex partner Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'

MALE_PARTNER_YN	Patient reports a male sex partner Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
CONDOM_YN	Patient reports condom use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
PILL_YN	Patient reports birth control pill use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
DIAPHRAGM_YN	Patient reports diaphragm use Y = Yes N = No X = Not Asked (*only use if source system identifies clinical settings where screening doesn't occur – do no impute this value)* U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
IUD_YN	Patient reports IUD use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'

SURGICAL_YN	Patient reports surgical birth control Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
SPERMICIDE_YN	Patient reports spermicide use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
IMPLANT_YN	Patient reports birth control implant use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
RHYTHM_YN	Patient reports rhythm method use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
INJECTION_YN	Patient reports birth control injection use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
SPONGE_YN	Patient reports birth control sponge use	CHAR(1) NOT NULL DEFAULT 'U'

	Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	
BC_INSERTS_YN	Patient reports birth control insert use Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
ABSTINENCE_YN	Patient reports birth control via abstinence Y = Yes N = No X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'
SEX_COMMENT	Comment on sexual activity	CHAR(255)
YEARS_EDUCATION	Number of years of education completed Text variable allows non-numeric responses.	CHAR(15)
SEXUALLY_ACTV	Is the patient sexually active? Y = Yes N = No W= Not Currently (Was) X = Not Asked U = Unknown, Missing *(no information collected)* Note: An answer of 'X' and 'U' differentiates who was asked the question by their provider.	CHAR(1) NOT NULL DEFAULT 'U'

Social History Data Quality Checks
- Is PERSON_ID or ENC_ID ever null?
- Is CONTACT_DATE in the future?

==INFORMATION PAST THIS POINT IS SUBJECT TO CHANGE==

Priority 3 Tables

ENROLLMENT

Each individual in any VDW table has one record in the demographics table and <u>at least</u> one record in the enrollment table. The VDW enrollment table can serve as the patient index.

The Enrollment table was designed to use health plan member information to track start and end dates of active coverage and coverage type. For a CHORDS data partner that is a health plan or payer, a member's dates of active coverage are represented by enrollment start and end date and a member's type of insurance coverage (or lack of insurance) during that time is described by the insurance variables. Each record in the enrollment table represents one enrollment span. When coverage ends or coverage type changes, a new enrollment span was generated.

For a non-payer healthcare provider, enrollment start and end dates can be populated to represent periods of time where the individual is considered an <u>active patient</u> (i.e. has been seen in the last 18 months). For these data partners, enrollment_basis should be marked with a 'P'.   Using this definition, Individuals with multiple records in the enrollment table may have multiple spans of time when they were an active patient <u>or</u> multiple payers during a period of time when they were being seen. If data partners do not have insurance information on a patients payer, that is ok, and a data partner can leave insurance variables empty if that information is unknown.

Enrollment spans may not overlap one another--there should be one and only one record covering any given day and individual. No enrollment end dates can occur in the future. If the last encounter in the enrollment span is within 18 months of the enrollment table refresh date, then the enrollment end date is the date of the last VDW load.

Example Enrollment Logic from a non-payer provider:

- The enrollment start date is populated at January 1 of the calendar year of the first encounter in that active patient span.
- The enrollment end date is populated as 18 months (540 days) following the last encounter in the enrollment span.
- If an individual has a series of encounters between 2007 and 2016 that are never separated by more than 18 months, they would have one long enrollment span:
  - Enr_start= 1/1/2007
  - Enr_end= refresh date
- If an individual has 2 encounters: 2/1/2009 and 3/1/2016, then the individual will have two enrollment spans:
  - Enrollment span 1:
    - Enr_start= 1/1/2009
    - Enr_end= 7/30/2010
  - Enrollment span 2:
    - Enr_start= 1/1/2016
    - Enr_end= refresh date

For enrollment spans in the past, insurance information from the last encounter will be used to populate the insurance fields. For enrollment spans that are current (where the table refresh date is the enr_end), insurance information from the most recent visit will be used to populate the insurance fields.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
ENR_START	Beginning date of the enrollment period.	DATE NOT NULL
ENR_END	End date of the enrollment period.	DATE NOT NULL
INS_MEDICAID	Flag indicating whether the person had any Medicaid coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_COMMERCIAL	Flag indicating whether the person had any Commercial coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_PRIVATEPAY	Flag indicating whether the person had Private Pay coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_STATESUBSIDIZED	Flag indicating whether the person had any State Subsidized coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_SELFFUNDED	Flag indicating whether the person had any coverage through an Employer group that insures itself. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_HIGHDEDUCTIBLE	Flag indicating whether the person had any coverage in a high deductible plan. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_MEDICARE	Flag indicating whether the person had any Medicare coverage, including Medicare Working Aged Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_MEDICARE_A	Flag indicating the person had Medicare Part A coverage during the period.	CHAR(1) NOT NULL

	Y - Yes N - No U - Unknown	
INS_MEDICARE_B	Flag indicating the person had Medicare Part B coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_MEDICARE_C	Flag indicating the person had Medicare Part C coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_MEDICARE_D	Flag indicating the person had Medicare Part D coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
INS_OTHER	Flag indicating whether the person had an Other type of coverage during the period. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
PLAN_HMO	Had at least some coverage under an HMO plan Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
PLAN_POS	Had at least some coverage under a point-of-service plan Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
PLAN_PPO	Had at least some coverage under a preferred provider organization plan Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
PLAN_INDEMNITY	Had at least some coverage under a traditional indemnity plan. Y - Yes N - No U - Unknown	CHAR(1) NOT NULL
DRUGCOV	Flag indicating whether the insurance coverage included at least some payment/coverage for prescription drugs. Y - Yes N - No	CHAR(1) NOT NULL

	U - Unknown	
OUTSIDE_UTILIZATION	Is this person/period likely to have substantial health care utilization outside the HMO system, under circumstances that would not result in a claim being submitted to the HMO? Y - Yes, we expect the utilization capture for this person/period is incomplete. N - No, there is no reason to suspect we have incomplete capture of utilization. U - Completeness of capture is unknown.	CHAR(1) NOT NULL
ENROLLMENT_BASIS	The basis for the claim (if we are making one—see outside_utilization above) that we have something like complete capture of the medical record in the VDW. G - Geographic Basis I - Insurance Basis B - Both Insurance and Geographic bases P - Non-enrollee Patient	CHAR(1) NOT NULL
PCC	The clinic to which the patient is paneled in administrative record. Same as utilization.facility_code	CHAR(4)
PCP	The clinician to which the patient is paneled in administrative data. Same as utilization.provider.	CHAR(15)

Priority 4 Tables

CAUSE_OF_DEATH

CHORDS data partners will not need to populate the Cause of Death table at this time.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
COD	Cause of death, A valid ICD diagnosis code (Please include decimal points in your ICD codes)	Char(6) NOT NULL
DX_CODETYPE	Refer to the Dx_Codetype variable on the Diagnosis file for type, length, and value set	Char(2)
CAUSETYPE	The type of cause of death. I - Immediate/Primary U - Underlying C - Contributory O - Other	Char(1)
SOURCE	Source of the information on this record S – State Death files	Char(1) NOT NULL

	N – National Death Index	
	T – Tumor data	
	B – Social Security Admin	
	Others as locally defined	
CONFIDENCE	How you rate the accuracy of the observation based on source, match, # of reporting sources, discrepancies, etc. E - Excellent F - Fair P - Poor	Char(1) NOT NULL

DEATH

CHORDS data partners will not need to populate the Death table at this time.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
DEATHDT	Date of death	DATE
DTIMPUTE	How DeathDt was imputed M - Month of date imputed D - Day of date imputed B – Both month and day imputed N – Not imputed	Char(1)
SOURCE	Source of the information on this record S – State Death files N – National Death Index T – Tumor data B – Social Security Admin Others as locally defined	Char(1) NOT NULL
CONFIDENCE	How you rate the accuracy of the observation based on source, match, # of reporting sources, discrepancies, etc. E - Excellent F - Fair P - Poor	Char(1) NOT NULL

LAB_NOTES

The second table, LAB_NOTES, links to LAB_RESULTS by the variable ROW_ID. Not all rows in the LAB_RESULTS data set will have corresponding records in LAB_NOTES. Only a record which requires more than 8 bytes character storage for results, normal low or normal high or a record that has notes associated with it will have one or more corresponding observations in this data set.

