



Description of the MHS Health Level 7 Chemistry Laboratory Database for Public Health Surveillance

NMCPHC-EDC-TR-055-2019

By Beth Poitras MPH, Jessica Spencer MPH, and Kathleen Creppage MPH DrPH
EpiData Center
18 April 2019

Approved for public release. Distribution is unlimited. The views expressed in this document are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. I am an employee of the U.S. Government. This work was prepared as part of my official duties. Title 17, U.S.C., §105 provides that copyright protection under this title is not available for any work of the U.S. Government. Title 17, U.S.C., §101 defines a U.S. Government work as a work prepared by a military Service member or employee of the U.S. Government as part of that person's official duties.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 18 Apr 2019		2. REPORT TYPE Technical		3. DATES COVERED (From - To) 01 May 2004 - 31 Oct 2018
4. TITLE AND SUBTITLE Description of the MHS Health Level 7 Chemistry Laboratory Database for Public Health Surveillance			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Beth Poitras MPH, Jessica Spencer MPH, Kathleen Creppage MPH DrPH			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The EpiData Center (EDC) Navy and Marine Corps Public Health Center (NMCPHC) 620 John Paul Jones Circle, Suite 1100 Portsmouth, VA 23708-2103			8. PERFORMING ORGANIZATION REPORT NUMBER NMCPHC-EDC-TR-055-2019	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) The EpiData Center (EDC) Navy and Marine Corps Public Health Center (NMCPHC) 620 John Paul Jones Circle, Suite 1100 Portsmouth, VA 23708-2103			10. SPONSOR/MONITOR'S ACRONYM(S) EDC, NMCPHC	
			11. SPONSORING/MONITORING AGENCY REPORT NUMBER NMCPHC-EDC-TR-055-2019	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT The EpiData Center (EDC) at the Navy and Marine Corps Public Health Center (NMCPHC) evaluated the Health Level 7 (HL7) formatted chemistry data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry/serology records, describes the pathway of data from healthcare provider to the EDC, provides detailed descriptions of all variables within the database, and assesses the database's strengths and limitations. Given an understanding of the strengths and limitations of the data, HL7 chemistry laboratory data have proven to be a valuable source of health information for surveillance purposes. The data allow the creation of a timeline of events corresponding to a specific disease occurrence. Furthermore, data are received in a timely fashion, allowing for near real-time surveillance of diseases.				
15. SUBJECT TERMS Health Level 7 (HL7) formatted data, chemistry laboratory database, Military Health System (MHS), public health disease surveillance, military health surveillance				
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT		18. NUMBER OF PAGES
U		UU		46
a. REPORT	b. ABSTRACT	c. THIS PAGE	19a. NAME OF RESPONSIBLE PERSON	
U	U	U	Beth Poitras MPH	
			19b. TELEPHONE NUMBER (include area code)	
			757-953-0970	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18



Abstract

The EpiData Center (EDC) at the Navy and Marine Corps Public Health Center (NMCPHC) evaluated the Health Level 7 (HL7) formatted chemistry data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry/serology records, describes the pathway of data from healthcare provider to the EDC, provides detailed descriptions of all variables within the database, and assesses the database's strengths and limitations. Given an understanding of the strengths and limitations of the data, HL7 chemistry laboratory data have proven to be a valuable source of health information for surveillance purposes. The data allow the creation of a timeline of events corresponding to a specific disease occurrence. Furthermore, data are received in a timely fashion, allowing for near real-time surveillance of diseases.



Contents

Abstract	iii
Executive Summary	6
Background	7
Public Health Surveillance Applications.....	7
Data Origination and Flow Process	8
Data Structure and Analysis.....	10
Structure	10
Analysis	11
Key Fields for Public Health Surveillance.....	12
Duplicates.....	12
Unique Patient/Specimen	12
Test Definition.....	13
Laboratory Test Result	14
Date/Time References	14
Strengths	16
Timeliness	16
Completeness	16
Organization and Structure.....	16
Public Health Applications.....	16
Limitations	17
Completeness	17
Inclusion	17
Generalizability	17
Inconsistency	17
Data Interpretations	18
All Data Fields (Variables)	19
Automatically Populated Fields	19
Field Descriptions	21
Abbreviations	43
USAFSAM U.S. Air Force School of Aerospace Medicine	43
Example Projects using Chemistry Laboratory Data.....	44



List of Tables and Figures

Table 1. Example of an HL7 Chemistry Data Feed for one Patient showing ACCESSION NUMBER, TEST NAME, and TEST RESULT.....	13
Table 2. Methods Used to Determine Timeframes in HL7 Chemistry Data	15
Table 3. Description of Fields in HL7 Chemistry Data (N=1,751,211,787 records as of 31 October 2018)	21
Figure 1. HL7 Laboratory Data: Origination and Exclusions.....	9
Figure 2. Laboratory Data Flow from MTF to EDC	10
Figure 3. HL7 Chemistry Record Structure	11
Figure 4. Timeline of Events in HL7 Chemistry Record.....	15



Executive Summary

The EpiData Center (EDC) at the Navy and Marine Corps Public Health Center (NMCPHC) was tasked by the Department of Defense (DOD) Global Emerging Infections Surveillance and Response System (GEIS) with the evaluation of the Health Level 7 (HL7) formatted chemistry data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry records, describes the pathway of data from provider to the EDC, provides a detailed description of all variables, and assesses the database's strengths and limitations.

The HL7 chemistry database is used extensively by the EDC for a variety of tasks, including daily case finding of reportable diseases, preparation of health reports, and response to congressional requests for disease burden (e.g., prevalence of sickle cell anemia among DOD beneficiaries). Positive laboratory test results can be matched with outpatient or inpatient records to identify how specific laboratory tests correlate with encounters, which may help with case validation and confirmation.

Using multiple fields within the HL7 chemistry database assists in queries and surveillance based on disease context. Combinations of sponsor's Social Security number (SSN), Family Member Prefix (FMP), and/or unique specimen identifiers are used to identify unique cases or individuals. The date and time variables distinguish timeframes between different laboratory events.

HL7 data are generated within the Composite Health Care System (CHCS) at fixed military treatment facilities (MTFs). A typical analysis would not include chemistry records from purchased care providers, shipboard facilities, battalion aid stations, or in-theater facilities.

Chemistry data are useful for identifying laboratory-confirmed cases of illness. However, cases where a physician chooses to treat presumptively without laboratory confirmation are not captured. Clinical practice with regards to chemistry testing varies between providers and facilities. Examples of situations where chemistry testing may not be performed include confirmatory tests for patients with influenza-like illness symptoms, or patients with superficial infections who are treated presumptively.

HL7 chemistry data are valuable as a source of health information for surveillance purposes. The data are organized in a logical and informative manner. The database is particularly useful for surveillance due to its event time stamping, timely data, efficient data structure, and accurate, complete, and certified entries from the Defense Health Services System (DHSS).



Background

The EpiData Center (EDC) of the Navy and Marine Corps Public Health Center (NMCPHC) evaluated the Health Level 7 (HL7) formatted chemistry data source for its usefulness in public health surveillance at the request of the Department of Defense (DOD) Global Emerging Infections Surveillance and Response System (GEIS). HL7 data include radiology records, anatomic pathology reports, and pharmacy transactions; however, chemistry and microbiology laboratory test results were identified by the EDC as the most useful data type for detection of emerging health threats among the military population. Therefore, extensive work was done to examine the laboratory datasets, determine their completeness and reliability, identify areas to improve surveillance, and establish methods for the surveillance of specific conditions. Available data fields have changed throughout the course of HL7 development; this document reflects variables (also referred to as “fields”) available as of February 2019. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry records, describes the pathway of data from the healthcare provider to the EDC, provides a detailed descriptions of all variables within the database, and assesses the database’s strengths and limitations.

Records for all DOD military service members (Army, Navy, Marine Corps, Air Force, Coast Guard), US Public Health Service personnel, National Oceanic and Atmosphere Administration (NOAA) personnel, overseas civilian personnel, TRICARE eligible dependents, and others who undergo chemistry testing at a military treatment facility (MTF) are included in the HL7 chemistry dataset. Chemistry records for the Department of Navy (DON) have been collected at NMCPHC since 01 May 2004 and were integrated into health surveillance in 2005. Due to additional GEIS funding to support DOD pandemic influenza surveillance, HL7 chemistry laboratory data for the entire DOD have been collected at NMCPHC since 21 July 2006.

