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

NMCPHC-EDC-TR-054-2019

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

# REPORT DOCUMENTATION PAGE

*Form Approved*
**OMB No. 0704-0188**



**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 18 Apr 2019 | Technical | 01 May 2004 - 31 Oct 2018 |

**4. TITLE AND SUBTITLE**
Description of the MHS Health Level 7 Microbiology 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-054-2019

**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-054-2019

**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 microbiology data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 microbiology database and its contents, explains the creation of microbiology records, describes the pathway of data from healthcare provider to the EDC, provides a detailed description 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 microbiology data have proven to be a valuable source of health information for surveillance purposes. The data allow the creation of a time line 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, microbiology 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 | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Beth Poitras MPH |
| U | U | U | UU | 43 | 19b. TELEPONE 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 microbiology data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 microbiology database and its contents, explains the creation of microbiology records, describes the pathway of data from healthcare provider to the EDC, provides a detailed description 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 microbiology 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

## List of Figures and Tables

## Executive Summary

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

The HL7 microbiology database is used extensively by the EDC for a variety of tasks, including daily case finding of reportable diseases, identification of antibiotic-resistant organisms, preparation of health reports, and responding to congressional requests for disease burden. Disease burden evaluations have included respiratory infections (e.g., pandemic influenza, pertussis), skin and soft tissue infections (e.g., methicillin resistant *Staphylococcus aureus*), and gastrointestinal infections (e.g., salmonellosis, norovirus). Positive microbiology test results can be matched with outpatient or inpatient encounter records to identify whether laboratory tests correlate with encounters, which may help with case validation and confirmation.

The HL7 microbiology database is limited such that records from purchased care providers, shipboard facilities, battalion aid stations, or in-theater facilities are not available. Cases where a physician chooses to treat presumptively without laboratory testing will not be captured. In addition, the use of rapid or polymerase chain reaction (PCR) testing may have altered clinical practices to rely on more rapid testing, rather than culturing for organism identification with susceptibility results. Clinical practice with regards to culturing varies between providers and facilities. Examples of situations where cultures may not be performed include confirmatory tests for patients with influenza-like illness symptoms, or patients with superficial infections who are treated presumptively.

HL7 microbiology data are 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. Data are received in a timely fashion, allowing for near real-time surveillance of diseases.

## Introduction

The EDC at the NMCPHC evaluated the HL7 data source for its usefulness in health surveillance activities at the request of the Department of Defense (DOD) Global Emerging Infections Surveillance and Response System (GEIS). The HL7 data source includes records from anatomic pathology, chemistry, microbiology, pharmacy, and radiology sources. Laboratory results (particularly microbiology and chemistry) were initially identified by the EDC as the most useful type of HL7 data for improving military health surveillance activities. Therefore, extensive work was done to examine the laboratory databases, determine completeness and reliability, identify areas for improvement in surveillance, and establish methods for the surveillance of specific conditions. This technical document describes the fields (also referred to as "variables") in HL7 microbiology data, the extent of missing values within these fields, modifications made to the HL7 data flow and processing schema, data cleaning rules, and other comments regarding surveillance activities.

When HL7 data were first received by the EDC in 2004, a significant amount of work was devoted to ensuring messages were parsed and organized properly. Sample extracts for review were received from Defense Health Services System (DHSS) (now known as Solution Delivery Division, Defense Health Agency (DHA-SDD)) from 01 September 2003 to 30 April 2004. Initial sample extracts showed data were sparse. Conversations with personnel at local military treatment facilities (MTFs) and analysis of particular fields in the sample files revealed several observations. First, not all microbiology results for a given culture were seen as expected in the sample extracts. Further investigation and discussions led to comparison of local MTF results, the original HL7 message for those results, and the records in the DHSS staging database. Based on these comparisons, DHSS reconfigured the HL7 process such that microbiology data were not lost when messages were parsed and submitted.

The second issue discovered during the initial review process involved fields that identified which MTFs requested and performed laboratory test orders. Frequencies of these fields showed many MTFs were not represented in the data. As this information was passed onto DHSS, the Military Health System (MHS) Helpdesk was contacted to remedy this situation. Within several months, missing MTFs began appearing in the HL7 data. In addition, DHSS began to monitor incoming message traffic by the Composite Healthcare System (CHCS) sending facility. By May 2004, all major MTFs and most clinics were represented in the HL7 data extracts.

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 EDC has used the HL7 microbiology data to support Department of the Navy (DON) and DOD preventive medicine activities since 2005. More recently, data have been used to support surveillance and preventive medicine activity for the Defense Health Agency (DHA)l. Examples of support include case findings of particular diseases (e.g., malaria, meningococcal meningitis, or influenza) and identification of antibiotic resistance patterns. The data are used for analyses in support of collaboration with MTFs for local infection control activities.

Epidemiologic analysis of these data focuses on defining trends of illness by reviewing laboratory test orders and results. These results may be linked to other databases for a more comprehensive description of a disease event. For example, positive laboratory results are matched to International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM[1]) codes in encounter records to identify the association between specific laboratory tests and clinical encounters and comorbid conditions. Pharmacy data may also play a significant role in conjunction with laboratory results data.

Applications of these data are not limited to traditional medical event surveillance. These data fill a significant gap in the DOD's ability to track and describe antibiotic resistance throughout the MHS. The MHS needed the ability to track patients with antibiotic-resistant infections across MTFs, especially as patients with infections were transferred between facilities. The HL7 data stream now fills these gaps as all information is collected centrally, without the need to access local CHCS data.