The LAB_NOTES table includes unstructured text associated with a lab test result. CHORDS data partners will not be expected to populate lab notes.

Fields

Name	Description	Definition
ROW_ID	Data partner identifier that links a LAB NOTES record to the corresponding LAB RESULTS record. It is possible for more than one record in LAB NOTES to link back to the same record in LAB RESULTS. Therefore it is possible for multiple records in LAB NOTES to have the same ROW_ID value.	Type and length are data partner specific
RESULT_NOTE	This field could contain: result, lowest value of normal range, highest value of normal range, notes. To accept text string results, comments or normal low or high range values that are longer than 8 bytes	CHAR(80)
NOTE_TYPE	Describes what information is held in the RESULT_NOTE field. R = Result L = Low Normal Value H = High Normal Value N = note(other type values can be added as need arises)	CHAR(1)
LINE	Counter starting at 1 and incrementing by 1 as needed for each ROW_ID by Type. As many rows as necessary can be associated with a ROW_ID/NOTE_TYPE combination. For example, a ROW_ID might have 5 records of Notes associated with it. The NOTE_TYPE value would = N and the line value would be 1 for the first note, 2 for the second note and so on through 5.	NUM(4)

TUMOR

Tumor contains records documenting neoplasms (nearly always malignant) diagnosed in patients or enrollees of the health plan. There is one record per separately diagnosed neoplasm. CHORDS data partners will not need to populate the Tumor table at this time.

Fields

Name	Description	Definition
PERSON_ID	See Keys section above	CHAR(12) NOT NULL
DXDATE	Date of cancer diagnosis	DATE NOT NULL
ICDODATA PARTNER	Data partner of the tumor (e.g. breast, prostate)	CHAR(4) NOT NULL
STAGEGEN	General Stage	CHAR(1) NOT NULL
SS1977	General Stage, per the 1977 SEER summary staging guidelines.	CHAR(1)
SS2000	General Stage, per the 2000 SEER summary staging guidelines.	CHAR(1)
STAGEAJ	AJCC summary stage or "best AJCC stage"	CHAR(4) NOT NULL

AJCC_Ed	AJCC Staging Scheme Edition	CHAR(1)
AJCC_GRP	Best available AJCC stage. one digit is allowed	CHAR(2)
AJCC_FULL_DER	Expanded staging elements including neoadjuvant flags	CHAR(20)
MORPH	Morphology/histology (tissue type of cancer)	CHAR(4) NOT NULL
BEHAV	Behavior	CHAR(1) NOT NULL
GRADE	Histologic grading and differentiation. Describes tumor's resemblance to normal tissue.	CHAR(1) NOT NULL
DXYEAR	The year the tumor was diagnosed.	NUMERIC(4) NOT NULL
DXAGE	The patient's age on the date the tumor was diagnosed.	NUMERIC(3) NOT NULL
BDATE	Birth Date	DATE NOT NULL
GENDER	The patient's Sex or Gender	CHAR(1) NOT NULL
RACE1	Race(s) of the person. HP – Native Hawaiian or Other Pacific Islander IN – American Indian/Alaska Native AS - Asian BA - Black or African American WH - White MU - More than one race, particular races unknown or not reported UN - Unknown or Not Reported	CHAR (2) NOT NULL
RACE2	Race(s) of the person. HP – Native Hawaiian or Other Pacific Islander IN – American Indian/Alaska Native AS - Asian BA - Black or African American WH - White MU - More than one race, particular races unknown or not reported UN - Unknown or Not Reported	CHAR (2) NOT NULL
RACE3	Race(s) of the person. HP – Native Hawaiian or Other Pacific Islander IN – American Indian/Alaska Native AS - Asian BA - Black or African American WH - White MU - More than one race, particular races unknown or not reported UN - Unknown or Not Reported	CHAR (2) NOT NULL

RACE4	Race(s) of the person. HP – Native Hawaiian or Other Pacific Islander IN – American Indian/Alaska Native AS - Asian BA - Black or African American WH - White MU - More than one race, particular races unknown or not reported UN - Unknown or Not Reported	CHAR (2) NOT NULL
RACE5	Race(s) of the person. HP – Native Hawaiian or Other Pacific Islander IN – American Indian/Alaska Native AS - Asian BA - Black or African American WH - White MU - More than one race, particular races unknown or not reported UN - Unknown or Not Reported	CHAR (2) NOT NULL
RACE1	SEER race classification	CHAR(2) NOT NULL
RACE2	Added to record multiple races.	CHAR(2) NOT NULL
RACE3	Added to record multiple races.	CHAR(2) NOT NULL
RACE4	Added to record multiple races.	CHAR(2) NOT NULL
RACE5	Added to record multiple races.	CHAR(2) NOT NULL
HISPANIC	Hispanic or Spanish Origin/Ethnicity	CHAR(1) NOT NULL
CLASS	Class of case (analytic vs. non-analytic)	CHAR(2) NOT NULL
SEQUENCE	Tumor sequence (order in which patients with multiple cancers are diagnosed)	CHAR(2) NOT NULL
VITAL	Vital status at last contact	CHAR(1) NOT NULL
DCAUSE	Cause of death for expired patients	CHAR(6)
DOD	Date of death	DATE
DT_FU	Date of last contact or death	DATE
LATERALITY	Laterality of primary data partner for paired organs (e.g. breast)	CHAR(1) NOT NULL
IDPLAN	Implementing data partner ID code	CHAR(2) NOT NULL
DCNFRM	Diagnostic confirmation -method used to diagnosis cancer (this is hierarchical)	CHAR(1) NOT NULL
DSTZ	Tumor size – size of largest dimension of primary lesion in centimeters (except for skin melanoma, which records depth of lesion in *centimeters*).	CHAR(3)
DAJC1T_P	T - AJCC pathological staging for tumor size.	CHAR(5)
DAJC1N_P	N – AJCC pathological staging for lymph nodes.	CHAR(5)
DAJC1M_P	N – AJCC pathological staging for metastisis.	CHAR(5)
DAJC1T_C	T - AJCC clinical staging for tumor size.	CHAR(5)