In February 2018, MTFs began the transition to a new platform for electronic medical records, GENESIS. GENESIS will replace other systems of record in the MHS and, after implementation is completed, be the source of laboratory data for surveillance purposes. At present, the EDC is unable to obtain data feeds from GENESIS; a process for obtaining this data is under review within the EDC, but, for now, as initial facilities complete the transition to GENESIS, visibility on related laboratory occurrences is lost.

Public Health Surveillance Applications

The HL7 chemistry database is used to support DON and DOD Preventive Medicine activities. Support includes reportable disease case identification, analysis of case burden, estimation of occupational exposures (as indicated by organic biomarkers such as lead and cadmium), and identification of testing for genetic diseases and conditions (such as sickle cell anemia, G6PD deficiency, and familial hypercholesterolemia). Epidemiologic analysis of HL7 chemistry data focus on trend identification of illness or condition. Positive laboratory results may be matched to encounter records to identify how specific laboratory tests relate to clinical encounters. Alternatively, International Classification of Diseases, 9th and 10th Editions, Clinical Modification



(ICD-9-CM, ICD-10-CM) codes in encounter data are also matched with HL7 chemistry laboratory data to identify how specific encounters relate to laboratory testing. HL7 chemistry laboratory data may also be matched to HL7 pharmacy records in order to determine compliance with clinical treatment guidelines or to assess the burden of disease by assessing common treatment trends for specific diseases. Many variables in a data-based disease surveillance model depend on clinical practice. These approaches will provide a more comprehensive description of how laboratory tests and results compare with information provided in encounter and pharmacy data. The integration of multiple data sources provides a robust, multifaceted understanding of military beneficiary health trends.

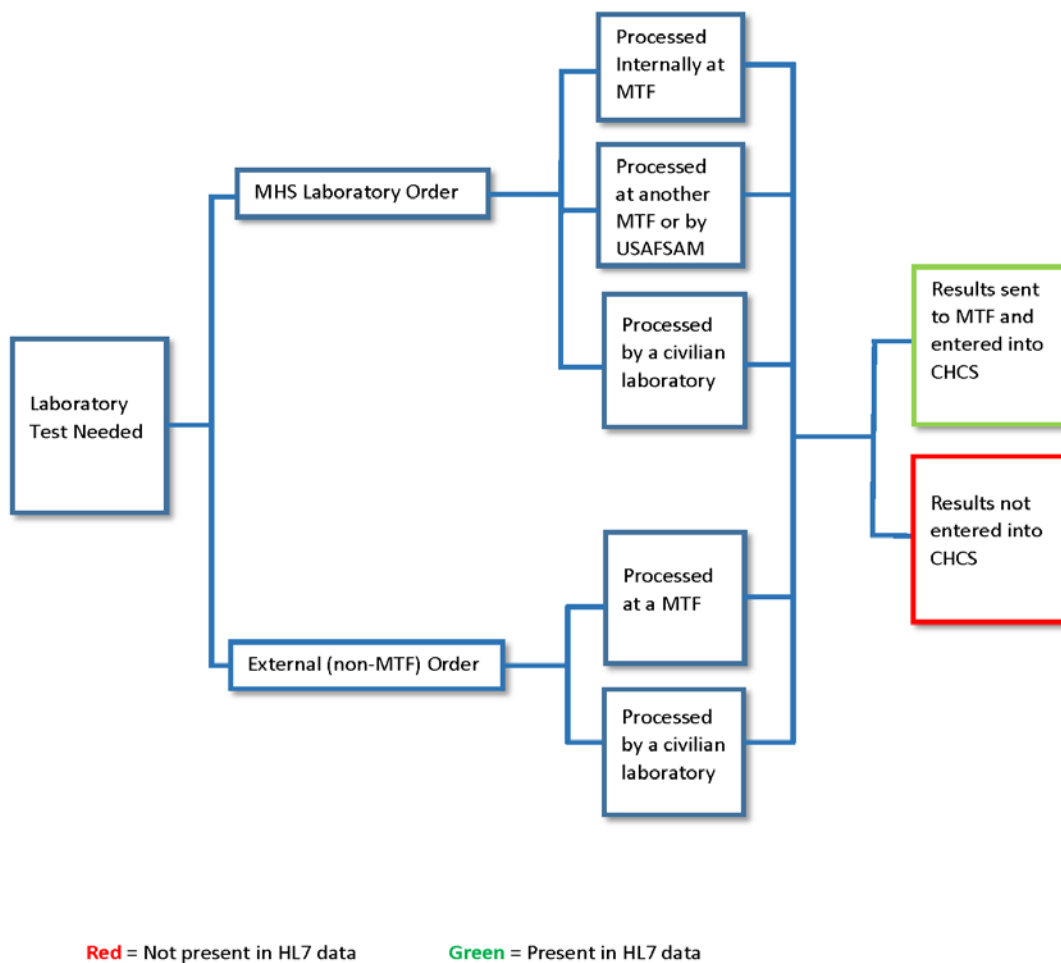
Additional public health applications involve investigation into the types and subtypes of disease and their impact on public health measures or policies. For example, knowing the subtype of a disease can assist in the predictions for the upcoming season, or aid in the development of vaccinations. Applications of these data are not limited to traditional medical event surveillance. These data can fill a significant gap in the DOD's ability to track conditions identified through chemistry laboratory testing.

Data Origination and Flow Process

Figure 1 provides an overview of the flow of information into the HL7 chemistry database. HL7 laboratory data originates at the point of care, when a patient is seen by a provider at a fixed MTF and the provider orders a laboratory test via the Composite Health Care System (CHCS). The provider generally selects from a list of MTF-specific tests. The list in CHCS is generated by the MTF laboratory Management Information Department (MID) based on laboratory tests that can be conducted at the local facility or at contracted outside laboratories. Therefore, lists of available laboratory tests in CHCS vary by facility. Additionally, the same test may be performed at different facilities with different names based on how they are set up in CHCS. Laboratory orders can also be entered into CHCS from providers outside the military healthcare system or through a referral from a non-CHCS participating military provider (e.g., ship-board clinic, battalion aid station) if the patient provides a laboratory order slip. Depending on the patient's status, specimens can be drawn within the hospital ward and sent to the laboratory (inpatient), or can be drawn directly at the laboratory during clinical encounters (outpatient). A laboratory test can be performed on site, or may be outsourced if the local facility is unable to perform the test indicated. When a test is completed, results are entered into CHCS either automatically by the machine performing the test or manually by the laboratory technician. Results are then certified by a laboratory technician or supervisor. The laboratory results must be entered into CHCS for visibility in the chemistry database regardless of whether the laboratory test was conducted within the MHS or outsourced to a civilian facility.