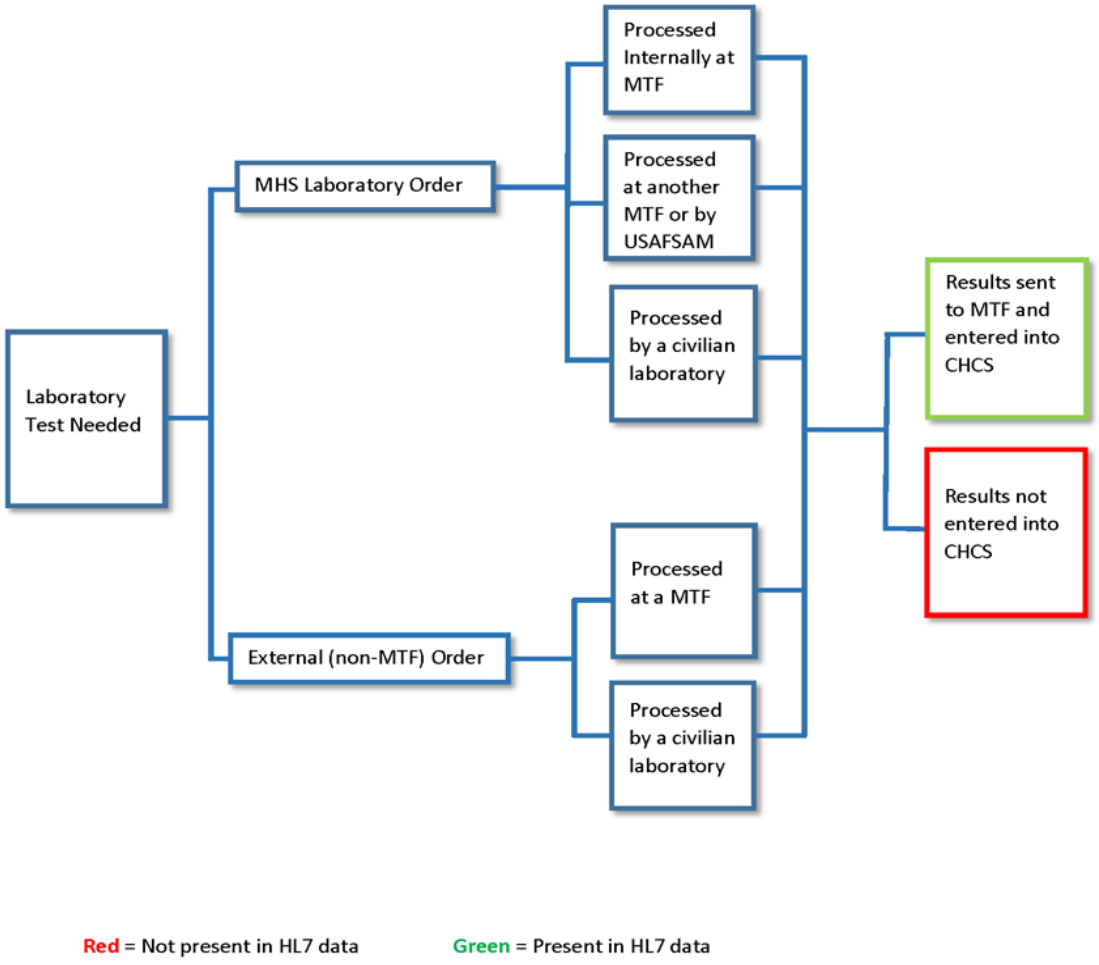
For surveillance purposes, a laboratory result often yields information useful for identifying epidemiologic characteristics of a health event such as the time frame of potential illness, pathogenic agent, patient demographics, and geographic location. Timely, comprehensive surveillance of these characteristics and others may help describe a disease cluster or other issues of concern to leadership, providers, and policy makers in preventive medicine, infection control, and patient safety sectors.

[1] On October 01, 2015, the Military Health System (MHS) transitioned from the Ninth (ICD-9-CM) to Tenth (ICD-10-CM) Revision of the International Classification of Diseases, Clinical Modification. Data remain available, however, caution should be used in comparison of diagnostic data before and after the transition.

## Methods and Procedures

Figure 1 provides an overview of the flow of information involved in the generation of HL7 laboratory data. The process originates at the point of care. When a patient is seen by a healthcare provider in a MTF, the provider can order a laboratory test via local CHCS. Generally, the provider selects the desired test from a pre-populated list of options in CHCS. This list is generated by the MTF laboratory's Management Information Department (MID) based on laboratory tests that can be performed at the facility or at contracted outside laboratories. Laboratory orders performed by non-MHS laboratories should be entered into CHCS when a patient provides the laboratory order slip or when the supporting laboratory communicates results with the requesting MTF. The consistency of this practice is not well-known.

**Figure 1.** HL7 Laboratory Data: Organizations and Exclusions

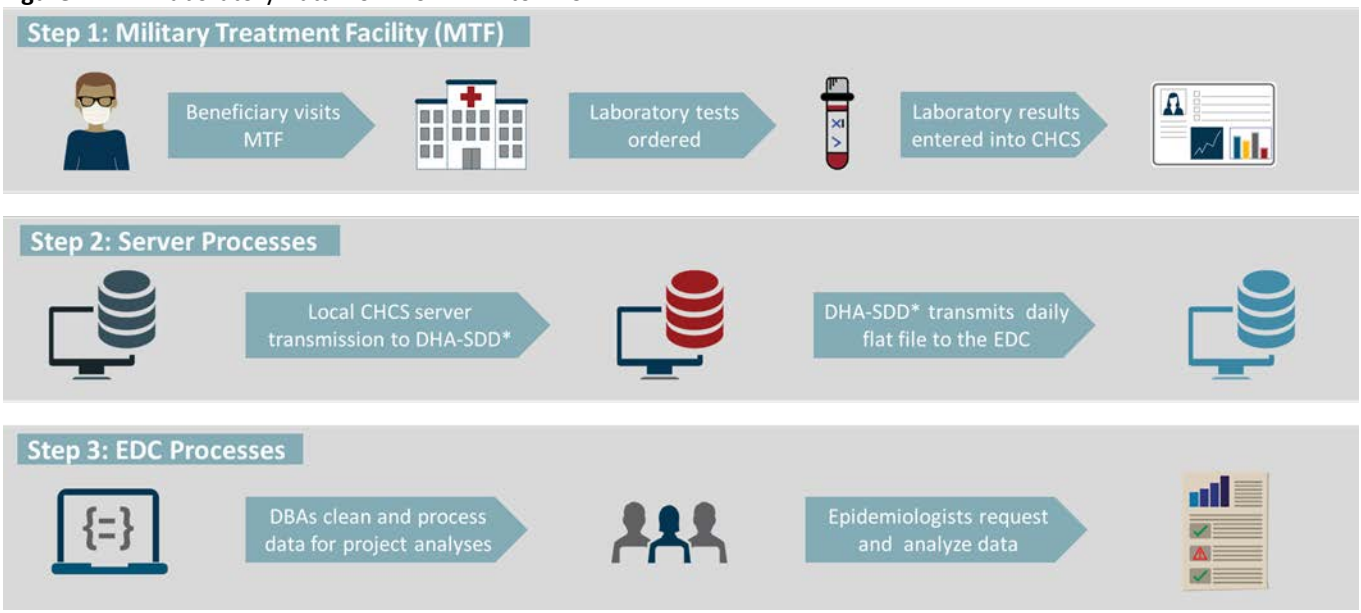


In some circumstances, laboratory tests may not be ordered directly from the provider. For example, the laboratory can initiate another test for a sample based on results of a previous test necessitating confirmation of the disease; this practice depends on standard protocols of the local laboratory. Depending on the patient's status, specimens could be drawn within the hospital ward and sent to the laboratory or could be drawn directly at the laboratory. Most hospitals can do microscopic readings in-house for tests such as malaria and acid-fast bacilli (AFB) smears for tuberculosis. However, MTFs may also outsource culture tests for some conditions. Initial culture growth may be performed at the MTF and then sent out for typing if an organism grows. If a test is outsourced, MTFs are required to enter the results into CHCS for clinical evaluation by the provider. Laboratory results are mapped to their appropriate HL7 database based on a value that defines the type of laboratory setting that performs the test (chemistry, microbiology, pathology, etc). Once labeled, the information is sent to the CHCS database where it is stored based on this value.