DAJC1N_C	N – AJCC clinical staging for lymph nodes.	CHAR(5)
DAJC1M_C	N – AJCC clinical staging for metastisis.	CHAR(5)
DSRG_FAC	1st course of surgical treatment at this facility – see primary data partner-specific coding from SEER.	CHAR(2) NOT NULL
DRAD_FAC	1st course of radiation treatment (this facility) – refers to modality	CHAR(2) NOT NULL
DCHM_FAC	1st course of chemotherapy at this facility	CHAR(2) NOT NULL
DHRM_FAC	1st course of hormone therapy at this facility	CHAR(2) NOT NULL
DIMM_FAC	1st course of immunotherapy (biological response therapy) at this facility	CHAR(2) NOT NULL
DOTH_FAC	1st course of any other treatment at this facility	CHAR(2) NOT NULL
DNDI	Regional lymph nodes positive	CHAR(2) NOT NULL
DNDX	Regional lymph nodes examined	CHAR(2) NOT NULL
DTMRK1	Tumor Marker One	CHAR(1)
DTMRK2	Tumor Marker Two	CHAR(1)
DTMRK3	Tumor Marker Three	CHAR(1)
CLN_STG	TNM Clin Stage Group (if available)	CHAR (4) NOT NULL
EOD	Extent of Disease	CHAR(12)
DT_SURG	Date Definitive Surgery First Performed	DATE
DT_CHEMO	Date chemotherapy started	DATE
DT_HORM	Date hormone therapy started	DATE
DT_RAD	Date radiation therapy started	DATE
DT_BRM	Date immunotherapy started	DATE
DT_OTH	Date other therapy started	DATE
R_N_SURG	Reason no surgery was performed on the primary data partner	CHAR(1) NOT NULL
R_N_CHEMO	Reason no chemotherapy started	CHAR(2) NOT NULL
R_N_HORM	Reason no hormone therapy started	CHAR(1) NOT NULL
R_N_RAD	Reason no radiation therapy started	CHAR(1) NOT NULL
R_N_BRM	Reason no immunotherapy started	CHAR(1) NOT NULL
R_N_OTH	Reason no other therapy started	CHAR(1) NOT NULL
DSRG_SUM	1st course of surgical treatment at this (any facility)	CHAR(2) NOT NULL
DRAD_SUM	1 ST course of radiation treatment (any facility)	CHAR(2) NOT NULL
DCHM_SUM	1st course of chemotherapy at any facility	CHAR(2) NOT NULL
DHRM_SUM	1st course of hormone therapy at any facility	CHAR(2) NOT NULL
DIMM_SUM	1st course of immunotherapy at any facility	CHAR(2) NOT NULL
DOTH_SUM	1st course of any other treatment at any facility	CHAR(2) NOT NULL
CS_SZ	Collaborative Staging Tumor Size	CHAR(3)
CS_EXT	Collaborative Staging Extent of Disease	CHAR(3)
CS_NODES	Collaborative Staging Lymph Nodes	CHAR(3)

CS_NODES_EVAL	Collaborative staging Lymph Nodes evaluation method.	CHAR(1)
CS_METS	Collaborative Staging Metastases	CHAR(2)
CS_METS_EVAL	Collaborative staging method of metastasis evaluation.	CHAR(1)
SSF1	Tumor Data partner-specific factors.	CHAR(3)
SSF2	Tumor Data partner-specific factors.	CHAR(3)
SSF3	Tumor Data partner-specific factors.	CHAR(3)
SSF4	Tumor Data partner-specific factors.	CHAR(3)
SSF5	Tumor Data partner-specific factors.	CHAR(3)
SSF6	Tumor Data partner-specific factors.	CHAR(3)
SSF7	Tumor Data partner-specific factors.	CHAR(3)
SSF8	Tumor Data partner-specific factors.	CHAR(3)
SSF9	Tumor Data partner-specific factors.	CHAR(3)
SSF10	Tumor Data partner-specific factors.	CHAR(3)
SSF11	Tumor Data partner-specific factors.	CHAR(3)
SSF12	Tumor Data partner-specific factors.	CHAR(3)
SSF13	Tumor Data partner-specific factors.	CHAR(3)
SSF14	Tumor Data partner-specific factors.	CHAR(3)
SSF15	Tumor Data partner-specific factors.	CHAR(3)
SSF16	Tumor Data partner-specific factors.	CHAR(3)
SSF17	Tumor Data partner-specific factors.	CHAR(3)
SSF18	Tumor Data partner-specific factors.	CHAR(3)
SSF19	Tumor Data partner-specific factors.	CHAR(3)
SSF20	Tumor Data partner-specific factors.	CHAR(3)
SSF21	Tumor Data partner-specific factors.	CHAR(3)
SSF22	Tumor Data partner-specific factors.	CHAR(3)
SSF23	Tumor Data partner-specific factors.	CHAR(3)
SSF24	Tumor Data partner-specific factors.	CHAR(3)
SSF25	Tumor Data partner-specific factors.	CHAR(3)
PAL_FAC	Palliative care at this facility	CHAR(1) NOT NULL
PAL_SUM	Palliative care at all facilities	CHAR(1) NOT NULL
DER_T6	Derived AJCC6 T	CHAR(3)
DER_T6_D	Derived AJCC6 T description	CHAR(1)
DER_N6	Derived AJCC6 N	CHAR(3)
DER_N6_D	Derived AJCC6 N description	CHAR(1)
DER_M6	Derived AJCC6 M	CHAR(3)
DER_M6_D	Derived AJCC6 M description	CHAR(1)
DER_T7	Derived AJCC7 T	CHAR(3)
DER_T7_D	Derived AJCC7 T description	CHAR(1)
DER_N7	Derived AJCC7 N	CHAR(3)
DER_N7_D	Derived AJCC7 N description	CHAR(1)
DER_M7	Derived AJCC7 M	CHAR(3)
DER_M7_D	Derived AJCC7 M description	CHAR(1)
DER_SS2000F	Describes summary stage 2000	CHAR(1)
RECUR_DT	Recurrence date	DATETIME

RECUR_TYPE	Type of 1st recurrence	CHAR(2) NOT NULL
RECUR_FL	1st recurrence date flag	CHAR(2)
DATA_SOURCE	Compound code to indicate source of tumor data	CHAR(4)

Appendix 1: SQL DDL

Appendix 2: Instructions for Connecting to CDPHE's FTP data partner

CDPHE CHORDs Project FTP Data partner: Secft2.dphe.state.co.us\CHORDS_tobacco_[agency name]

Via FTP Client Software: There are many FTP clients available for you to install. Which you choose is up to your organization. To have full functionality with CDPHE's FTP data partner, the FTP client you install must support SFTP (Microsoft's command line FTP client is not supported).

The following settings are common on the clients:

Host/IP/URL:	secft2.dphe.state.co.us or 165.127.8.22
Username and password:	Active Directory username and password assigned by CDPHE ITS
Port:	22
PASV	yes
Connection:	SSH/SFTP

Via Web Browser: Instead of installing FTP client software, you can access CDPHE's FTP data partner through your web browser.

1. Open a web browser (Internet Explorer and Firefox have been successfully tested).

2. In the address/URL/location field enter: https://secft2.dphe.state.co.us/thinclient/login.aspx

3. You will receive a certificate error due to:
    a. The webdata partner name, secft2.dphe.state.co.us, does not match the name on the certificate, dphevs21.
    b. The certificate for the secft2.dphe.state.co.us data partner is an internal certificate (meaning it is not from a large public certification authority such as Verisign or Entrust and was created from within the application or server) and the certificate store on the local system most likely does not have the root certificate for CDPHE certificates.

4. Even though a certificate error appears, it is safe to proceed into the FTP data partner. Verify that the certificate:
    a. Firefox:
        i. Expand technical details and you should see:
            The certificate is not trusted because it is self-signed.
            The certificate is only valid for DPHEVS21.
        ii. Expand "I Understand the Risks"
        iii. Click the "Add Exception…" button
        iv. Click the "Confirm Security Exception" button

    b. Internet Explorer (IE):
        i. Click "Continue to this webdata partner (not recommended.)"
        ii. Next to the URL you should see a box that states "Certificate Error"
        iii. Click on "Certificate Error" and you should see a message that there is a "Mismatched Address"
        iv. Select "View Certificates"
            The "Issued to:" field should be dphevs21
            The "Issued by:" field should be dphevs21

If there are still concerns, please contact the "Hosted Services Team" at CDPHE ITS (303-692-2222) and a technician can verify the thumbprint of the certificate.

5. In the username field enter the username, assigned by CDPHE ITS, followed by "\secft2.dphe.state.co.us" (without the quotes, see image).

6. In the password field enter the password assigned to your username. If you do not know your password, contact your program liaison.