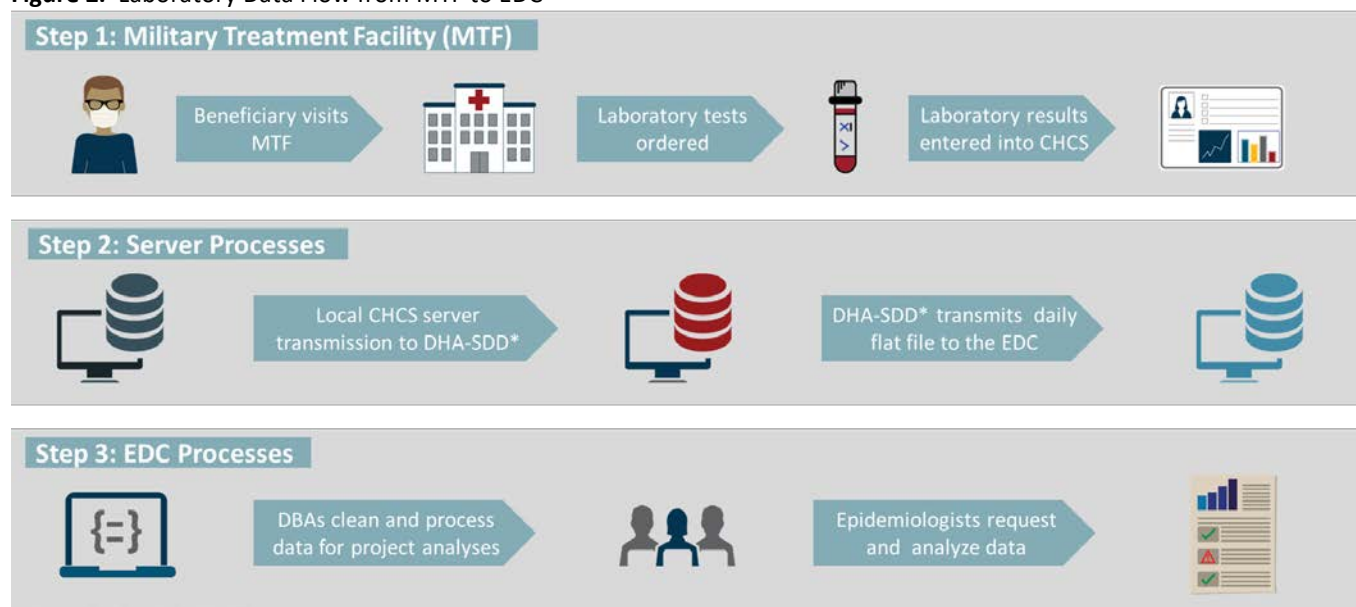


Figure 1. HL7 Laboratory Data: Origination and Exclusions



After certification in CHCS, a script is run to generate an HL7 message for that test result. The HL7 message is then archived and batched with other HL7 messages on the local CHCS host. At least once a day, these HL7 messages are forwarded to Defense Health Agency-Solutions Delivery Division (DHA-SDD) main servers. Once forwarded and receipt is verified by DHA-SDD, HL7 messages at the local host are deleted. These records are then retrieved from the main servers and parsed into a database design four times a day. The EDC receives flat file extracts of the raw parsed data from DHA-SDD on a daily basis using a secure connection, as dictated by the Interface Control Document (ICD). The EDC database administrators clean and process the data for use by analysts. Figure 2 outlines the flow of laboratory data from the MTF to the EDC.

Figure 2. Laboratory Data Flow from MTF to EDC



Data Structure and Analysis

Structure

HL7 chemistry data are retrieved by the EDC in a standard, pipe-delimited flat file from DHA-SDD. Fields (or “variables”) provided are described in detail in a subsequent section of this document ([Field Descriptions](#)). Records are grouped at the highest level by ACCESSION NUMBER. The ORDER NUMBER can refer to multiple test orders within the ACCESSION NUMBER. The results for the tests within the ORDER NUMBER are grouped using MESSAGE ID and SET ID. Multiple tests may be seen grouped together within the same ACCESSION NUMBER (e.g., panels). See Figure 3, HL7 Chemistry Record Structure.



Figure 3. HL7 Chemistry Record Structure

Sponsor_Id	FMP	Accession_Number	Order_Number	Msg_Id	Set_Id
012345678	20	160926 ABC 11	012345-11111	A111-012345678	1
012345678	20	160926 ABC 11	012345-11111	A111-012345678	2
012345678	20	160926 ABC 11	012345-11111	B111-012345678	1
012345678	20	160926 ABC 11	023456-22222	F111-012345678	1
012345678	20	160926 ABC 11	023456-22222	F111-012345678	2
012345678	20	160926 DEF 22	034567-33333	H222-234567890	1

In this example, this record has 2 unique specimens, 3 unique orders, 4 unique messages, and 6 unique rows.

Analysis

Most military laboratory facilities can perform routine chemistry testing, such as complete blood counts (CBC), Chem 7s, rapid plasma reagins (RPR), and hepatitis panels. More specialized tests (e.g., western blots) are usually outsourced to a contracted private laboratory. If a test is outsourced, MTFs are required to ensure those results are entered into CHCS for clinical evaluation by the provider following the receipt of results from the contracted laboratory. However, all MTFs may not be compliant with this policy.

Laboratory tests that are not traditionally classified as chemistry tests may be found in the chemistry dataset due to MTF-specific situations. This may occur if an MTF acquires microbiology specimens through the chemistry section of the laboratory. This practice is more common in smaller laboratories without a microbiology-specific sector, which may perform non-chemistry tests in the chemistry laboratory. Therefore, when searching for a particular outcome, both laboratory datasets (chemistry and microbiology) should be included to ensure capture of all applicable test results.

A laboratory test can be used as a diagnostic tool but is not required for many illnesses, and may not estimate the true burden of disease. For instance, in some cases a provider may identify and treat patients based on clinical symptoms alone. Other times, based on epidemiologic characteristics such as extent of disease in the community, seasonality, or disease characteristics, the disease may be identified without laboratory confirmation. Thus, a disease surveillance model based on laboratory results alone may skew illness patterns, unless these and other factors are taken into consideration.

Several considerations must be made before analysis of a particular outcome or disease. It is critical to understand the pathogen or disease, its clinical presentation, provider practices with regard to treatment and laboratory testing for the disease/pathogen, available testing methods, and



laboratory testing or outsourcing procedures. It is also important to note that test names and results are not standardized in CHCS or in the HL7 data, but are often standardized at each MTF. The CHCS administrator at an MTF may set up drop-down menus and panels based on the facility's capabilities. Multiple test names and result formats may be associated with a particular disease depending on the testing method and recording practices of the facility. Therefore, in conducting a query of particular tests in the HL7 chemistry table, misspellings and abbreviations should be included. Formatting variations should also be considered, including spacing and use of periods. Multiple fields may be searched for flagged tests associated with a particular outcome as well. The subset of data created for the disease of interest should then be reviewed to assess completeness and reliability of the data with regard to that particular outcome.

To re-evaluate and assess the HL7 chemistry dataset and provide insight into data fields as described below, data were extracted for DON beneficiaries from the EDC's HL7 chemistry database from 01 May 2004 – 31 October 2018. All data were reviewed and analyzed to ensure the dataset addresses the disease surveillance needs of the Navy and Marine Corps. The data extract reflects information from the US DOD military personnel and their dependents whose test results were certified in the CHCS at a DOD MTF.

Key Fields for Public Health Surveillance

HL7 chemistry data are continually reviewed and analyzed to identify changes in data structure as well as updates in clinical practices to address the disease surveillance needs of the EDC. Data standardizations and methods for identifying duplicate and unique records have been established.

Duplicates

True duplicates are records in which all fields have identical values. If true duplicates exist, one record is kept and the duplicates are eliminated by EDC database administrators prior to providing data to the analyst. Each record that remains after removing true duplicates is considered a unique record; there is at least one variable value different than all other records in the database.

Unique Patient/Specimen

Unique patients are identified in the HL7 chemistry data through a combination of SPONSOR ID and FMP. This combination creates a unique identifier that can be used to track individual patients through all chemistry records. Another variable called the Electronic Data Interchange Personal Number, or EDIPN, is a DOD identification (ID) number specific to each beneficiary; this field was included in the HL7 chemistry data since 2016, but not all beneficiaries have received an EDIPN. The field is not consistently populated to date, but will replace the SPONSOR ID/FMP for identification of unique patients after implementation is complete. It is important to note that it is possible for individuals to have two separate SPONSOR IDs over time. For example, if the child of a sponsor becomes active duty, then that child will have his/her own SPONSOR ID. Each unique patient can have multiple laboratory orders in the HL7 chemistry data.



A unique specimen is defined as each sample collected from each SPECIMEN SOURCE from each person. A unique ACCESSION NUMBER is assigned to each unique specimen collected. Patients may have multiple samples taken at the same time. Each sample has a different ACCESSION NUMBER, even if many of the other fields are the same. ACCESSION NUMBER may be reused by the MTF, but when combined with SPONSOR ID, FMP, and ORDER EFFECTIVE DATE, the combination is unlikely to be reused. Because multiple tests can be performed on the same sample, there can be several records with the same ACCESSION NUMBER.

Test Definition

The TEST ORDERED and the TEST NAME fields are used to identify the test performed. HL7 chemistry tests are often ordered in panels. The TEST ORDERED will remain constant while the TEST NAME will change to reflect the individual tests within the panel (See Table 1). TEST NAMES and TEST ORDERED are not standardized due to variation between computer coding within the regional CHCS and the capabilities of each laboratory. It is important to consider this in analysis of these data.

When a test is canceled, inaccurate, or misread, the details are recorded within the TEST RESULT or the RESULT NOTES field; a new row is not created. When a test is reordered, separate rows are generated to show the newest results, and references the previous test.