Laboratory results are certified by a laboratory technician in CHCS before a script based on system triggers can generate an HL7 message. The HL7 message is archived and batched with other HL7 messages on the local CHCS host. At least once a day, these HL7 messages are forwarded to the central CHCS server. After receipt is verified by the central server, HL7 messages at the local host are deleted. These records are then retrieved by DHA-SDD and parsed into a database design four times a day.

Extracts are retrieved by the EDC using a secure connection to the Defense Health Agency-Solutions Delivery Division (DHA-SDD) main servers. Flat file extracts of the raw, parsed data are received from DHA-SDD on a daily basis. The data are cleaned and true duplicate records are removed. Figure 2 demonstrates this linear process from MTF to EDC.

**Figure 2.** HL7 Laboratory Data Flow from MTF to EDC



# Data Structure and Analysis

## Structure

The EDC receives HL7 microbiology data in a pipe-delimited flat file from DHA-SDD. Table 1 shows the general structure of microbiology data with one entry per line. An individual test can have multiple entries or records as part of a series. For example, names in the TEST ORDERED field are not always disease-specific because of the nature of the tests. While physicians may suspect a specific viral or bacterial etiology, they order a general test for culture (e.g., respiratory or wound culture). As a result, an individual sample can have multiple entries for each test performed. In addition, it is important to note that microbiology data may contain more than culture records and include tests like PCR, though this is unusual.

**Table 1.** Example of HL7 Microbiology Record Sorted by Message (MSG) ID and SET ID

| MSG ID | ACCESSION NUMBER | SET ID | TEST ORDERED | TEST NAME | TEST RESULT |
|---|---|---|---|---|---|
| 999-253 | 123-45-7 | 1 | VIRAL CULT | Flu Cult | |
| 999-253 | 123-45-7 | 2 | BACT CULT | Bact | Neg A |
| 999-253 | 123-45-7 | 3 | BACT CULT | Bact | Pos A |
| 999-253 | 123-45-7 | 4 | | Inf A | |
| 562-235 | 123-45-7 | 1 | CULT | Influenza A&B | |
| 562-235 | 123-45-7 | 2 | | Bact | Neg A |
| 562-235 | 123-45-7 | 3 | | Bact | Neg A |
| 562-235 | 123-45-7 | 4 | | Virus Cult | Neg A |
| 562-235 | 123-45-7 | 5 | | Haemophilus Influenzae | |
| 562-235 | 123-45-7 | 6 | | Bact | No Growth |
| 562-235 | 123-45-7 | 7 | | Bact | Moderate Growth |
| 562-235 | 123-45-7 | 8 | CULTURE | Virus Cult | Haemophilus Influenza-Beta Lactamase Positive |

**Summary of MSG ID 999-253 (Orange):**

- SET ID 1 identifies the test as an Influenza Culture
- SET ID 2 shows no growth of Influenza A at first time interval in the test result field
- SET ID 3 shows growth of Influenza A at second time interval in the test result field
- SET ID 4 identifies the agent that grew as Influenza A in the TEST NAME field

**Summary of MSG ID 562-235 (Blue):**

- SET ID 1 identifies the test for Influenza A and B
- SET ID 2-4 show no growth at first three time intervals in the test result field
- SET ID 5 identifies the second test for the same specimen as *H. influenzae*
- SET ID 6-7 show growth progression at first two time intervals in the test result field
- SET ID 8 identifies growth of *H. influenzae* in the test result field

Multiple entries may also be observed for culture results. Culture results are typically provided at 24 and 48 hour intervals until the final growth is read (usually at the end of a 72 hour growth period), although the growth period may vary depending on disease. Separate records are created for results at each time interval. All of these records will be certified in CHCS and are visible in the HL7 data. These records should be read in chronological order to understand the progress of bacterial or viral growth.

The two examples in Table 1 should be considered separately because one (MSG ID 999-253) indicates a positive test for a particular species, and the other (MSG ID 562-235) describes growth of a culture over time. Methods have been established in the EDC to distinguish between these instances. Records can be grouped together by ACCESSION NUMBER, SPECIMEN SOURCE, and SET ID to group laboratory results for an individual sample.

Results may be corrected after final results are certified. Corrected results are visible in the HL7 microbiology data as separate lines for each corrected result. The hierarchy for selecting a final line entry for analysis purposes is as follows: C (Corrected), F (Final), P (Pending).

## Analysis

Before using HL7 microbiology data for analysis of a particular outcome, it is critical to understand the disease of interest, its symptomatic course, local provider practices with regard to treatment and testing, available tests, and local laboratory testing or outsourcing procedures. It is also important to note that test names and results are not standardized in CHCS or the HL7 microbiology data. Multiple variations of test names and results may be associated with a particular disease, depending on the testing method and recording practices. Misspellings, abbreviations, and variations in spacing, periods, or other punctuation within text should all be considered when conducting a query of particular tests or results in the HL7 microbiology database. Queries may be performed across multiple fields to search for tests and results for a particular outcome. All query results should be thoroughly analyzed throughout the process to assess data completeness and accuracy and to ensure data are classified correctly.

Test orders that are not traditionally considered microbiology tests may be found in HL7 microbiology data for particular MTFs, depending on how the particular laboratory is configured at the local level. MTFs may accession some specimens through the microbiology section of the laboratory, even though the order requests a chemistry test. Therefore, when searching for a particular agent or species, both chemistry and microbiology tables should be examined to ensure full capture of test orders and results.

## Key Fields for Public Health Surveillance

Data were pulled from the HL7 microbiology database for all DOD beneficiaries who received care in the MHS in calendar years 2004 through 31 October 2018 . All data were reviewed and analyzed in order to modify the datasets to more accurately address the disease surveillance needs of the EDC. Methods for identifying duplicate and unique records were 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 value different than all other records in the database.

### Unique Patient/Specimen

*Unique patients* are identified through a combination of SPONSOR ID (sponsor Social Security number [SSN]) and family member prefix (FMP). This unique identifier can be used to track individual patients through all HL7 microbiology records. There is also a field called PATIENT ID (patient SSN); however, the field is not consistently populated and is unreliable as a way to identify patients within or across databases. Another field called EDIPN is a DOD ID number specific to each beneficiary; this field was included in the HL7 microbiology data since 2016, but not all beneficiaries have received an EDIPN. The field is not consistently populated yet, but will replace the SPONSOR_ID/FMP as a way to identify unique patients after implementation is complete. It is possible for individuals to have two different SPONSOR IDs in the database over time. For example, if the child of a sponsor becomes active duty, then that child will have his/her own SSN as the SPONSOR ID instead of the parent's SSN.

A patient will often have multiple samples taken at one time or over a period of time. Each sample, or specimen, would have a different ACCESSION NUMBER, even if many of the other fields are the same. A *unique specimen* is defined using ACCESSION NUMBER, SPONSOR ID, and FMP. Unique patients can also have multiple laboratory orders in the HL7 chemistry data.