7. Click on the "Login" button
   a. If you are unable to successfully log in, contact your program liaison.
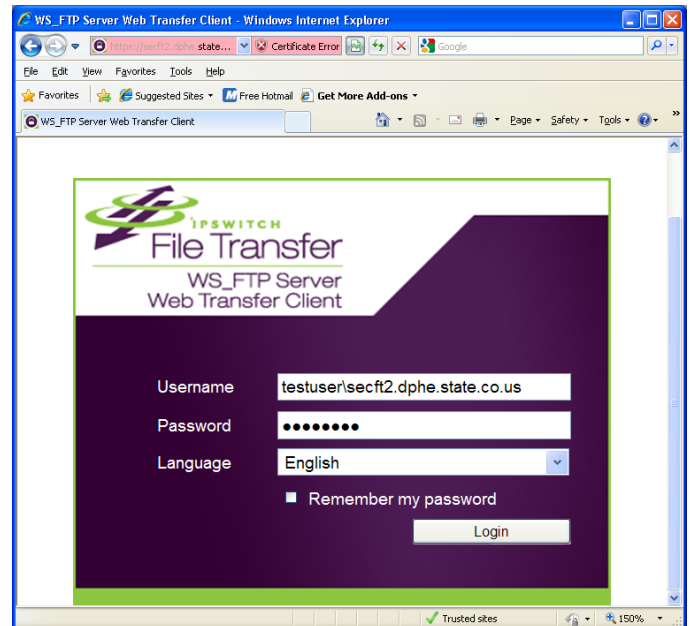
What happens when Addresses do not match or a patient is homeless?

Geocoding typically requires a valid address (at least valid within the available zips, street names, and numeric ranges possible for these streets). If in the data an address is provided for a homeless individual-say a day or night shelter, valid address for a mail box, etc.-then that address will get geocoded to the appropriate geographies.

Appendix 3: CENSUS_DEMOG

The CENSUS_DEMOG table is a static reference table that will be provided to each data partner through the PopMedNet data sharing client. This data is used to provide community level attributes for each census tract or county of a patient's residence. The CENSUS_DEMOG table for CHORDS will include the American Community Survey data from 2010-2014 to start. Five-year files are required for census tract level information.

Inclusion: All available data from the American community survey should be included in this table.

Exclusion: None

Fields

Name	Description	Definition
BLOCK	A three-character numeric code assigned to census blocks.	CHAR (3) NULL
CENSUS_DATA_SRC	Source of the data in the record	CHAR (16) NULL
GEOCODE	A numeric code signifying the geographic location of the address where the person was found.	CHAR (15) NOT NULL
CHORDS_GEOLEVEL		CHAR (10) NULL
STATE	A two-character numeric code assigned to US states, districts, territories and protectorates.	CHAR (2) NULL
COUNTY	A three-character numeric code assigned to census counties.	CHAR (3) NULL
TRACT	A six-character numeric code assigned to census tracts.	CHAR (6) NULL
BLOCKGP	A one-character numeric code assigned to census block groups.	CHAR (1) NULL
HOUSES_N	Number of housing units in geography	INT NULL
RA_NHS_WH	Percent white, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_BL	Percent black or African American, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_AM	Percent American Indian or Alaska native, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_AS	Percent Asian, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_HA	Percent native Hawaiian or other pacific islander, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_OT	Percent other, non-Hispanic	NUMERIC(6,4) NULL
RA_NHS_ML	Percent two or more races, non-Hispanic	NUMERIC(6,4) NULL
RA_HIS_WH	Percent white, Hispanic	NUMERIC(6,4) NULL
RA_HIS_BL	Percent black or African American, Hispanic	NUMERIC(6,4) NULL
RA_HIS_AM	Percent American Indian or Alaska native, Hispanic	NUMERIC(6,4) NULL
RA_HIS_AS	Percent Asian, Hispanic	NUMERIC(6,4) NULL
RA_HIS_HA	Percent native Hawaiian or other pacific islander, Hispanic	NUMERIC(6,4) NULL
RA_HIS_OT	Percent other, Hispanic	NUMERIC(6,4) NULL
RA_HIS_ML	Percent two or more races, Hispanic	NUMERIC(6,4) NULL
HOUSES_OCCUPIED	Proportion of HOUSES_N that are occupied.	NUMERIC(6,4) NULL

HOUSES_OWN	Proportion of occupied housing units that are occupied by owners.	NUMERIC(6,4) NULL
HOUSES_RENT	Proportion of occupied housing units that are occupied by renters.	NUMERIC(6,4) NULL
HOUSES_UNOCC_FORRENT	Proportion of unoccupied housing units that are for rent.	NUMERIC(6,4) NULL
HOUSES_UNOCC_FORSALE	Proportion of unoccupied housing units that are for sale	NUMERIC(6,4) NULL
HOUSES_UNOCC_RENTSOLD	Proportion of unoccupied housing units that are rented or sold, but still unoccupied.	NUMERIC(6,4) NULL
HOUSES_UNOCC_SEASONAL	Proportion of unoccupied housing units that are used seasonally or other occasional use.	NUMERIC(6,4) NULL
HOUSES_UNOCC_MIGRANT	Proportion of unoccupied housing units that are used for migrant workers.	NUMERIC(6,4) NULL
HOUSES_UNOCC_OTHER	Proportion of unoccupied housing units that are vacant for other reasons.	NUMERIC(6,4) NULL
EDUCATION1	Less than 9th grade.	NUMERIC(6,4) NULL
EDUCATION2	9th -12th grade	NUMERIC(6,4) NULL
EDUCATION3	High school graduate	NUMERIC(6,4) NULL
EDUCATION4	Some college, no degree	NUMERIC(6,4) NULL
EDUCATION5	Associate degree	NUMERIC(6,4) NULL
EDUCATION6	Bachelor degree	NUMERIC(6,4) NULL
EDUCATION7	Graduate or professional degree	NUMERIC(6,4) NULL
EDUCATION8	Doctorate degree	NUMERIC(6,4) NULL
MEDFAMINCOME	Median family income	INT NULL
FAMINCOME1	Less than $10,000	NUMERIC(6,4) NULL
FAMINCOME2	$10,000 - $14,999	NUMERIC(6,4) NULL
FAMINCOME3	$15,000 - $19,999	NUMERIC(6,4) NULL
FAMINCOME4	$20,000 - $24,999	NUMERIC(6,4) NULL
FAMINCOME5	$25,000 - $29,999	NUMERIC(6,4) NULL
FAMINCOME6	$30,000 - $34,999	NUMERIC(6,4) NULL
FAMINCOME7	$35,000 - $39,999	NUMERIC(6,4) NULL
FAMINCOME8	$40,000 - $44,999	NUMERIC(6,4) NULL
FAMINCOME9	$45,000 - $49,999	NUMERIC(6,4) NULL
FAMINCOME10	$50,000 - $59,999	NUMERIC(6,4) NULL
FAMINCOME11	$60,000 - $74,999	NUMERIC(6,4) NULL
FAMINCOME12	$75,000 - $99,999	NUMERIC(6,4) NULL
FAMINCOME13	$100,000 - $124,999	NUMERIC(6,4) NULL
FAMINCOME14	$125,000 - $149,999	NUMERIC(6,4) NULL
FAMINCOME15	$150,000 - $199,999	NUMERIC(6,4) NULL
FAMINCOME16	$200,000+	NUMERIC(6,4) NULL
MEDHOUSINCOME	Median household income	INT NULL
HOUSINCOME1	Less than $10,000	NUMERIC(6,4) NULL
HOUSINCOME2	$10,000 - $14,999	NUMERIC(6,4) NULL
HOUSINCOME3	$15,000 - $19,999	NUMERIC(6,4) NULL
HOUSINCOME4	$20,000 - $24,999	NUMERIC(6,4) NULL