Table 1. Example of an HL7 Chemistry Data Feed for one Patient showing ACCESSION NUMBER, TEST NAME, and TEST RESULT

ACCESSION NUMBER	SET ID	TEST ORDERED	TEST NAME	TEST RESULT
070101 – ACH – 045	1	CBC W/O DIFF	HEMATOCRIT	36.9
070101 – ACH – 045	2	CBC W/O DIFF	HEMOGLOBIN	12.7
070101 – ACH – 045	3	CBC W/O DIFF	MCH	28.7
070101 – ACH – 045	4	CBC W/O DIFF	MCHC	34.5
070101 – ACH – 045	5	CBC W/O DIFF	MCV	83.2
070101 – ACH – 045	6	CBC W/O DIFF	MPV	7.9
070101 – ACH – 045	7	CBC W/O DIFF	RBC	4.43
070101 – ACH – 045	8	CBC W/O DIFF	WBC	5.5
070101 – ACH – 045	9	CBC W/O DIFF	PLATELET CT	255
070101 – ACH – 047	1	CHEM 7	CREATINE	0.6
070101 – ACH – 047	2	CHEM 7	UREA NTROGEN	13
070101 – ACH – 047	3	CHEM 7	GLUCOSE FASTING	172
070101 – ACH – 047	4	CHEM 7	SODIUM	136
070101 – ACH – 047	5	CHEM 7	POTASIUUM	3.6



070101 – ACH – 047	6	CHEM 7	CHLORIDE	103
070101 – ACH – 047	7	CHEM 7	CARBON DIOXIDE	23

Laboratory Test Result

Due to the structure of the laboratory datasets, the result or its subgenus (or subtype) could be recorded in multiple fields. Most test results for chemistry are seen in the TEST RESULT field. When a result is not present within this field, usually identified by terms such as ‘See Results’, ‘See Note’, etc., a search within the RESULT NOTES could indicate a positive or negative result.

An entry could have multiple results per ACCESSION NUMBER, either through errors, or multiple tests performed. Therefore, it is important to review the records based on the most recent results certified by the laboratory (CERTIFY DATE), the most certain RESULT STATUS OBX (Corrected, Final, then Pending), and the SET ID to select the most accurate results.

Date/Time References

Figure 4 illustrates the four date fields with associated time fields in the chemistry laboratory dataset. ORDER EFFECTIVE DATE is the date the provider requests the laboratory test be performed. Data pulls for epidemiologic projects primarily use the ORDER EFFECTIVE DATE to capture the timeframe when the patient is likely to be symptomatically ill, which often leads to tests ordered to assist with diagnosis. It is possible for a provider to enter a laboratory test order to be performed in the future; this may be common practice for chronic illnesses or long term inpatient visits. It is also possible that the patient did not immediately report to the facility to have the specimen collected. Therefore, the appropriateness of using ORDER EFFECTIVE DATE should be considered. An alternative proxy may be the sample COLLECTION DATE, which is the date that the specimen was taken from the patient. The CERTIFY DATE is the date that a laboratory technician certifies the results into CHCS, or makes changes within the system. The CERTIFY DATE may identify when a condition or disease was first identified or confirmed and may be used to evaluate proper disease reporting. The MSG DATE is the date when the record was sent from CHCS to DHA-SDS.



Figure 4. Timeline of Events in HL7 Chemistry Record

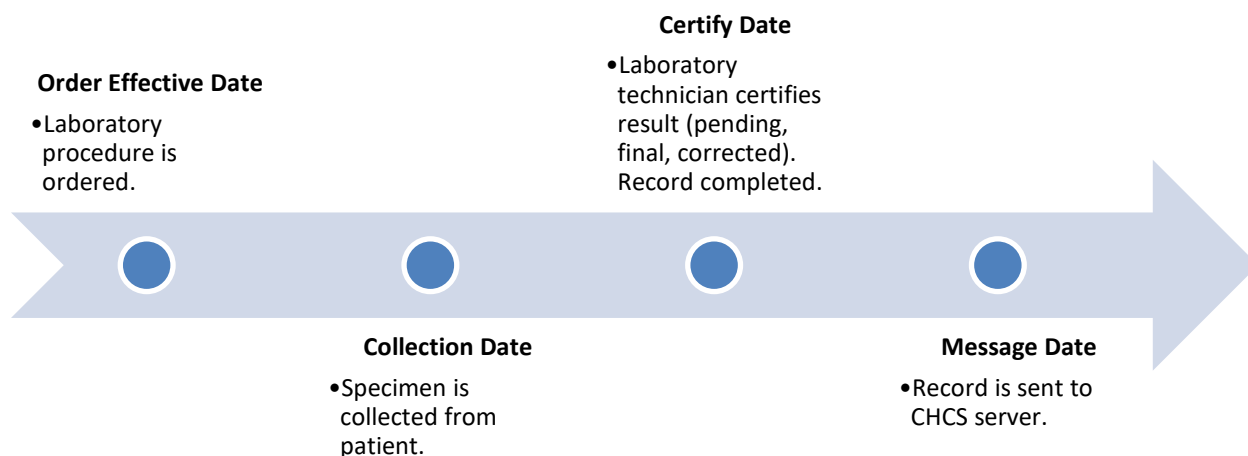


Table 2 reflects methods used to determine timeframes within the HL7 chemistry database. Each date and time pair references one action towards the completion of the laboratory test. A time span was created when comparing one date/time variable to another date/time variable.

Table 2. Methods Used to Determine Timeframes in HL7 Chemistry Data

Date/Time Variables	Interpretation
Message – Order Effective	Timeframe from test ordered date/time until HL7 message date/time generation at the local CHCS host
Order Effective – Date of Birth	Age of member at date of test order
Collection – Date of Birth	Age of member at specimen collection date
Certify – Order Effective	Timeframe from test ordered date to result date
Certify – Collection	Timeframe from sample collection date to result date



Strengths

Timeliness

DHSS includes several date and time fields in the data file provided to the EDC, including: MSG DATE/TIME, ORDER EFFECTIVE DATE/TIME, REQUESTED DATE/TIME, COLLECTION DATE/TIME, and CERTIFY DATE/TIME. A timeline of dates is provided in Figure 4. In order to assess the timeliness of the data, the MESSAGE DATE (the date that the record is sent to the local CHCS server) was compared to the DHSS LOAD DATE (the date that local records are sent to the central DHSS server). Over 94% of records were transferred from the local facility to the central DHSS server and available to the EDC within one day.

Completeness

Records are received from the majority of shore-based fixed MTFs connected to the CHCS, but gaps in data may exist. Gaps in data may occur due to server failure at location or due to functional errors. The completeness of individual fields varies and the characteristics of each are described in detail in Table 3 of this document. In general, some fields of particular interest, such as SPONSOR ID, FMP, and SERVICE, are highly populated due to the business rules of CHCS. Entries are only received by DHSS after the test has been completed and certified, which increases the accuracy of data for outcome analysis and surveillance.

Organization and Structure

Data are organized in a useful manner. Data fields of highest importance for surveillance are complete and contain expected variables. Test results are located on single rows of the data table, which limits the need for restructuring. Single line entries allow for flexibility with stratifications, filtering, and aggregating. Timelines can then be determined within each test from the date and time variables.

Public Health Applications

The HL7 chemistry database provides the ability to establish a disease timeframe based on actual laboratory results. As technology advances, laboratory tests have improved specificity and sensitivity compared to previous methods. Although a laboratory result alone may not indicate a diagnosis, it can be used in combination with clinical symptoms or conditions to determine a more accurate diagnosis. Existing surveillance methods suffer from severe and systematic underreporting and undercounting. Understanding the limitations of the HL7 chemistry data has led to more accurate estimation of case burden for several diseases of interest.



Limitations

Completeness

Several of the identified data fields of public health interest are highly populated, but others are not. The completeness of each data field, as described in Table 3, should be considered. Incomplete demographic information can limit the ability to describe disease burden demographically. Complementary databases, such as personnel data, may need to be utilized to supplement HL7 data for more complete and accurate demographic information.

When the DHSS node fails, HL7 messages may be lost and those that had been sent may not be retrievable unless the network outage was planned in advance. Gaps may exist in the data received at NMCPHC, though these gaps are limited and quality assurance checks are in place to determine if data received are of sufficient quantity and quality.

Inclusion

The chemistry data only include records generated by MTFs that use CHCS servers. MTFs that do not use CHCS (including forward deployed clinics, ship-board clinics, and battalion aid stations) are not captured in this dataset. Records for tests ordered by these clinicians are present in the dataset if a service member is referred to an MTF that uses CHCS. Additionally, the severity of illness or lack of on-location capabilities may require the service member to be transferred from a non-CHCS MTF to a shore-based CHCS facility. Comparatively, should a beneficiary seek care outside the MHS, such as the use of TRICARE Prime, his/her records would not be captured within CHCS or HL7 unless the laboratory testing was completed at an MTF laboratory.