The use of ACCESSION NUMBER and TEST NAME, in combination with the SPONSOR ID and FMP, is the most accurate way to identify a *unique test order* for the microbiology database (for chemistry, it is slightly different). Because samples may be continuously tested throughout the day, several records may have the same ACCESSION NUMBER. By using the combination of selected fields to determine a unique test order, the analyst eliminates duplicate records per each patient's ordered tests. Additionally, a person could have multiple samples taken, each with a different ACCESSION NUMBER. The unique test order will determine which test was ordered for a particular specimen per patient. The most recent CERTIFY DATE and CERTIFY TIME for a unique test order generally correspond to the results of the test.

### Test Definition

The TEST ORDERED and the TEST NAME fields are used to identify the laboratory test performed. Both fields may contain non-standardized values, so additional fields and searches of

free text may be needed to capture all tests. Inaccurate reporting could occur if the analyst does not know all parameters of the test prior to analyzing the data. Using influenza as an example, a laboratory test could be defined as taxa-specific (e.g., order, family, subfamily, genus, or species), or based on test type (e.g., antibody staining test or convalescent test). Each test could have a different specimen source (e.g., serum or nasopharyngeal wash), result type (e.g., numeric value or "positive/negative" text), and timeframe that determines each testing method. The outcome of interest determines how the data are organized in the microbiology dataset. It is important to understand the dataset as it relates to a specific project. For many diseases and pathogens, the observations as described above hold true. For others, the observations may be vastly different (where test results are recorded, how test results are recorded, how to determine what test was ordered). For example, influenza tests are recorded differently than tests for other pathogens such as *Acinetobacter baumannii* and *Staphylococcus aureus*.

## Laboratory Test Result

Due to the structure of the laboratory data, results could be identified across multiple fields and records (Table 1). For example, culture results are recorded at time intervals to describe growth patterns, which leads to multiple lines for each test ordered. The TEST RESULT field could show growth per the number of colony forming units (CFU), positive or negative growth or results, the laboratory technician's comments or methods, or null values. Particularly in the HL7 microbiology data, the order of these lines of data is important in interpreting a laboratory result. The SET ID field signifies the line number of an HL7 message, and records should be sorted in descending order for proper analysis.

The actual test result could be found in either the TEST NAME or TEST RESULT field (Table 1). Due to the nature of the test and recording methods, analysis of the HL7 microbiology data will vary when determining case definition for each disease. For records with a result of "Positive," the TEST NAME will typically identify the positive pathogen's name.

The HL7 microbiology data are not optimally arranged to identify a pathogen and its associated antibiotic susceptibility tests. If a message is sorted by MSG ID and SET ID, a culture result is generally organized as follows:

1. The test results of the laboratory pathogen in terms of growth after 24 and 48 hours. This may appear over several rows of data.
2. The final test result determining either no growth or the name of the pathogen. This may appear over several rows of data.
3. The resulting antibiotic susceptibility pattern of the identified pathogen with one antibiotic result per row of data.
4. If present, the name of a second pathogen.
5. The antibiotic susceptibility results for the second pathogen.

Information for each of these segments is recorded differently. Data recorded in the TEST NAME and TEST RESULT fields are different depending on the segment of the message. If several organisms grow from one specimen, the name of the second pathogen usually appears after the

antibiotic susceptibility results of the first pathogen and is followed by its own antibiotic susceptibility results. This pattern continues for multiple organisms identified from one specimen. Many pathogens are not tested for antibiotic susceptibility; in these cases, only the pathogen name appears in the record.

The structure of an HL7 microbiology message allows the system to place the name of an organism (if culture positive) in the TEST NAME field. In these records, the TEST RESULT field will be blank or contain extraneous characters (%, &, etc). Methods can be developed to identify pathogens of interest. In Table 1, there were two tests ordered for one specimen. The first SET ID for each test labels the test performed in the TEST NAME field and does not indicate a result. The subsequent SET IDs below the first SET ID are text fields showing what pending results laboratory technicians have entered during interim growth periods (for example, 24 hour and 48 hour read). An ACCESSION NUMBER can have numerous TEST ORDERED fields associated with it, corresponding to the multiple tests performed per sample.

Results of quality assurance tests are also included in the HL7 microbiology dataset. Such records may include text strings such as *QA*, *QC*, *LIO*, *CAP*, or *INTEROP* in the TEST ORDERED, TEST NAME, TEST RESULT, or CLINICAL COMMENTS fields. Other indicators of quality assurance testing include “ZZZ” or “QQQ” appended to the TEST NAME value. Records of laboratory quality assurance tests are not typically included in analyses for public health surveillance.

## Date/Time References

There are five date and time fields within the HL7 microbiology dataset. Historically, the MSG DATE has been used as part of the criteria to query data from CHCS. The MSG DATE and MSG TIME fields indicate when the data were sent to the central server (fields not described in Figure 3). Data for epidemiologic projects primarily use the ORDER EFFECTIVE DATE to capture the timeframe when the patient is likely to have been symptomatically ill, thus prompting the order of a laboratory test to support the diagnosis. ORDER EFFECTIVE DATE is the date the provider makes the laboratory order effective. It is possible for a provider to enter a laboratory test order to be effective for a future time frame, which may be common practice for chronic or inpatient illnesses. It is also possible that the patient may wait several days before going to the laboratory to have a specimen collected. Therefore, the applicability of using ORDER EFFECTIVE DATE may vary. An alternative date that could be used for analysis is COLLECTION DATE. Figure 3 provides a timeline of dates associated with laboratory tests.

**Figure 3.** Timeline of Events in HL7 Microbiology Record

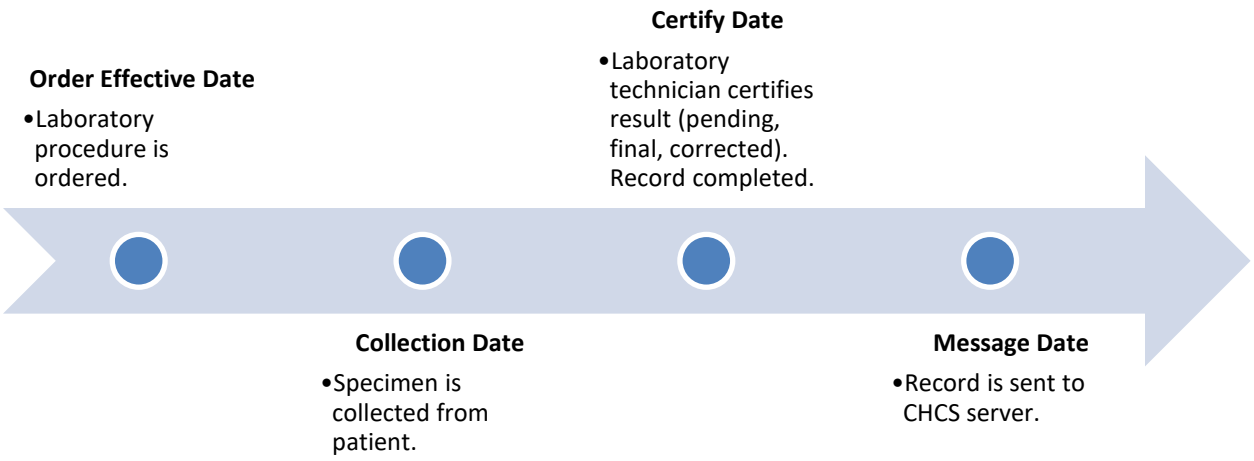



Table 2 reflects methods used to determine timeframes within the HL7 microbiology database. Each date/time pair references one action towards the completion of the laboratory test. A time span was created when comparing one date/time field against another date/time field.