HOUSINCOME5	$25,000 - $29,999	NUMERIC(6,4) NULL
HOUSINCOME6	$30,000 - $34,999	NUMERIC(6,4) NULL
HOUSINCOME7	$35,000 - $39,999	NUMERIC(6,4) NULL
HOUSINCOME8	$40,000 - $44,999	NUMERIC(6,4) NULL
HOUSINCOME9	$45,000 - $49,999	NUMERIC(6,4) NULL
HOUSINCOME10	$50,000 - $59,999	NUMERIC(6,4) NULL
HOUSINCOME11	$60,000 - $74,999	NUMERIC(6,4) NULL
HOUSINCOME12	$75,000 - $99,999	NUMERIC(6,4) NULL
HOUSINCOME13	$100,000 - $124,999	NUMERIC(6,4) NULL
HOUSINCOME14	$125,000 - $149,999	NUMERIC(6,4) NULL
HOUSINCOME15	$150,000 - $199,999	NUMERIC(6,4) NULL
HOUSINCOME16	$200,000+	NUMERIC(6,4) NULL
POV_LT_50	Less than 50% of poverty level	NUMERIC(6,4) NULL
POV_50_74	Between 50% and 74.99% of poverty level	NUMERIC(6,4) NULL
POV_75_99	Between 75% and 99.99% of poverty level	NUMERIC(6,4) NULL
POV_100_124	Between 100% and 124.99% of poverty level	NUMERIC(6,4) NULL
POV_125_149	Between 125% and 149.99% of poverty level	NUMERIC(6,4) NULL
POV_150_174	Between 150% and 174.99% of poverty level	NUMERIC(6,4) NULL
POV_175_184	Between 175% and 184.99% of poverty level	NUMERIC(6,4) NULL
POV_185_199	Between 185% and 199.99% of poverty level	NUMERIC(6,4) NULL
POV_GT_200	Greater than 200% of poverty level	NUMERIC(6,4) NULL
ENGLISH_SPEAKER	Proportion of people over age 5 that speak only English or speak English "very well"	NUMERIC(6,4) NULL
SPANISH_SPEAKER	Proportion of people over age 5 who speak only Spanish or speak Spanish "very well"	NUMERIC(6,4) NULL
BORNINUS	Proportion of people over age 5 that were born in the US	NUMERIC(6,4) NULL
MOVEDINLAST12MON	Proportion of households that moved in the last 12 months	NUMERIC(6,4) NULL
MARRIED	Proportion of people over age 15 who are married	NUMERIC(6,4) NULL
DIVORCED	Proportion of people over age 15 who are divorced	NUMERIC(6,4) NULL
DISABILITY	Proportion of people over age 18 living with any disability	NUMERIC(6,4) NULL
UNEMPLOYMENT	Proportion of civilian non-institutionalized population between 18 and 64 years old who are unemployed	NUMERIC(6,4) NULL
UNEMPLOYMENT_MALE	Proportion of civilian non-institutionalized males between 18 and 64 years old who are unemployed	NUMERIC(6,4) NULL
INS_MEDICARE	Proportion of people covered by Medicare	NUMERIC(6,4) NULL
INS_MEDICAID	Proportion of people covered by Medicaid	NUMERIC(6,4) NULL
HH_NOCAR	Proportion of households with no car (owner and renter occupied)	NUMERIC(6,4) NULL
HH_PUBLIC_ASSISTANCE	Proportion of households on public assistance	NUMERIC(6,4) NULL
HMOWNER_COSTS_MORT	Proportion of households with monthly owner costs more than 50% household income, in homes with mortgages	NUMERIC(6,4) NULL

HMOWNER_COSTS_ NO_MORT	Proportion of households with monthly owner costs more than 50% household income, in homes without mortgages	NUMERIC(6,4) NULL
HOMES_MEDVALUE	Median value of homes	INT NULL
PCT_CROWDING	Proportion of households with more than 1 person per room	NUMERIC(6,4) NULL
FEMALE_HEAD_OF_H H	Proportion of households headed by females (no male present)	NUMERIC(6,4) NULL
MGR_FEMALE	Proportion of female management occupations	NUMERIC(6,4) NULL
MGR_MALE	Proportion of male management occupations	NUMERIC(6,4) NULL
RESIDENTS_65	Proportion of population over 65 years old	NUMERIC(6,4) NULL
SAME_RESIDENCE	Proportion of persons in same residence since year 2005	NUMERIC(6,4) NULL
FAMPOVERTY	Proportion of family households in the geography with below-poverty level income.	NUMERIC(6,4) NULL
HOUSPOVERTY	% Households with below-poverty level income	NUMERIC(6,4) NULL
CENSUS_YEAR	Year the census data was collected	INT NOT NULL
ZIP	A five-character numeric code assigned by the US Postal Service to various regions where it delivers mail.	CHAR (5) NULL

Appendix D: CHORDS Roles and Settings

September 2016

Definitions:

Patient-level data: Data in which each row represents one patient's information. May also use terms: record-level, individual-level data.

Summary data: Data in which counts of the patient records meeting given criteria are returned. May also use term: counts.

Aggregate data: Data in which results are pooled/consolidated across multiple sites before being returned to investigators and analyzed.

Site-specific data: Data in which results come from only one site.

Roles and Settings for Data Exchange:

Role	Description	Type of Agreement Governing Role
DataMart Client Administrator (At each data partner site.)	Review and respond to requests via the DataMart Client (both incoming requests and outgoing results).	DUA
	DataMart Administrators may also manage the metadata for their DataMart(s) and submit requests to their own DataMart(s).	
	Regular data uploads and data quality checks. Ensure timing of data updates does not interfere with query submissions and returns.	
Investigator (Data user)	May submit requests and review/export aggregated (not site-specific) results within a Project. Investigators can not select specific DataMarts; they can only set geographic/demographic selection criteria.	DUA
Enhanced Investigator (Data user)	May submit requests and review/export site-specific results for all requests within a Project.	DUA that allows site-specific data sharing. Additional user agreements as required by site.

Role	Description	Type of Agreement Governing Role
Network Administrator (UCD)	Manage the network, including creating network entities, managing access controls, and approving or creating users..	
	Submit requests and review results to ensure operations and functionality. Assist investigators and enhanced investigators.	Must be listed on project IRBs/protocols

Appendix E: CHORDS Access Request Form

1. By signing, I agree that I have read, understand and will abide by the CHORDS Data Use Agreement
2. Fill in the information below, print and sign where indicated, obtain your manager/supervisors signature and submit to: Rachel Zucker/Jessica Bondy/Bryant Doyle, email: CHORDS.Support@ucdenver.edu

By signing, I agree to use CHORDS only for work-related business. I will not share my account password with anyone or let others use CHORDS while I am logged in. I agree to use recommended security practices where possible, including encrypting computers used to connect to CHORDS; and physically securing computers by working in access-controlled or locked areas and using automatic screen-saver time outs. I agree to share the information obtained within CHORDS only with personnel who have a legitimate work-related need for the information. I understand failure to comply with the institution's policies and procedures may result in termination of access to CHORDS and other employment consequences depending on the nature of the violation. Misuse of data for purposes other than those agreed to will result in 1) Termination of access to CHORDS, 2) Report of the misuse to the study's IRB, if applicable, and/or COMIRB, and 3) Additional consequences as outlined in study-specific DUAs with partner sites.

Validation of work-related need for anyone requesting CHORDS access is required. If requesting access to electronic protected health information (ePHI) I attest to the need to use electronic devices that access, create, store and receive ePHI AND that the electronic devices that I use with ePHI are fully encrypted with all security provisions applied. I agree not to store any reports from CHORDS on removable hardware (ex. USB drives). I understand that information about my use of the CHORDS system is logged and that these logs are audited on a regular basis. I agree to access only the minimum necessary data required to perform my duties. I agree to immediately report any suspected policy violation to my supervisor.