Generalizability

Data are generated from the medical laboratory records of military service members and their dependents, a highly specific patient population that differs from the general US population in many ways, including average age, gender distribution, physical fitness, and health status. Further, this population has universal access to medical care, which is not true of the US population. Unfortunately, reservists are not well captured in HL7 unless they are activated for duty or seek medical care within the MHS during training. These differences limit the comparability to the general US population. Extra precautions are required when extrapolating data to larger populations and comparing the disease rates and trends of military and non-military populations.

Inconsistency

The format of CHCS allows for free text fields (e.g., TEST NAME, TEST RESULT, RESULT NOTES), resulting in variations between entries and difficulties in HL7 data interpretation. Free text fields limit the ability to determine specific test type (e.g., enzyme immunoassays or culture), test results, and reference ranges. Non-standardized naming and result conventions hinder the ability to use standardized syntax. However, methods have been developed at the EDC to overcome these barriers and quality assurance is ongoing.



Inconsistencies between MTFs limit data use. Tests performed are based on the capability of each laboratory. The MTF laboratories determine testing capabilities and control what providers may order on their CHCS laboratory order screens. Also, TEST NAME and TEST RESULT choices in CHCS (pull down menus) are determined and controlled at the local CHCS host, introducing variation between facilities. Due to this variation, extensive coding and methodical review ensure comprehensive case capture.

Data Interpretations

While HL7 chemistry data may be useful to identify laboratory-confirmed cases of illness, not all cases will be present if physicians elect to treat patients presumptively without laboratory confirmation. Clinical practices, and reliance on symptomatic examination versus diagnostic testing, may vary between providers. For example, during the influenza season, providers may not order confirmatory tests for patients who present with influenza-like illness symptoms, but may choose to treat presumptively instead.



All Data Fields (Variables)

Automatically Populated Fields

Several types of automatically populated fields in the chemistry data are created at different sources, including the CHCS host, each MTF (a division of the host), and each laboratory work section.

When a facility registers within CHCS, several fields are created for use in identification of the facility per DHSS. These fields include: PERFORMING DMIS ID, PERFORMING DMIS FACILITY NAME, PERFORMING FACILITY SERVICE, REQUESTING DMIS ID, REQUESTING FACILITY NAME, and REQUESTING FACILITY SERVICE. Additionally, DHSS populates the SET ID and SERVICE variables.

Each patient or dependent is registered in the Defense Eligibility Enrollment Reporting System (DEERS), which feeds into CHCS. When a patient presents at a medical facility, the SPONSOR ID (usually the Social Security number) is entered and their name is chosen from a drop down list. The following patient demographic fields are automatically populated after this selection, if they were entered when the patient was registered in DEERS: DATE OF BIRTH, ETHNICITY, FMP, GENDER, MARITAL STATUS, PATIENT CATEGORY, PATIENT ID, RACE, EDIPN, and SPONSOR ID. If these data are not present in the system, a designated unknown value is entered, and, therefore, records should have no missing values in these fields. Administrative personnel at the MTF have the ability to edit these fields at the time of visit. It is the sponsor's responsibility to update his/her family's DEERS information when they report to a new duty station or when demographic information changes.

As records are created, edited, and completed, several fields are created by the CHCS clock when a specific action is taken, such as order entry or certification of a test result. These variables include: ORDER EFFECTIVE DATE/TIME, SPECIMEN COLLECTION DATE/TIME, REQUESTED DATE/TIME, and CERTIFY DATE/TIME. These can be changed if necessary by the laboratory staff, but is not common practice.

MSG DATE, MSG ID, MSG TIME, and MSG SENDING FACILITY are created and assigned when the message (record) is sent to the CHCS server. The MSG ID value is created when the certified laboratory record is processed by the CHCS server at the local level for batching. The time and date are also automatic values created by the system. The MSG SENDING FACILITY is a code which identifies the server and facility creating the message.

The CHCS host or MTF limits some content by pre-setting drop-down menus or lookup tables for some fields (e.g., the TEST ORDERED menu is often limited to only those tests the MTF is authorized to perform onsite or outsource). Should a physician choose a test from the TEST ORDERED field, a variation of tests within the TEST NAME could be selected. Based on the physician's request, the SPECIMEN SOURCE, BODYSITE COLLECTION SAMPLE, CPT



CODE DATA, and NO OF CPT CODES are selected for each TEST NAME. Drop-down lists and lookup tables can be changed if necessary by the laboratory staff.

The laboratory results and each reference range can be entered by the laboratory staff, automated by the laboratory CHCS coordinator, or automated by the specific laboratory equipment used. Specific models of laboratory equipment can be programmed to indicate the normal ranges of values, label abnormalities, and create syntax per each specimen's result. Depending on each laboratory facility and equipment used, the TEST RESULT, RESULT NOTES, REFERENCE RANGE, ABNORMAL FLAG, SENSITIVE RESULT FLAG, and UNITS OF MEASURE fields can be automated or manually entered by staff.



Field Descriptions

Observations are based on DOD data. Within the chemistry dataset, there are 56 standard variables. Frequency distributions from 01 May 2004 through 31 October 2018 were run on select data fields from the chemistry database to describe completeness. All fields were available since 2004 unless otherwise noted. These fields are presented in alphabetical order.

Table 3. Description of Fields in HL7 Chemistry Data (N=1,751,211,787 records as of 31 October 2018)

Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
ABNORMAL FLAG	Automated via regional CHCS location or lab equipment OR entered manually	String field	Three (3) possible values: H (High) L (Low)	85% (N=1,496,267,321)	Indicates whether the test result is abnormally high or low. The majority of the entries have missing values.
ACCESSION NUMBER	Automatically populated by CHCS	Combination of: 1) date in YYYYMMDD 2) a two or three character alpha setting 3) a numeric listing of how many tests of that specific type were run in one day	Last numeric digits can range from 1 to 9999	0% (N=2,582)	Created for each unique biological sample collected from the patient. Different chemistry tests from the biological sample can have the same ACCESSION NUMBER. These numbers could be recycled throughout a day's time, and should not be used solely to identify a record. They may be used to determine tests ordered per patient in conjunction with the SPONSOR ID, FMP, and ORDER EFFECTIVE DATE.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
BODYSITE COLLECTION SAMPLE	Physician can select from drop down menu or look up table; staff can edit	Character		0% (N=37,845)	Refers to the place on the body where the specimen is collected from the patient. This field is used with SPECIMEN SOURCE to determine where the sample is taken. A patient can have numerous samples taken from one area (e.g., a lung can have numerous biopsy specimens, thus having a different ACCESSION NUMBER for each specimen). But, like SPECIMEN SOURCE, it can be used to determine if proper protocol was used for a test, or can be used to determine the type of test performed (e.g., CBC would not have urine or bladder as a sample type).
CERTIFY DATE	Automatically populated by CHCS clock; staff can edit but rarely do	YYYYMMDD		0% (N=70,970)	Date when a laboratory technician certifies the results into CHCS, or makes changes within the system. Unlike the ORDER EFFECTIVE DATE, there can be deviations between the values within SET ID, due to different test run dates. Contains limited missing values. All entries have valid dates. The values of the timeframe are between ORDER EFFECTIVE DATE and MSG DATE.
CERTIFY TIME	Automatically populated by CHCS clock; staff can edit but rarely do	HHMM (Standard 24 hour clock)	0000 - 2359	0% (N=64,322)	Represents the time component of the CERTIFY DATE. All times are valid entries.
CLINICAL COMMENTS	Manually entered by provider or lab technician	Free text	Free text	84% (N=1,474,321,562)	Allows the provider or laboratory technician to add additional information regarding the patient's symptoms, CAP* survey information, contact phone numbers, specimen media, or instructions on test procedures. This field is not primarily used in case definition, but is added to eliminate superfluous entries. *College of American Pathology (CAP) conducts surveys for quality assurance within the laboratories.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
COLLECTION DATE	Auto generated by CHCS clock, staff can edit	YYYYMMDD	N/A	0% (N=786)	Date when the specimen is extracted from the patient. The value for this entry should be between the values of the ORDER EFFECTIVE DATE and the CERTIFY DATE. Includes limited missing values. All records without missing values have valid dates. Because the field approximates the time that the laboratory sample is collected, it may be useful for analysis. It can be used for time analysis between the specimen collection and test result certification. By knowing the timeframe of each test conducted, an approximation of the type of test used can be determined.
COLLECTION TIME	Auto generated by CHCS clock, staff can edit	HHMM (Standard 24 hour clock)	0000 - 2400	0% (N=26)	As with COLLECTION DATE, the COLLECTION TIME is the time when the specimen is extracted from the patient, and follows a standard 24-hour clock. Unlike ORDER EFFECTIVE TIME, the timeframe is from 0000 to 2400. All times are valid entries. There are no missing values.
CPT_Count	Provider can select from drop down menu or look up table; staff can edit	Numeric		0% (N=12,621,149)	Lists the number of CPT codes for each test performed. The number of CPT codes is determined at each regional location.
CPT_List	Provider can select from drop down menu or look up table; staff can edit	Numeric		0% (N=12,613,632)	CPT-4 code data for laboratory procedure. NOTE: This field is formed by concatenating three pieces of data from each ZL1 segment contained on a raw HL7 message. The data pieces include: CPT Code, CPT Modifier, and CPT Status.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
DATE OF BIRTH	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	YYYYMMDD		0% (N=1,313)	It is possible to have inaccurate values for DOB. If the DOB is unknown and the year is, then CHCS enters zeros for the month and day. Not all dates for this field are valid (e.g., dates with a year in the early 1900s or a date with a year in the future). This field is required within CHCS.
DHSS LOAD DATE	Automatically populated by the CHCS clock; staff can edit but rarely do	YYYYMMDD		32% (N=565,260,357)	Date when DHSS loads the data from the central CHCS server. The field is used to determine the timeliness of reporting and to identify lags in reporting times from certain MTFs.
DHSS LOAD TIME	Automatically populated by the CHCS clock; staff can edit but rarely do	HHMM	0100-2312	32% (N=565,260,357)	Time component of the DHSS LOAD DATE field.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
EDIPN	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Numeric		4% (N=82,522,315) Since 2016	Electronic Data Interchange Personal Number, or EDIPN, is a DOD ID number specific to each beneficiary; this field was included in the HL7 chemistry data since 2016, but not all beneficiaries have received an EDIPN. The field is not consistently populated to date, but will replace the SPONSOR ID/FMP for identification of unique patients after implementation is complete.
ETHNICITY	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Alphanumeric	Six (6) possible values: 1 – Hispanic 2 – South Eastern Asian 3 – Filipino 4 – Other Asian Pacific Islander 9 – Other Z – Unknown	13% (N=237,594,662) No missing values prior to 2014	Language or cultural group that the patient claims. The majority of records indicated Other (49%). About 30% of records from chemistry were categorized as Unknown. Hispanic ethnicity made up 3% of the records within the DOD extract. The results indicated that the field of ETHNICITY may be self-identified and not consistently reported; those not reported are labeled as Unknown. The Unknown responses are assumed to be pre-populated in order to eliminate blanks within the database. It limits the ability to identify disease trends in minority groups and to identify diseases that have a disproportionate burden on these groups.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
FMP	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Numeric (2 digits)	See Notes	0% (N=0)	Family Member Prefix – designates the relationship of the patient to the sponsor. In 2018, 58.9% of records had an FMP of 20 (sponsor) followed by 27.8% with an FMP of 30 (spouse of sponsor). Other possible values include 01-19 (child of sponsor, numbered in age order). No records were missing an FMP value. Unknown entries are labeled as 99. List of values and % populated is available upon request.
GENDER	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Single alpha character	Three (3) possible values: M – Male F – Female X – Unknown	0% (N=10,005) No missing values prior to 2014	Patient's sex. Values derived from the DOD Standard Gender Table. The majority of records were populated. The distribution between females and males is similar, F 50% (N=879,746,830), M 49% (N=871,441,685), X 0% (N=13,267).