**Table 2.** Methods Used to Determine Timeframes in HL7 Microbiology Data

| Date/Time Range | Interpretation |
|---|---|
| Message – Order_Effective | Determines timeframe from when test was ordered to when an HL7 message was generated 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 date when specimen was extracted |
| Certify – Order_Effective | Timeframe from test ordered to results |
| Certify – Collection | Timeframe from when sample was taken to results |

# Strengths

## Timeliness

Several date and time fields are included in the HL7 microbiology data, including MSG DATE/TIME, ORDER EFFECTIVE DATE/TIME, REQUESTED DATE/TIME, COLLECTION DATE/TIME, and CERTIFY DATE/TIME. To assess the timeliness of the data, the ORDER EFFECTIVE DATE (date the order was placed into CHCS by the provider) was compared to the MSG DATE (date the HL7 message was generated by CHCS) to estimate the time between patient encounter and receipt of data at DHSS. For microbiology data, approximately 94% of all records are available on the DHSS server within one day. This is consistent from 2014 forward.. This indicates that the timeliness of reporting is within acceptable ranges for DON and DOD surveillance activities. Future analysis and assessment may further define lag times in relation to particular MTFs, testing, or outcomes of interest.

## Completeness

Systematic limitations, such as missing records due to atypical data processing or data feed issues, can be adjudicated through collaboration with the MHS Helpdesk to correct issues. HL7 data are required to be sent to the central CHCS server by MTFs in support of Veterans Administration (VA) and other data sharing initiatives. Improvements on completeness and reliability of the data, therefore, are vital in the MHS record-keeping process. Isolates can be tracked MHS-wide from a central location, which is critical for local MTFs as it is difficult to retrieve pathogen data from the local CHCS.

Cultures generally provide a high sensitivity and specificity for laboratory testing. Disease and case burden, therefore, can be determined for many pathogens of interest. It should be noted that culture results may yield pathogens not directly responsible for a patient's symptoms. For example, bacteria isolated from a wound culture may be colonizers and not the infectious agent. Finally, antibiotic resistance and susceptibility results can be reviewed for those pathogens where laboratories have run these tests.

## Organization and Structure

The fields within HL7 microbiology data that are of highest utility for disease surveillance are generally complete and contain expected values. When designing analyses to identify and describe pathogens of interest, analysts must consider factors unique to each pathogen (e.g., misspellings, presentation, etc.) to maximize capture of test results. The EDC has developed several algorithms to restructure the HL7 microbiology data such that rapid analysis of diseases, pathogens, and antibiotic resistance patterns can be completed.

## Public Health Applications

The HL7 microbiology database provides the ability to establish a disease time line 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 microbiology data has led to more accurate estimation of case burden for several diseases of interest.

# Limitations

## Complexity

The layout of the HL7 microbiology database requires familiarity of test methodology and parameters in order to conduct proper analysis. For example, a single test record can be composed of multiple lines of data due to multiple tests within a panel or results at specific time intervals (as with cultures). Customized methods need to be applied to ensure the final test results are captured and that each test is analyzed appropriately.

## Completeness

Incomplete demographic information can limit the ability to describe disease burden demographically. Complementary data, such as personnel records, may be used to supplement HL7 microbiology data for more complete demographic information.

Because MTFs need to be connected to CHCS so that records are captured in the HL7 database, medical laboratory procedures conducted aboard ships or in the field are not captured. Laboratory tests for a service member aboard a ship may be seen in the HL7 databases if 1) the ship-based provider referred the service member to a shore-based facility, or 2) the severity of illness deems the service member to be transported to a shore-based CHCS facility. Shipboard or forward deployed laboratories may also send specimens to shore-based facilities for testing; these tests would be captured in the HL7 database. Additionally, laboratory tests performed outside the MHS are not captured in CHCS unless results are entered manually by laboratory technicians.

## Generalizability

HL7 microbiology data are generated from the medical laboratory records of military service members, their family members, and other beneficiaries. This population differs from the general US population in many ways, including average age, gender distribution, and access to healthcare. Active duty service members also differ from the general population in terms of overall physical fitness and health. Reservists are also not well-captured in HL 7 microbiology data, as they only become visible in the system when they are activated for duty or if they utilize the medical system while in training. These differences limit the comparability of MHS findings to the general US population.

## Inconsistency

The format of CHCS includes free text fields (e.g., TEST NAME, TEST RESULT, RESULT NOTES, CLINICAL COMMENTS), resulting in variations between entries and potential difficulties in data interpretation. Free text fields limit the ability to easily determine testing type, test results, and reference ranges. Non-standardized naming and resulting conventions hinder the ability to use standardized syntax. Methods have been developed to overcome these barriers and flag results of interest. The coding involved in this process can be extensive and requires regular re-evaluation.

Inconsistencies also exist across MTFs. MTF laboratories determine testing capabilities and control what providers can order on the CHCS laboratory order screen. Also, TEST NAME and TEST RESULT options in CHCS (drop down menus) are controlled at the local CHCS host, introducing variation in syntax of test names and results across MTFs. Extensive coding and methodical review ensure comprehensive case capture.. Current efforts are underway to help standardize this process at the MTF level.

## Data Interpretations

While HL7 microbiology data may be useful to identify laboratory-confirmed cases of illness, not all cases will be identified if physicians elect to treat patients presumptively without laboratory confirmation. Clinical practice, 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, and would opt to treat presumptively instead.

# All Data Fields (Variables)

## Automatically Populated Fields

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

When a facility registers within CHCS, several fieldss are automatically 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.

Each DOD beneficiary is registered in the Defense Eligibility Enrollment Reporting System (DEERS) under the SPONSOR ID, which feeds into CHCS. When a patient presents at a medical facility, the SPONSOR ID is entered and the patient's name is chosen from a drop down list of dependents associated with that SPONSOR ID. 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, SERVICE, PATIENT ID, RACE, EDIPN, and SPONSOR ID. If these data are not present in the system, a designated unknown value is entered; therefore, records should not have missing values for these fields. Administrative personnel at the MTF have the ability to edit records 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 generated by the CHCS time-keeping system. For example, this occurs during order entry or certification of a test result. These fields include: ORDER EFFECTIVE DATE/TIME, COLLECTION DATE/TIME, REQUESTED DATE/TIME, and CERTIFY DATE/TIME. These values 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 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).After a physician chooses a test among the TEST ORDERED choices, for example, a variety of options for TEST NAME are presented next. Based on the physician's request, the SPECIMEN SOURCE, BODYSITE COLLECTION SAMPLE, MEPRS CODE, CPT CODE DATA, and NO OF CPT CODES are selected for each TEST NAME. These can be changed if necessary by the laboratory staff.