DATA USER INFORMATION	
Name:	
Organization:	
Email:	
Phone:	
Title:	
Business Address:	
CHORDS Point of Contact Name:	
CHORDS Point of Contact Email:	
CHORDS Point of Contact Phone:	
AFFILIATION	
	I represent a healthcare organization contributing data to CHORDS

Place an 'X' next to the statement that best describes you.	I represent a local public health agency accessing public health information for community/population health monitoring	
	I am a researcher using CHORDS for research purposes	
Applicant Signature:		Date
CHORDS Point of Contact Signature:		Date
ACCORDS Representative Signature		Date

TO BE COMPLETED BY CHORDS/PopMedNet ADMINISTRATION

CHORDS/PopMedNet ACCESS ACTIVATION			
Access activation process status	User info provided to CHORDS/PopMedNet OIT Support		Date: Initials:
	Role(s) added to Requester in CHORDS/PopMedNet		Date: Initials:
	Requester activated		Date: Initials:
	Role changes made		Date: Initials:

USER ACCESS ROLES

*A CHORDS user account must be created for each individual that requires CHORDS access. Once logged in, the CHORDS user is assigned an **Access Role**. Access Roles are used for security, and define which data the user may see and which parts of CHORDS the user may access.*

CHORDS Role

	DataMart Administrator: You will be monitoring your site's datamart client (DMC), including the queries submitted and the results returned. You represent a data contributing site
	Investigator: You are submitting one or more queries and accessing data using CHORDS. You are requesting data that is aggregated from multiple partners.
	Enhanced Investigator: You are submitting one or more queries and accessing data using CHORDS. You are requesting data that is aggregated from multiple partners. You also need to see site-specific data.

Appendix F: CHORDS Incidence Response Plan

This section identifies the CHORDS Data Incident Response Plan.

A **data incident** is a situation in which CHORDS data are released, shared, and/or accessed
in a way that is inconsistent with processes approved by COMIRB/IRB of record or executed data use agreements.

Should a data incident occur, this Response Plan will be followed along with appropriate
mitigative actions to address the incident. All CHORDS data partners will be notified, within one
business day, by the CHORDS board chairperson if a data incident occurs so they can follow their local sites' policies
and procedures for reporting and mitigation, if required. A data incident may occur at a data partner site, data user
site, or at the University of Colorado's Anschutz Medical Campus.

Depending on the severity of the data incident (as determined by the Board chair in consultation with the Privacy
Officer or others as required), procedures implemented can range from communication/education in the case of a low
risk incident; up to contacting CHORDS Network staff to shut down the CHORDS instance of PopMedNet™ in the case
of a request that was submitted through PopMedNet™ and resulted in a very high-risk incident.

For data incidents occurring at a data partner site:
a. A DataMart Administrator is responsible for executing all CHORDS queries. Queries will be sent through
PopMedNet™ (PMN). Administrators have accountability for returning the query results to the the PMN client. If a data
incident occurs at a participating site, the DataMart Administrator will follow all applicable local policies and
procedures for reporting and mitigation of the data incident (i.e., notifying their institution's Privacy Officer, local IRB,
and other institutional officials as appropriate). The Data Mart Administrator will also contact the CHORDS Network
Administrator as soon as possible or within one business day of the incident occurring.

For data incidents occurring at a data user site:
b. After notifying their local IRB, privacy officer or others as required, the data user will, within one business day, notify
the chair of the CHORDS Network Operations Work Group and Board chair of the data incident issue and any
mitigative actions taken at their institution including the final resolution of the data incident. The Board chair will be
responsible for reporting the data incident with all
relevant information and within one business day to data partners and COMIRB.

CHORDS also adheres to the following security guidelines of the University of Colorado Anschutz Medical Campus
regarding data incidents. References to University documents in this section can be found here:
http://www.ucdenver.edu/research/ORC/HIPAA/Pages/policies.aspx.

Security

- 9.1 Security Management

See Attachments 1 and 2: Risk Analysis plan, including threat matrix. This plan includes our descriptions of our Risk
Management procedures. CHORDS will maintain and review our risk analysis and management documentation in
accordance with University policy as described in Chapter 9.1.

- 9.2 Security Incidents

CHORDS will adhere to CU Incident Reporting policies as outlined in this Chapter.

- 9.3 Auditing

Audit Control and Review Plan:

1. Systems and applications to be logged: CHORDS activity is logged for the central portal (web application) hosted at the Anschutz Medical Campus.
2. Information logged in each system: All data requests submitted through our system are fully logged. CHORDS staff is working with software developer Harvard Pilgrim on incorporating additional logging capabilities into future open source versions of the software.
3. Activity reports for each system: CHORDS logs are accessible to our server network administrators. CHORDS staff request these logs and review them on a quarterly schedule; logs are also available upon request.
4. Procedures to review all audit logs and activity reports, including workforce member responsible for performing the audit, the frequency the audit is to be performed, and escalation procedures if suspicious activity is detected: CHORDS Network Administrators are responsible for regular audits of available logs and activity reports. These audits will be performed quarterly and following any security incidents. If suspicious activity is detected, staff members will report the activity to CHORDS Network Administrators who will report the activity to the University's OIT and adhere to University Incidence Report policies as described in this Chapter's Section 9.3.C.

Audit Trail and Audit Trail Mechanisms

1. Logs contain the information outlined in Chapter 9.3, including user login, login date/time, and activity time.

Workforce Accountability

1. Users are trained on HIPAA accountability through university mandated HIPAA training. Additionally, users agree to adhere to CHORDS and university policies when submitting user access request forms. The policies are located and/or referenced on each form and serve as documentation that staff are trained on these policies.

- 9.4 Workforce Security
1. Access to Electronic Protected Health Information (ePHI): All individuals accessing ePHI through CHORDS will have the appropriate permission through their project to access ePHI. Each site contributing data to CHORDS must have a DUA in place (or other agreement as required by the institution); each project requesting data from CHORDS must receive IRB approval or exemption and acquire any necessary DUAs from sites (or other agreement as required by the institutions). All individuals accessing CHORDS will be confirmed by their supervisor and by CHORDS administrators as being authorized members of an approved project before being granted access.

   Individuals shall only be granted access to the minimum necessary ePHI that they require to perform their duties.

All individuals will complete a CHORDS access request form (see Attachment 3). CHORDS adheres to all University policies regarding granting, modifying, and terminating access. Supervisors of approved CHORDS projects will inform CHORDS administrators of any changes in staffing. Individuals no longer associated with an approved project will have their account disabled within one week of termination. In addition, CHORDS requires password changes on a regular basis (6 months); individuals who lose their access to their institutional email account will be unable to change their password and will therefore be locked out of their CHORDS account.

2. Workstation Use and Security
Individuals accessing ePHI through CHORDS agree to adhere to all policies regarding workstation use and security. CHORDS requires strong passwords and unique user names. Individuals agree to ensure that their workstation settings for all computers used to access ePHI through CHORDS adhere to OIT policy (including regular security patches, standard anti-virus product use, using workstations located in areas with controlled access, etc.). Users agree to use recommended security practices where possible, including encrypting computers used to connect to CUPID; and physically securing computers by working in access-controlled or locked areas and using automatic screen-saver time outs.

- 9.5 Facility and Device Security

Data will be stored on the Cancer Center VMWare infrastructure, on premise, currently in the old data center on the first floor of the Building 500. Physical access is controlled with two-factor authentication (something you have and something you know) – key card issued by the badging office and a 4-digit pin issued by the Electronic Building Security department of the University Police Department.

Information is stored on virtual servers running Windows Server 2012. Application runs on the app servers running Microsoft IIS, and the data is stored on the SQL servers running SQL Server 2012. Access is controlled with Active Directory Groups, and with roles within the SQL Server.

University data is stored in a SQL server and files in the file server at University of Colorado Cancer Center. The servers sit within university racks behind the university firewalls and have the university's security protections. The system is configurable to purge data at a set number of days. Data is encrypted at the tape backup level by the tape drives, and over the wire by TLS in IIS. IIS will be configured to use TLS 2.0 with 2048 key length. The tapes use AES256-GSM. In case of data loss and the need to restore data, these backups would be used to re-populate CHORDS data. CHORDS data is not used for clinical treatment and therefore loss of access to data would not pose an urgent emergency nor a threat to patient safety or wellbeing.

- 9.6 Transmission Security

1. Transmission Security: CHORDS securely transmits data from client datamarts at each data site to a central portal for access by an approved researcher. Only approved projects will receive data through CHORDS transmission. Human review takes place at each site when each query is received and when data is ready for transmission to the researcher. An individual's access to data is limited to the approved project and the minimum necessary data for that individual's role on the project. CHORDS data is accessed via secure file transfer and remote login protocols. CHORDS data is never transmitted via fax.