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
MARITAL STATUS	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Single alpha character	Nine (9) possible values: A – Annulled D – Divorced I – Interlocutory Decree L – Legally Separated M – Married N – Never Married S – Single/Not Married W – Widow/Widower Z – Unknown	11% (N=202,394,161) No missing values prior to 2014	Patient's marital status. Derived from DoD Standard File. The majority of records are classified as Married (42%, N=750,714,062) followed by Unknown (27%, N=480,809,110)) and Single/Not Married (12%, N=226,265,849).
MEPRS CODE	Automatically populated when the record is created	Four alpha characters	The first letter indicates the most general area: A – inpatient B – outpatient C – dental D – ancillary E – support services F – special programs G – medical readiness	0% (N=54,557)	Medical Expense and Performance Reporting System CODE that indicates where within the MTF the laboratory order was entered. It is automatically populated when the record is created. This field is useful for tracking where people are seen within the MTF. It can indicate ambulatory care, special dialysis clinics, the maternity ward, etc. which can affect the interpretation of the data. The majority of records present in the chemistry dataset have a MEPRS code that begins with B.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
MSG DATE	Automatically populated when the message (record) is sent to CHCS server	YYYYMMDD		0% (N=0)	This date approximates the transaction time between the MTF and the regional CHCS site, but it can vary based on location. Some MTFs send messages in batches, therefore the time or date portions may not correlate to the actual transaction time. There are no missing values and all are valid dates.
MSG ID	Automatically populated when the message (record) is sent to CHCS server	Alphanumeric. Varies by MTF and may include numbers, letters, or numeric code that identify the MTF, or it can identify the function of the message (e.g, RESCHED-057342).		0% (N=0)	The Message ID is an alphanumeric code assigned to each batch of messages based on when the message is sent from CHCS to the server. The MSG ID is not unique to each record ; each batch of messages is assigned one MSG ID.
MSG SENDING FACILITY	Automatically populated when the message (record) is sent to CHCS server	Alphanumeric Four possible formats: A#### F#### HP#### N####		0% (N=64,741) No missing values since 2009	This field allows analysts to identify and track problems that arise in the transfer of messages from the MTFs to DHSS and the EDC. There are no missing records within this dataset.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
MSG TIME	Automatically populated when the message (record) is sent to CHCS server	HHMM	0001 - 2359	0% (N=0)	Time when the message is sent from the MTF to the regional CHCS site. All times are valid entries. There are no missing values.
ORDER EFFECTIVE DATE	Automatically populated by CHCS clock; staff can edit but rarely do	YYYYMMDD		0% (N=27,861) No missing values since 2009	Date that the laboratory order enters the CHCS system and indicates when the laboratory test was actually ordered. All entries are valid dates. ORDER EFFECTIVE DATE may be used to approximate when a patient was ill, to analyze the difference in time between when the order was issued and when the sample was collected, and to assess the length of time between the dates the test was ordered and when the data are available for use in the EDC.
ORDER EFFECTIVE TIME	Automatically populated by CHCS clock; staff can edit but rarely do	HHMM	0000 - 9000	0% (N=27,861) No missing values since 2009	The field represents the time component of the ORDER EFFECTIVE DATE and is formatted using a standard 24-hour clock. Unlike MSG TIME. All times are valid entries.
ORDER NOTES COMMENTS	Populated by provider	Free text	Free text	99% (N=1,751,209,575)	Allows the provider to provide comments or notes which accompany the test ordered. This field is rarely populated.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
ORDER NUMBER	Automatically populated by CHCS	#####-##### (11 numerical digits with a hyphen: The first 6 digits include the date and the last 5 digits include consecutive numbers for tests provided at that specific location)		0% (N=4)	The ORDER NUMBER is unique to each order, but not unique for each record. An order can have multiple records that correspond to changes made to the order (e.g., changes in test, cancellations), or refer to multiple parts to a test (e.g., results for influenza A and influenza B). All changes appear as individual records with the same ORDER NUMBER. It is a plausible way to track a patient but not useful for identifying unique records.
ORDERING PROVIDER	Automatically populated by CHCS	Last Name, First Name, Middle Initial (three components separated by commas)	Last, First, MI	0% (N=231)	Indicates the name of the ordering physician. It is structured to facilitate analysis but could be separated if necessary.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
PATIENT CATEGORY CODE or PATCAT CODE	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Alphanumeric X## First letter refers to the sponsor's branch of service Following 2 digits correspond to the patient's relationship to the sponsor	Nine (9) possible values for sponsor branch: A – Army B – National Oceanic and Atmospheric Administration C – Coast Guard F – Air Force K – Other beneficiaries of the federal government M – Marine Corps N – Navy P – US Public Health Service R – NATO recipient	0% (N=1,016,577)	Indicates the patient's status with the uniformed services. For example: A11=Army Active Duty Member, A41=Army Dependents of Active Duty Member, etc. A complete list should be obtained from DOD resources. List of PATCAT values and % populated is available upon request.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
PATIENT ID	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	##### Nine digit numeric value Patient's SSN when available		0% (N=210,282)	The PATIENT ID is intended to serve as a unique identifier for each patient. The PATIENT ID is the patient's Social Security number (SSN) when available; however, the accuracy of this field cannot be assured based on the EDC's observations and analyses. In place of PATIENT ID, SPONSOR ID and FMP should be used to identify individual patients. This value is missing in 0.01% of records overall. It is important to preserve the entire PATIENT ID when importing the data into SAS or other analysis programs. The PATIENT ID field needs to be imported as a character field so that leading zeros are not dropped.
PERFORMING DMIS FACILITY NAME	Automated from DHSS	Four numeric digits, import to SAS as character ####		6% (N=110,396,202) Not consistently completed until 2015	Text translation of the DMIS ID provided in the PERFORMING DMIS ID field. This field indicates the laboratory facility name where the specimen was sent, not where the specimen was collected from the patient. This field is assigned by DHSS at the request of the EDC. The translation of the DMIS code on the official list is often more accurate than the Performing Facility field in CHCS. Use of this field allows for more accurate analysis of geographic information. Because the field is also a translation of the performing facility field in CHCS, it will be missing when that variable has a missing value.
PERFORMING DMIS ID	Automated from DHSS	Four numeric digits, import to SAS as character ####		0% (N=1,888,542)	Identifies the MTF that performed the laboratory test. This code allows for grouping of MTFs based on geographic location, as well the ability to identify parent/child relationships between installations.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
PERFORMING FACILITY SERVICE	Automated from DHSS	Single alpha character	Five (5) possible values: A – Army C – Coast Guard F – Air Force N – Navy P – DHA	6% (119,883,555) Not consistently completed until 2015	Indicates the branch of service with which the MTF is associated. This value is determined from the DMIS code list provided to DHSS by the EDC. It is missing when the Performing Facility information is missing. This field is useful for limiting the observations included in an investigation. Often, the data available for use are limited by branch of service for the MTF or patient.
PERFORMING FACILITY NAME	Automated from CHCS	Text field		0% (N=386)	The performing facility field in CHCS indicates the name of the MTF where the test was performed. This is a relatively standard text field. Problems are encountered if the text is entered incorrectly when the facility is registered in the system (e.g., misspellings). There are numerous facility names listed within the chemistry dataset but no values missing.
PERFORMING WORK CENTER NAME	Automated from CHCS	Unstructured text field		0% (N=386)	Indicates the work center within the laboratory that provided the service. This field is a relatively unstructured text field with many possible values. Over 400 performing work centers were identified in 2011. These locations are usually laboratories mapped according to the PERFORMING DMIS ID.
RACE	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Single alpha character	Six (6) possible values: C – White M – Asian or Pacific Islander N – Black R – American Indian or Alaskan Native X – Other Z – Unknown	10% (N=186,820,616) Approximately 26% missing values between 2014 and 2018	Patient's racial classification. Derived from the DOD Standard File. 27% of the records are classified as Unknown, followed by White (35%), Black (11%), Other (11%), and Asian (3%). Those entries that have no reported RACE are labeled as Unknown. The Unknown responses are assumed to be pre-populated, to eliminate blanks within the database. These records are distributed across DMIS locations, which indicate that the problem is not a data quality problem at one MTF or on a specific CHCS server. This limits the ability to use the data to look at diseases or conditions that disproportionably affect one race.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
RECORD TYPE	Automatically populated from CHCS	Three alpha characters, a standard value (LCH)	LCH – Chemistry dataset	0% (N=0)	The value “LCH” for RECORD TYPE identifies the chemistry dataset. All entries in this dataset have the value of LCH in this field. There are no missing data in this field.
REFERENCE RANGE	Automated via regional CHCS location or lab equipment OR entered manually	Alphanumeric Examples: 13.0 -16.0 G/DL 7 - 20.5 U/G		14% (N=260,456,359)	The REFERENCE RANGE is an alphanumeric variable that 1) shows a comparative numeric value indicating a result or lesser than the norm, 2) a character string (ex: “less than” or “greater than 8”), or 3) a positive or negative reference. These values are used in conjunction with the TEST RESULT to allow clinical interpretation of the results. The positive or negative readings have a character string entry in the TEST RESULT. Listings akin to “less than” have a titer result within the TEST RESULT, which are usually a derivative of 8. Other values show a normal range or a normal titer level (1:8 or <8). There are different expressions of REFERENCE RANGE, and the majority of the entries are missing values.
REQUESTED DATE	Automatically populated by CHCS clock; staff can edit but rarely do	YYYYMMDD		0% (N=8)	This field is not frequently used within the data analysis, as the ICD does not provide a detailed definition. All entries for this field have valid dates. The values of the REQUESTED DATE are between ORDER EFFECTIVE DATE and COLLECTION DATE.
REQUESTED TIME	Automatically populated by CHCS clock; staff can edit but rarely do	HHMM	0000 – 2359	0% (N=8)	Represents the time component of the REQUESTED DATE formatted using a standard 24-hour clock. All times are valid entries.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
REQUESTING DMIS FACILITY NAME	Automated from DHSS	Text field		5% (N=89,145,904) Approximately 15% missing values from 2012 – 2014; not consistently completed until 2015	Text translation of the DMIS ID provided in the REQUESTING DMIS ID field. This field indicates the laboratory facility name that is requesting laboratory service to be completed. This allows for more accurate investigations when geographic information is used, because it is created using an official DOD DMIS list. This field is a translation of the Requesting Facility field in CHCS; therefore, it will be missing when that field is missing in the record.
REQUESTING DMIS ID	Automated from DHSS	Four numeric digits ####		0% (N=2,140,540)	Four digit code assigned by the DOD to all units in all installations to uniquely identify them. The code allows grouping of MTFs based on geographic location, as well as to identify parent/child relationships between installations. Importing this field in character format can prevent the loss of leading zeros, which may produce complications when producing summary statistics. Because this field is calculated based on the Requesting Facility field, all records missing a value for that field are missing a value for the REQUESTING DMIS ID field. Missing values are limited and seen at few specific MTFs. Other records include test entries from inactive sites.
REQUESTING FACILITY NAME	Automated from DHSS	Text field		0% (N=83,573)	Field in CHCS that indicates the name of the MTF where the order originated, and is a relatively standard text field. Problems are encountered if the text is entered incorrectly when the facility is registered in the system (e.g., misspellings). The field allows tracking of orders from origin to where they are filled.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
REQUESTING FACILITY SERVICE	Automated from DHSS	Single alpha character	Five (5) possible values: A – Army C – Coast Guard F – Air Force N – Navy P – DHA	7% (N=132,769,997) Approximately 16% missing values from 2010 - 2014	Indicates the branch of service with which the MTF is associated. This value is determined from the DMIS code list provided to DHSS by the EDC. It is missing from a record when the Requesting Facility information is missing. Because this field is mapped to the REQUESTING DMIS FACILITY NAME and the REQUESTING DMIS ID, the REQUESTING FACILITY SERVICE is missing when the other two fields are blank. This field is useful for limiting the observations by the branch of service. This may be necessary for comparison, as data compared to the HL7 datasets are often service branch specific.
REQUESTING WORK CENTER NAME	Automated from DHSS	Unstructured text field		0% (N=27,985)	The Requesting Work Center is the ward or clinic within the MTF that requests the laboratory test. Entries are labeled as DMIS ID number, clinic wards, service centers, and unknown/other MTF locations.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
RESULT NOTES	Automated via regional CHCS location or lab equipment OR entered manually	Automatic/drop down menu that is created when the TEST NAME is selected via CHCS or free-text		81% (N=1,429,127,013)	The RESULT NOTES field is a character string which allows the laboratory technician to provide additional information about the result, a recommendation for additional testing, or the interpretation of the laboratory result. This field is either an automatic/drop down menu that is created when the TEST NAME is selected via CHCS or free-text. A frequency of the variable shows numerous variations of the same basic note. There are duplicates of long interpretations of a particular test by its DMIS location, which signifies syntax mapped out for the CHCS location and the test ordered. Other text entries show limited comments, such as the laboratory technician's initials, which genus type was found, or the address where the sample is tested. The RESULT NOTES field is useful when the TEST RESULT field does not indicate the outcome of a test (i.e., when Test Result shows either a reading of 'SEE NOTES' or 'SEE COMMENTS'). The RESULT NOTES field is a hindrance due to the variability of the entries as well as having elongated text fields which limits the use of wildcards and searches. If the field is not populated, it indicates either a result's notes is not mapped to that particular test by the DMIS location or the laboratory technician did not manually enter information.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
RESULT STATUS OBX	Automated via lab equipment OR entered manually	Single alpha character	Three (3) possible values: P – Preliminary F – Final C – Correction	0% (N=9,390)	The RESULT STATUS OBX field is a character string which shows the status of the test performed. These tests are used in a timely fashion, and always follow the order P, F, and then C. A test always has an F within a group of SET IDs of a test, but may also have a P or a C. Should a test have more than one RESULT STATUS OBX, it has the same SET ID, TEST NAME, and TEST ORDERED, but is on a separate entry line. An entry of “C” is entered when the record is amended due to operator error, wrong test ordered, or the test is performed for the wrong patient. An entry of “C” may also be entered if an error is identified in what was originally entered as the result.
SENSITIVE RESULT FLAG	Automated via regional CHCS location or lab equipment OR entered manually			99% (N=1,750,662,913)	Unpopulated but is included in the final daily feeds.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
SERVICE	Automated from DHSS	Single alpha character	Nine (9) possible values: A – Army B – National Oceanic and Atmospheric Administration C – Coast Guard F – Air Force K – Other beneficiaries of the federal government M – Marine Corps N – Navy P – US Public Health Service R – NATO recipient	0% (N=744,879)	Refers to the service branch of the sponsor. The value is determined from the first component of the PATCAT field and the values are the same. Therefore, there are an equal number of records that are missing the branch of service and PATCAT CODES. In 2018, the highest proportion of records belonged to the Army, Navy/Marine Corps, and Air Force, respectively.
SET ID	Automated from DHSS	Numeric	1 – 99	0% (N=477)	Automatically populated by CHCS. The numbers show the logical order of arrival of OBX segment data within an HL7 message. There could be numerous types of tests and controls within the same ACCESSION NUMBER and SPECIMEN SOURCE, thus would have a different number associated to each test listed. Should an entry have a change in its resulting status (from pending to final, from final to corrected), the SET ID is the same for that test entry.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
SPECIMEN SOURCE	Physician can select from drop down menu or look up table; staff can edit	Text field		0% (N=37,762)	Text field that indicates where the specimen was taken from on the patient. This field is useful to determine if the proper protocol was used for a laboratory test. Also, the laboratory does not always differentiate between the specimen source and the specimen location. This is seen with entries that indicate tissue of the patient's body, such as Both Eyes, Buttocks, Mouth, Oral Cavity, or Toe. Less than 1% of records had missing values in this field.
SPONSOR ID	Automatically populated from DEERS after user enters SPONSOR ID; staff can edit; sponsor responsible for maintaining DEERS information	Nine numeric digits with no dashes: #####	Sponsor's SSN with no dashes	0% (N=56,216)	The SPONSOR ID is not sufficient to identify a unique patient, but may be used in conjunction with the FMP as a unique patient identifier. It is important to preserve the entire SSN when importing the data into any analysis program. The SPONSOR ID variable needs to be imported as a character field so that leading zeros are not dropped. Not all Sponsor IDs are Social Security Administration SSNs. If the patient does not have a valid SSN, a pseudo SSN is created. The pseudo Sponsor ID begins with 800 or 900, followed by the date. If the number is already assigned to another patient, the primary three numbers will change to 801 or 901 consecutively depending on the number created with the same date. Additionally, quality assurance testing is conducted in laboratories. Quality assurance procedures utilize SSN-like identifiers in the SPONSOR ID field. The Sponsor ID for these procedures may resemble a pseudo-SSN, arbitrary identifiers such as 777777777, or three consecutive zeros. These tests will have labels such as: Ztest, Quality Control, PSR, CAP, Non-human (NH,#), etc.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
TEST NAME	Physician can select from drop down menu or look up table	Text field		0% (N=482)	Shows which test is performed on the sample provided. This value is usually from a drop-down list of tests related to the TEST ORDERED variable. This field does not have missing values because TEST NAME is automated by the regional CHCS. The TEST NAME includes entries such as tests to be performed, quality controls, temperature, and even alerts for positive results. Quality control tests are within this field, and are noted via a ZZZ prior to the actual test name. The variance between tests names suggests the fields are automated by a local CHCS host, not at the main location. A test procedure can be specific or general. A test name can refer to the procedure type, e.g. convalescent, cultures, or antibody testing. Therefore, a test name should match the type of test result present (a titer should be numeric, organism specific tests can reflect positive or negative or even sub-typing).
TEST ORDERED	Physician can select from drop down menu or look up table	Text		0% (N=8)	Identifies the requested observation, test, or panel. Each regional CHCS location has the autonomy to determine the criteria for each test ordered. Therefore, the TEST ORDERED field can have different grouping of tests per DMIS locations. The TEST ORDERED value is repeated among all records for tests associated with it according to the ORDER NUMBER. A provider can use a pull-down menu to determine the test(s) to be performed on a specimen. This shows all available tests per each test ordered.