The laboratory results and each reference 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 specialized 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, and SENSITIVE RESULT FLAG could be automated or manually entered by staff.

## Descriptions of Fields (Variables)

Observations and frequencies below are DOD data. 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 Microbiology Data (n= 133,328,250 records as of 31 October 2018)

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| ACCESSION NUMBER | Automatically populated by CHCS | Combination of:<br>1) collection date in YYMMDD format<br>2) two or three alpha characters<br>3) a numeric value | Last numeric digits can range from 1 to 9999. | 0%<br>(N = 86) | Created for each unique biological specimen collected from a patient. Different microbiology tests from the biological specimen can have the same ACCESSION NUMBER. These numbers could be recycled throughout the day; therefore, **ACCESSION NUMBER should not be used on its own to identify a unique record**. ACCESSION NUMBER may be used to determine tests ordered for a patient in conjunction with the SPONSOR ID, FMP, and TEST ORDERED. |
| BODYSITE COLLECTION SAMPLE | CHCS host or MTF may offer drop-down lists or lookup tables; staff can edit if desired | Character | Example: Urine^urine cath | 0%<br>(N = 19,987) | Refers to where the specimen was taken from the patient. This field is associated with the SPECIMEN SOURCE to determine specifically where the sample was taken. |
| CERTIFY DATE | Automatically populated by the CHCS clock; staff can edit but rarely do | YYYYMMDD | Any valid date | 0%<br>(N = 7,295) | Date when a laboratory technician certified the results into CHCS or certified changes to the results. Unlike the ORDER EFFECTIVE DATE, there may be differences between the values for CERTIFY DATE for each SET ID because there are differences in when tests are performed and results are available. The field does not have missing values and contains all valid dates. The value of CERTIFY DATE should be between ORDER EFFECTIVE DATE and MSG DATE. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| CERTIFY TIME | Automatically populated by the CHCS clock; staff can edit but rarely do | HHMM (Standard 24-hour clock) | 0000 – 2359 | 0% (N = 5,314) | The field represents the time component of the CERTIFY DATE formatted within a 24 hour cycle. All times are valid entries. |
| CLINICAL COMMENTS | Manually entered by provider or lab technician | Free text | Free text | 83% (N = 111,846,980) | Allows the provider or laboratory technician to add additional information regarding the patient's symptoms, contact phone numbers, specimen media, or instructions on procedures for a test. Quality assurance tests may also be identified using the CLINICAL COMMENTS field. Records containing text strings such as *QA*, *QC*, *CAP*, *LIO*, or *INTEROP* all indicate quality assurance tests and should be removed from analysis. |
| COLLECTION DATE | Automatically populated by the CHCS clock; staff can edit but rarely do | YYYYMMDD | Any valid date | 0% (N = 5) | Date when the specimen is taken from the patient. This value should be between ORDER EFFECTIVE DATE and the CERTIFY DATE. All records had valid dates. |
| COLLECTION TIME | Automatically populated by the CHCS clock; staff can edit but rarely do | HHMM (Standard 24-hour clock) | 0001 – 2400 | 0% (N = 5) | The field represents the time component of the COLLECTION DATE formatted within a 24 hour cycle. All records had valid times and there are no blank entries. |
| CPT COUNT | Provider can select from drop down menu or look up table; staff can edit | Numeric | Dependent on number of CPT codes listed in CPT LIST field | 2% (N = 3,146,265) | Lists the number of CPT codes used for each test performed. Over three-quarters of the entries only have one CPT code associated with them. The number of CPT codes is determined at each regional location. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| CPT LIST | Provider can select from drop down menu or look up table; staff can edit | Numeric | 00832\00\AD#87088\00\AD#87186\00\AD#87077\00\AD <--> 99499\00\AD#86480\90\AD | 2% (N = 3,142,893) | 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. |
| DHSS LOAD DATE | Automatically populated by the CHCS clock; staff can edit but rarely do | YYYYMMDD | Any valid date | 41% (N = 54,884,729)<br><br>Available since 2009 | Derived from the date portion of the DATE/TIME field that identifies when the Message Header content was loaded into the HL7 Object Delivery System (ODS) at DHSS. 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 | 41% (N = 54,884,729)<br><br>Available since 2009 | Derived from the date portion of the DATE/TIME field which identifies when the Message Header content was loaded into the HL7 ODS at DHSS. Time component of the DHSS LOAD DATE field. |
| DATE OF BIRTH | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | YYYYMMDD | Any valid date | 0% (N = 337) | Required field within CHCS. There are no blank values and limited false dates. If only the year is known, 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; therefore, there are no missing values. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| EDIPN | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Numeric value with 10 digits<br><br>########## | See notes | 3% (N = 4,889,489)<br><br>Available since 2015 | 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 2015, but not all beneficiaries have received an EDIPN. The field is not consistently populated to date, but will replace the SSN/FMP for identification of unique patients after implementation is complete. |
| ETHNICITY | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Alphanumeric | **Six (6) possible values:**<br>1 – Hispanic<br>2 – Southeast Asian,<br>3 – Filipino<br>4 – Other Asian Pacific Islander<br>9 – Other<br>Z – Unknown | 13% (N = 18,459,103)<br><br>No missing values prior to 2014 | In 2018, the majority of records indicated Other for ETHNICITY (51%), followed by Unknown (31%). Thirteen percent were missing. The most frequent value other than Unknown or Other was Hispanic (4%). These results indicated that ETHNICITY is not consistently reported and may be self-identified, which limits the ability to identify disease trends in minority groups. |
| FMP | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Numeric value with two digits<br><br>## | See Notes | 13% (N = 18,459,103) | Designates the relationship of the patient to the sponsor. In 2018, 46.8% of records had an FMP of 20 (sponsor) followed by 32.3% with an FMP of 30 (spouse of sponsor). Other possible values include 01-19 (child of sponsor, numbered in age order). Thirteen percent missing an FMP value. Unknown entries are labeled as 99.<br>List of values and % populated is available upon request. |
| GENDER | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Single alpha character<br><br>X | **Three (3) possible values:**<br>M – Male<br>F – Female<br>X – Unknown | 13% (N = 18,459,560)<br><br>No missing values prior to 2014 | As of 2018, 37% of records indicated male and 62% indicated female as gender. 0% were unknown. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| MARITAL STATUS | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Single alpha character<br><br>X | **Nine (9) possible values:**<br>A – Annulled<br>D – Divorced<br>I – Interlocutory Decree<br>L – Legally Separated<br>M – Married<br>N – Never Married<br>S – Single/Not Married<br>W – Widow/Widower<br>Z – Unknown | 22%<br>(N = 29,989,633)<br><br>No missing values prior to 2014 | As of 2018, 38% of records had a MARITAL STATUS of Married, followed by 28% of records with Unknown and 18% of records with Single/Not Married. Twenty-two percent were missing. |
| MEPRS CODE | CHCS host or MTF may offer drop-down lists or lookup tables; staff can edit if desired | Four alpha characters<br><br>XXXX | The first letter indicates the general areas:<br>A – Inpatient<br>B – Outpatient<br>C – Dental<br>D – Ancillary<br>E – Support Services<br>F – Special Programs<br>G – Medical Readiness | 0%<br>(N = 1,139) | The MEPRS (Medical Expense Performance Reporting System) CODE is a four letter code that indicates where within a MTF the patient was seen when the sample was collected. The entire MEPRS CODE is used to indicate the specific unit or ward, such as Family Practice, General Surgery, or Pediatrics. The MEPRS CODE field is useful for tracking where people were seen within the MTF, which can affect the interpretation of the data. The majority of records in the HL7 microbiology dataset have a MEPRS CODE that begins with B (Outpatient). The most frequent MEPRS code in 2011 HL7 microbiology data was BIAA (22.6%), indicating that the patient was seen in the Emergency Department. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| MSG DATE | Automatically populated when the message (record) is sent to the CHCS server | YYYYMMDD | Any valide date | 0%<br>(N = 1,139) | Date when records are sent from the local CHCS to the central server. There are no missing values and all are valid dates, and all dates are either the same date or after the CERTIFY DATE. Some MTFs send messages in batches, therefore, **the time or date portions may not correlate to the actual transaction time.** |
| MSG ID | Automatically populated when the message (record) is sent to the CHCS server | Alphanumeric<br><br>Varies by MTF and may include numbers, letters, numeric code that identifies the MTF, or the function of the message (e.g., RESCHED-057342). | See notes | 0%<br>(N = 1,139) | The MSG ID is an alphanumeric code assigned to each batch of messages based on when the message is sent from CHCS to the central 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 the CHCS server | Alphanumeric<br><br>Four possible formats:<br>A####<br>F####<br>HP####<br>N##### | See notes | 0%<br>(N = 7,296)<br><br>No missing values after 2007 | This field allows analysts to identify and track problems that may arise in the transfer of messages from the MTFs through DHSS to the EDC. There are no blank values. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| MSG TIME | Automatically populated when the message (record) is sent to the CHCS server | HHMM | 0001 - 2359 | 0%<br>(N = 7,296)<br><br>No missing values after 2007 | MSG TIME is the time component of the MSG DATE formatted within a 24 hour cycle. This field has four numeric characters, and there were no blank values. |