2. Transmission Security Measures: All transmissions of ePHI from CHORDS to a recipient outside the UCD network (e.g. over the Internet) utilize an encryption mechanism between UCD and the receiving entity. Files containing ePHI are transferred using a secure file transfer protocol. CHORDS does not send e-Mail messages containing ePHI for transmission outside the UCD network. . See the Secure E-Mail Transmission policy.
3. Integrity Controls: CHORDS has implemented transmission security measures to ensure that ePHI is not improperly modified during transmission. EPHI integrity is sustained using approved mechanisms in transmission from data partners (checksums, hashing algorithms) and to researchers (checksums and hashing algorithms) whenever available and feasible to protect against unauthorized alteration, tampering, corruption, or falsification of the ePHI.

- 9.7 Contingency Planning

Contingency Plan:

1. Data Backup and Storage Plan: Data contained within CHORDS is not used for patient treatment and therefore immediate recovery of the data is not required in case of an emergency. SQL server data is backed up every 15 minutes. We perform a full SQL backup nightly, differential backups every hour, and log backups every 15 minutes. This allows us to restore to any point in time. Full server backups are performed nightly. We take a copy of the VM in a consistent state and then have the ability to restore the entire VM, individual disks, or specific files as needed. Backups are performed from a snapshot of the VM while the VM is running. This data is then stored both on local disks for fast restore as well as offsite for long-term archival and DR. No backup data, or any data for that matter, is stored in the cloud.
2. Disaster Recovery Plan: CHORDS information is not critical for patient treatment and loss of access for a given period of time would not hinder operations. The physical equipment hosting CHORDS is in the OIT physical environment. CHORDS will adhere to the OIT disaster recovery plan and provide assistance for CHORDS and Cancer Center project-specific needs in the event of an emergency.
3. Emergency Mode Operation Plan: CHORDS does not provide critical business operations or functions and therefore this Plan is not necessary.
4. Testing and Revision Plan: This is not a real-time system and therefore a contingency plan in case of emergency and loss of system access is not necessary. CHORDS users would be able to go several weeks without access to the system without impeding business operations.
5. Applications and Data Criticality Plan: Does not apply, CHORDS functions as one system.
6. Emergency Access Controls: Does not apply, CHORDS functions as one system with several administrators who can provide access in case of absence of one administrator.

- 9.8 Data Integrity

1. Integrity Controls: CHORDS uses several standard integrity controls, including:
   a. Firewalls: University data is stored in a SQL server and files in the file server. The servers sit within university racks behind the university firewalls and have the university's security protections.

b. Encryption: Data is encrypted at the tape backup level by the tape drives, and over the wire by TLS in IIS. IIS will be configured to use TLS 2.0 with 2048 key length. The tapes use AES256-GSM. Data is encrypted while in transit and encrypted at rest on tape.

c. Password protection: Strong password requirements, user authentication through access request form.

d. Anti-virus software: All users adhere to University policies and update anti-virus software as requested.

e. Standards for change control, testing, documentation, approval, and rollback: Software is developed on a development system. Deployment packages for versions are created and turned over to the administrators to be run on the test system. They run regression testing before deploying a new version to the production server. Extensive documentation of the software development and additional documentation of this process exists.

2. Data Authentication Controls:
   a. Database integrity: SQL server data is backed up every 15 minutes. We perform a full SQL backup nightly, differential backups every hour, and log backups every 15 minutes. This allows us to restore to any point in time. Backups are performed from a snapshot of the VM while the VM is running. This data is then stored both on local disks for fast restore as well as offsite for long-term archival and DR. No backup data, or any data for that matter, is stored in the cloud.
   b. Message integrity: CHORDS will only transmit ePHI using https connections.
   c. Procedure integrity: The Cancer Center's servers are stored in OIT's Ground Floor datacenter. It has redundant cooling and power. The room is monitored by cameras and the door is protected by dual-factor authentication (badge and pin).

3. Software Controls: SQL server meets the requirements as defined in this Chapter. CHORDS does not allow for the modification of ePHI. Further, the original ePHI remains at its proprietary source and is not accessed at all through CHORDS. No modifications are possible for the ePHI stored in the EHRs at each source institution. Therefore, no modifications can be made which would impact patient treatment or alter a patient's original record.

- 9.9 Person or Entity Authentication

CHORDS uses unique user logins and passwords (which are encrypted when stored). CHORDS users agree to adhere to all policies outlined in this Chapter. CHORDS administrators will ensure prompt deletion of terminated users as outlined in this Chapter.

- 9.10 Device and Media Controls

All ePHI stored on hardware or electronic media will be destroyed prior to the decommissioning of the hardware or media itself in accordance with the policies and methods outlined in this Chapter. Prior to device or media re-use, all ePHI stored on a device or media will be securely removed. While CHORDS does not currently use hardware or electronic media to store ePHI (as ePHI is not transmitted through our system), once these practices are initiated CHORDS will keep a written inventory of hardware and electronic media used to store ePHI as outlined in this Chapter. All research projects using CHORDS data will need to have additional IRB and DUA documentation of the project's procedures and policies around data storage.

- 9.11 Portable Media Security

All CHORDS users agree to adhere to the policies outlined in this Chapter. ePHI will only be stored on portable media devices when necessary. All devices will have security controls in place in accordance with the University's policies, and only minimum necessary data will be stored. Data will be deleted/wiped and/or the device destroyed when the ePHI storage is no longer necessary.

- 9.12 Secure E-Mail Transmission

No ePHI is currently transmitted over email in the CHORDS system. In the future CHORDS will only transmit ePHI using https connections.

- 9.13 Security of ePHI on Home Computers

Initially, CHORDS users will not be accessing ePHI in the data and reports. When ePHI becomes accessible, all CHORDS users agree to adhere to University policies regarding anti-virus software, security patches, anti-spyware software, firewalls, etc. as outlined in this Chapter when using home computers. Each research project using CHORDS will need to document their data storage policies and procedures in study-specific IRB documents and Data Use Agreements. Additionally, CHORDS recommends that access from home computers take place only using a VPN connection to the University.

Chapter 10 Report a Breach

- 10.1 HIPAA Privacy Incident Notice
    - o CHORDS team members will use the appropriate forms to notify the HIPAA Compliance officer of potential HIPAA privacy incidents
- 10.2 Complaint Notification Form
    - o CHORDS team members will use the appropriate forms to notify the HIPAA Compliance officer of HIPAA Complain Notifications

Appendix G: CHORDS Use Cases

CHORDS Key Health Issues

Four key health issues have been prioritized for initial use cases: cardiovascular disease, diabetes, mental health and unhealthy weight.

These issues play an important role in determining overall health. Local public health partners have prioritized many of these issues for action in their community or public health improvement plans. Colorado's 2015-2019 Plan for Improving Public Health and the Environment "Healthy Colorado: Shaping A State of Health" identified obesity and mental health/substance abuse as its two flagship priority areas.

Understanding how the burden of these issues varies between demographic groups is an important public health priority. Demographic information can be used to produce prevalence estimates for specific age groups, by gender and within racial/ethnic groups. CHORDS provides information at sub-county levels (e.g., zip code, neighborhood or census tract) to define disparities as well as cross-county analyses to support cross-jurisdictional work. Data establish baseline measures to monitor key metrics for assessing and evaluating the impact of community and regional health improvement efforts, clinical interventions or new policy initiatives.

CHORDS Topic Areas

Cardiovascular Disease

Cardiovascular disease (CVD) is a leading cause of death among Coloradans, and disproportionately impacts males, older adults and Coloradans with lower incomes or less years of completed education. Rates of high cholesterol and blood pressure are more common among black and Hispanic Coloradans. CVD can be impacted by modifiable behaviors including tobacco cessation, physical activity and healthy eating. Public health programs address these heart-healthy activities for adults and children.

- CHORDS public health use cases: hypertension (e.g., screening for, prevalence, and control) or tobacco (e.g., second-hand smoke exposure or use in its various forms).