Field Name	Input Method	Field Format	Values	Percentage of Missing Values % (N)	Notes
TEST RESULT	Automated via each regional CHCS location OR entered manually	Alphanumeric		0% (N=2,626,523)	Shows either the pending information or the results of a test ordered. The TEST RESULT field can show positive or negative results, titer numbers, control values, dates, reorders, references to see comments, notice that a test was not performed due to inadequate results, or insufficient quality. There are multiple variations, including misspellings and slang language (i.e. NOPERS). Many of these variations show the same result, such as Positive, POSITIVE, POS, and so forth. Currently, CHCS is in the process of regulating the regional CHCS locations to create one specific result for each TEST NAME outcome. This will limit the variation of TEST RESULT significantly. When the TEST RESULT is missing, it is recommended to consult the RESULT NOTES field, which is a free text field and may include the results of the test.
UNIT OF MEASURE	Automated via regional CHCS location or lab equipment OR entered manually	Alphanumeric		21% (N=363,147,621)	The UNIT OF MEASURE is a character field which shows the concentration of the TEST RESULT.



Abbreviations

AMA	American Medical Association
CAP	College of American Pathologists
CBC	Complete Blood Count
CDC	Centers for Disease Control and Surveillance
CHCS	Composite Healthcare System
CPT	Current Procedural Technology
DHA	Defense Health Agency
DHA-SDD	Defense Health Agency Solutions Delivery Division
DHSS	Defense Health Services System
DMIS	Defense Medical Information System
DOD	Department of Defense
DON	Department of the Navy
EDC	EpiData Center
EIA	Enzyme Immunoassay
FMP	Family Member Prefix
GEIS	Global Emerging Infections Surveillance and Response System
HL7	Health Level 7
ICD-9-CM	International Classification of Diseases, 9 th Edition, Clinical Modification
ICD-10-CM	International Classification of Diseases, 10 th Edition, Clinical Modification
MEPRS	Medical Expense Performance Reporting System
MHS	Military Health System
MTF	Military Treatment Facility
NMCPHC	Navy and Marine Corps Public Health Center
SSN	Social Security Number
USAFSAM	U.S. Air Force School of Aerospace Medicine



Example Projects using Chemistry Laboratory Data

Project Name/Description	Population of Interest	Purpose	Impact
Daily Case Finding Identification of reportable medical events from laboratory data	DOD Beneficiaries	Identify global infectious disease trends in near real time	Identification of disease outbreaks among specific populations; identify disease trends; evaluate gaps in infectious disease reporting
Influenza SITREP Identification of influenza trends; an example of surveillance integrating multiple data sources (laboratory, pharmacy, encounter, and vaccination data)	DON Beneficiaries	Weekly update of laboratory positive influenza cases	Identification of weekly influenza trends and comparison to past seasons; identification of regional trends; impact to special populations such as active duty, recruits, and children
G6PD Deficiency Identification of individuals with G6PD, an enzyme deficiency that may cause hemolysis when exposed to specific anti-malarial drugs	DOD Active Duty	Identify incidence of G6PD deficiency among DOD service members	Support readiness through the identification of possible magnitude of hemolytic side effects from anti-malarial drugs
Blood Lead Level Surveillance	DOD Beneficiaries	Identify blood lead levels; identify at risk populations;	Identify occupational exposure from shooting ranges among specific military populations; identify geographic areas where children's blood lead levels exceed CDC guidelines
Burden of Hepatitis A Surveillance	DON Beneficiaries	Describe the burden of Hepatitis A using medical encounter data	Confirms case-burden of Hepatitis A in MHS.