| ORDER EFFECTIVE DATE | Automatically populated by the CHCS clock; staff can edit but rarely do | YYYYMMDD | Any valid date | 0%<br>(N = 471) | Date that the laboratory order enters the CHCS system and indicates when the laboratory test was actually ordered. All entries are valid dates, and there were no blank values.<br><br>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 the CHCS clock; staff can edit but rarely do | HHMM | 0000 – 2400 | 0%<br>(N = 471) | ORDER EFFECTIVE TIME represents the time component of the ORDER EFFECTIVE DATE formatted within a 24 hour cycle. All times are valid entries. |
| ORDER NOTES COMMENTS | Populated by provider | Free text | All records have blank values | 100%<br>(N = 133,328,250) | Allows the provider to provide notes or comments that accompany the test ordered. All records have blank values in this field. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| ORDER NUMBER | Automatically populated by CHCS | Numeric code of 11 digits<br><br>######-#####<br><br>First six numbers are the date<br><br>Last five numbers are consecutive per the location | Range: 000109-00327 – 991227-01964 | 0%<br>(N = 476) | 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 | Alpha characters, three components separated by commas<br><br>Last Name, First Name, Middle Initial | Last, First, I | 0%<br>(N = 86) | 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<br><br>or<br><br>PATCAT CODE | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Alphanumeric<br><br>X##<br><br>First letter refers to the sponsor's service branch affiliation<br><br>Following 2 digits correspond to the status of the sponsor within the service and the patient's relationship to the sponsor | **Nine (9) possible values for sponsor branch:**<br>A – Army<br>B – National Oceanic and Atmospheric Administration<br>C – Coast Guard<br>F – Air Force<br>K – Other beneficiaries of the federal government<br>M – Marine Corps<br>N – Navy<br>P – US Public Health Service<br>R – NATO recipient | 0%<br>(N = 193,121) | Indicates the patient's relationship to the Uniformed Services.<br><br>Examples: M11=Marine Corps Active Duty Service Member, A31=Army Retired Active Duty, N41=Navy Dependent of Active Duty.<br><br>**A complete list of PATCAT codes and % populated is available upon request.**<br><br>In 2018, the most frequent PATCAT was A41 (Army Dependent of Active Duty) (15.9%) followed by A11 (Army Active Duty Service Member) (9.5%). Less than 1% of records had missing values for this field. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| PATIENT ID | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Nine digit numeric value<br><br>(Patient SSN when available, but not reliable) | See notes | 0%<br>(N = 61,882) | 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.**<br><br>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 | Text | See notes | 12%<br>(N = 16,989,080)<br><br>Not consistently completed until 2015 | Text translation of the PERFORMING DMIS ID field and is assigned by DHSS. Because the field is a translation of PERFORMING DMIS ID, it will be missing when that field is missing in the record. This field was missing in 0.01% of records in 2018. |
| PERFORMING DMIS ID | Automated from DHSS | Four numeric digits, import to SAS as character<br><br>#### | See notes | 19%<br>(N = 26,030,526)<br><br>Not consistently completed until 2006 | 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 installationsThis field was missing 0.01% of records in 2018. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| PERFORMING FACILITY SERVICE | Automated from DHSS | Single alpha character | **Three (3) possible values:**<br>A – Army<br>F – Air Force<br>N – Navy<br>P - DHA | 13% (N = 17,897,025)<br><br>This field was missing 2.7% of values between 2015 and 2018 | Indicates the branch of service with which the MTF is associated. This value is determined from the DMIS ID code list provided to DHSS by the EDC. It will be missing from a record when the PERFORMING DMIS ID is missing. This field is useful for limiting observations to those specific to a service, allowing for analysis of disease burden in facilities of that particular service. |
| PERFORMING FACILITY NAME | Automated from CHCS | Text field | See notes | 6% (N = 8,172,038)<br><br>Not consistently completed until 2006 | Indicates the name of the MTF where the test was performed. Problems may be encountered if the text was entered incorrectly when the facility was registered in the system or due to inaccurate mapping between DMIS ID and facility name. |
| PERFORMING WORK CENTER NAME | Automated from CHCS | Unstructured text | See notes | 0% (N = 167) | Indicates the work center within the MTF that provided the testing service. This field is a relatively unstructured text field with many possible values. These locations are usually laboratories mapped according to the PERFORMING DMIS ID. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| RACE | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Single alpha character | **Six (6) possible values:**<br>C – White<br>M – Asian or Pacific Islander<br>N – Black<br>R – American Indian or Alaskan Native<br>X – Other<br>Z – Unknown | 8%<br>(N = 10,717,994)<br><br>Approximately 23.6% missing values between 2014 and 2018 | Twenty-nine percent % all of records were categorized as Unknown, followed by White (36%) and Other (12%). A high frequency of Unknown or Other values limits the ability to use the data to look at diseases or conditions according to race. |
| RECORD TYPE | Automatically populated from CHCS | Three alpha characters; a standard value (LMI) | **One (1) possible and standard value:**<br>LMI (microbiology laboratory) | 8%<br>(N = 10,717,994) | Identifies the database that the records are from (e.g., microbiology, chemistry). There are no blank values for this field and the field is three characters in length. All HL7 microbiology records have a value of "LMI" (for microbiology laboratory) in this field. |
| REQUESTED DATE | Automatically populated by the CHCS clock; staff can edit but rarely do | YYYYMMDD | Any valid date | 0%<br>(N = 583) | This field is not frequently used for data analysis, and does not have missing or invalid values. The timeframe of this value should be between ORDER EFFECTIVE DATE and COLLECTION DATE. |
| REQUESTED TIME | Automatically populated by the CHCS clock; staff can edit but rarely do | Four numeric digits<br><br>#### | 0000 – 2359 | 0%<br>(N = 5) | The field represents the time component of the REQUESTED DATE formatted within a 24 hour cycle. 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 | See notes | 6% (N = 8,955,024)<br><br>Not consistently completed until 2015. | Text translation of the DMIS ID provided in the REQUESTING DMIS ID field and is assigned by DHSS, although this may be done inconsistently. This field was missing in 12.0% of records in 2011. |
| REQUESTING DMIS ID | Automated from DHSS | Four numeric digits<br><br>#### | See notes | 6% (N = 9,216,479) | Identifies the MTF that requested the laboratory test. This code allows for grouping of MTFs based on geographic location, as well as the ability to identify parent/child relationships between installations. |
| REQUESTING FACILITY NAME | Automated from DHSS | Text | See notes | 0% (N = 4,057) | Indicates the name of the MTF where the order originated. Problems may be encountered if the text was entered incorrectly when the facility was registered in the system. The field allows tracking of orders from origin to where they were performed. |
| REQUESTING FACILITY SERVICE | Automated from DHSS | Single alpha character | **Three (3) possible values:**<br>A – Army<br>F – Air Force<br>N – Navy<br>P - DHA | 7% (N = 9,866,883)<br><br>Approximately 17% of values are missing between 2011 and 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 will be missing from a record when the REQUESTING DMIS ID is missing. This field is useful for limiting observations to those specific to a service, allowing for analysis of disease burden in facilities of that particular service. |
| REQUESTING WORK CENTER NAME | Automated from DHSS | Unstructured text | See notes | 0% (N = 452) | Indicates the ward or clinic within the MTF that requested the laboratory test. This field is an unstructured text field with many possible values. Possible entries include 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 | Free text | | 99% (N = 133,328,213) Available since 2010 | The RESULT NOTES field is a character string that allows the laboratory technician to provide additional information about the result, a recommendation for additional testing, or an interpretation of the laboratory result. This value is either from an automatic/drop down menu that is created when the TEST NAME is selected via CHCS or free-text. A frequency 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, genus type, or the address where the sample is tested.<br><br>The RESULT NOTES field is useful when the TEST RESULT field does not indicate the outcome of a test (e.g., 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 that limits the use of wildcards and searches. Nearly 99% of records are blank in this field, which 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. |
| RESULT SENSITIVITY | Automated via lab equipment OR entered manually | Singl\e alpha character | **Three (3) values:**<br>R – Resistant<br>I – Intermediate<br>S – Susceptible | 77% (N = 103,847,983) | The value is based on the numeric antibiotic susceptibility result in the TEST RESULT field for the corresponding antibiotic in the TEST NAME field. |