Diabetes

Colorado was one of eight states that saw a significant increase in the prevalence of diabetes between 2013-2014, according to the Colorado Department of Public Health and Environment. Diabetes is a costly condition, both in medical expenses and the toll it takes on an individual's health and well-being. Public health agencies are raising awareness of diabetes and offering evidence-based interventions to support residents with pre-diabetes.

- CHORDS public health use cases: pre-diabetes and diabetes diagnosis; diabetes control.

Mental Health

Mental and emotional well-being among adults and children is an important element of overall health. This area encompasses the depression use case described above.   Social and environmental factors and stressors can often exacerbate poor mental health. Improved screening and referral for mental health concerns, and efforts to combat stigma around seeking help, are public health priorities.

- CHORDS public health use cases: screening, referral and prevalence of anxiety, ADHD, depression and postpartum depression.

Unhealthy Weight

Obesity is a pressing health issue in communities across Colorado. Rising rates of youth overweight and obesity is especially concerning. Public health agencies are addressing unhealthy weight at the policy, systems and environmental levels with a focus on reducing health disparities.

- CHORDS public health use cases: overweight and obesity.

CHORDS Examples

Three brief examples describe how CHORDS data users (public health agencies and researchers) will interact with the CHORDS system to access data.

CHORDS can extract patient demographic, geographic and clinical information from electronic health records across health care providers to answer questions related to public health concerns.

1. Depression Prevalence

Depression is one of the most common adult health conditions and is linked to a number of comorbid conditions and death. Understanding the prevalence of depression and how burden varies between demographic groups or geographic areas can help public health agencies determine the levels of investments needed in this area, engage community partners in planning efforts and appropriately target prevention and outreach efforts.

The clinical information required for depression prevalence estimates are:

- encounter dates,
- encounter types (e.g., emergency department, inpatient, or ambulatory clinic),
- diagnosis codes,
- diagnosis dates, and
- information about the type of code (e.g. ICD-9 or ICD-10).

Demographic information about patients like gender, race and age can be used to produce prevalence estimates for each age-group, gender, and racial/ethnic group. Geographic information about the patient's residence can be used to estimate prevalence for each neighborhood or census.

Currently depression cases are identified from diagnosis codes for depression and depression related symptoms. In the future, the depression use case could be expanded to include a spectrum of depression-related services (e.g., screening questionnaires, admission or discharge information, prescription drug information or symptom surveys).

Because a care-seeking population represented in CHORDS may be different than the residents living in the community, information from the American Community Survey (ACS) are used to produce population weighted prevalence estimates. For instance, census tract level socio-economic information from ACS are used to compare levels of poverty, education, and insurance coverage to depression prevalence.

2. Congenital Heart Disease Prevalence and Treatment

Patients with congenital heart disease (CHD) are living longer than ever thanks to enhanced surgical care and medical management improvements. An estimated 85% of neonates born with CHD are now surviving into adulthood; about 12% of them have a severe CHD[1-3]. The lack of United States population-based monitoring for CHD outcomes among affected individuals, of all ages, prevents estimates of prevalence, health service utilization, health needs of these individuals, cost of services, and long-term outcomes, including comorbidities and survival. The ability to identify and track this new and growing population of adolescents and adults with congenital heart defects is critical for providing effective life-long care, accurate prevalence estimates on which to base allocation of health care resources, and evaluation of treatments and interventions.

The clinical information required for a case definition query includes:

- patient age,
- patient gender,
- encounter dates related to CHD,
- encounter types (e.g., emergency department, inpatient, or ambulatory clinic),
- diagnosis codes, and
- procedure codes.

CHORDS can identify cases of CHD across healthcare providers using EHR data. CHORDS provides information on patient demographics, location, and clinical details that can be used to determine accurate regional CHD prevalence rates. CHORDS can also obtain all encounters for patients with a CHD diagnosis (including those not stemming directly from CHD-related symptoms), providing a more complete picture of comorbidities, health outcomes, and overall healthcare utilization for patients with CHD.

1.Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, Pilote L. Congenital heart disease in the general population: changing prevalence and age distribution. Circulation 2007; 115(2):163-172

2.Opotowsky AR, Siddiqi OK, D'Souza B, Webb GD, Fernandes SM, Landzberg MJ. Maternal cardiovascular events during childbirth among women with congenital heart disease. Heart 2012; 98:145-151

3.Nieminen HP, Jokinen EV, Sairanen HI. Late results of pediatric cardiac surgery in Finland: a population-based study with 96% follow-up. Circulation 2001; 104:570-575

3. Obesity Monitoring

The epidemic of obesity is a pressing health issue in communities across the nation. However, community level information about the burden of obesity and local changes over time is limited. Resources to understand local trends in obesity are limited to obesity prevalence estimates calculated from self-reported height and weight measurements collected from the Behavioral Risk Factor Surveillance System and the Youth Risk Behavioral Survey. Investigation of survey data validity has shown that individuals are likely to overestimate their height while underestimating their weight creating inaccurate estimates of BMI and obesity prevalence. For any time period, CHORDS can extract patient demographic, geographic and height and weight information from electronic health records across healthcare providers to attribute individuals to a healthy weight, overweight, or obese status. CHORDS can repeat these queries over time to track obesity trends longitudinally for the entire Denver Metro region or in specific Denver neighborhoods. The clinical information required for longitudinal obesity estimates are:

- encounter dates,
- encounter types (e.g. emergency room, inpatient, or ambulatory encounter),
- height,
- weight, and
- calculated BMI.

Demographic information about patients like gender, race and age can be used to produce prevalence estimates for each age-group, gender, and racial ethnic group. For children, BMI percentile is used to assess weight status instead of BMI. A CDC standard calculation is used to identify implausible values, heights or weights outside of realistic ranges, and attribute a child to a weight status. Geographic information about the patient's residence can be used to estimate prevalence for each census tract or neighborhood. Annual estimates for each neighborhood could be used to study change in obesity prevalence over time and evaluate the impact of community level interventions.

As in our depression example, because a care-seeking population can be very different than the geographic population, information about residents from the American Community Survey (ACS) can be used to produce population weighted prevalence estimates. Census tract level socio-economic information from ACS can be used to compare levels of poverty, education, and insurance coverage to obesity prevalence.

Appendix H: CHORDS Work Group Contacts

Data Work Group (meetings monthly)
Emily McCormick, Denver Public Health, co-chair
Sara Schmitt, Colorado Health Institute, co-chair
Bryant Doyle, CU Anschutz
Cathy Hawkins, CCMCN
David Tabano, Kaiser Permanente Colorado
Greg Budney, Denver Public Health
Jeanette Waxmonsky, Jefferson Center for Mental Health
Jessica Bondy, CU Anschutz
Ken Scott, Denver Public Health
Kevin Matthews, Children's Hospital Colorado
Rachel Zucker, CU Anschutz
Sara Deakyne Davies, Children's Hospital Colorado
Sara Morgan, CCMCN
Will Carter, CU Anschutz

Network Operation Work Group (meets quarterly)
Rachel Zucker, CU Anschutz, co-chair
Robert Denson, CORHIO, co-chair
Bryant Doyle, CU Anschutz
Darko Teodorovic, UC Cancer Center
Jason Braddy, UC Cancer Center
Jessica Bondy, CU Anschutz
Peter Dunlap, Jefferson Center for Mental Health
Will Carter, CU Anschutz

Project Development Work Group (meetings monthly)
Greg Budney, Denver Public Health, co-chair
John Steiner, Kaiser Permanente Colorado, co-chair
Jessica Bondy, CU Anschutz
Ashlee Cerda, CCMCN
Bernadette Albanese, Tri-County Health Department
Jeanette Waxmonsky, Jefferson Center for Mental Health
Jennifer Barrow, Kaiser Permanente Colorado
Rachel Zucker, CU Anschutz
Sara Schmitt, Colorado Health Institute
Talia Brown, Boulder County Public Health