| 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) values:**<br>P – Preliminary<br>F – Final<br>C – Correction | 0%<br>(N = 7,456) | RESULT STATUS OBX is a free-text field that shows the status of the test performed. There are three entries which could be used, and always go in the following consecutive order: P (Preliminary), F (Final), and C (Correction). Each SET ID for a test should have a status of F. Records with a status of P or C for a test may also be present, either along with or in the absence of F records. Should a test have numerous RESULT STATUS OBX values, each record will have the same SET ID, TEST NAME, and TEST ORDERED values. If a SET ID has multiple test statuses available, the following hierarchy should be used to determine which record to use: C→F→P. An entry is corrected (C) when it is amended due to a change in interpretation, operator error, wrong test ordered, or test was performed under wrong patient. The value of "F" was the majority listed within the dataset, followed by "P", and then "C." |
| SENSITIVE RESULT FLAG | Automated via regional CHCS location or lab equipment OR entered manually | | | 99%<br>(N = 133,286,091) | Permits the laboratory technicians to record results-dependent codes for classifying the observation in CHCS. Nearly all records had missing values for this field. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| SERVICE | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Single alpha character | **Nine (9) possible values:**<br>A – Army<br>B – National Oceanic and Atmospheric Administration<br>C – Coast Guard<br>F – Air Force<br>K – Other beneficiaries of the federal government<br>M – Marine Corps<br>N – Navy<br>P – US Public Health Service<br>R – NATO recipient | 0%<br>(N = 159,291) | Refers to the service branch of the Sponsor.<br>In 2018, 41.5% of records had a value of A (Army) for SERVICE, followed by F (Air Force) (21.7%) and N (Navy) (20.1%). No records had missing values for this field. |
| SET ID | Automated from DHSS | Numeric | 1 – 99 | 0%<br>(N = 5) | Mandatory field populated by CHCS.The numbers show the logical order of arrival of data within an HL7 message. Should an entry have a change in its resulting status (from pending to final, or final to correct), the SET ID will remain the same for that test entry. |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| SPECIMEN SOURCE | CHCS host or MTF may offer drop-down lists or lookup tables; staff can edit if desired | Text field | | 0%<br>(N = 54,817) | 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. |
| SPONSOR ID | Automatically populated from DEERS after user enters Sponsor ID; MTF staff can edit | Nine (9) numeric digits with no dashes<br><br>######### | Sponsor's SSN with no dashes<br><br>Pseudo SSNs **may** begin with 800 or 900, followed by a date, or begin with 801 or 901 | 0%<br>(N = 10,528) | **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.** Only 540 of 8 million records in 2018 had blank values for SPONSOR ID. It is important to preserve the entire SPONSOR ID when importing the data into SAS or other analysis programs. The SPONSOR ID field needs to be imported as a character field so that leading zeros are not dropped.<br><br>Not all SSNs are ones given by the Social Security Administration. If the patient does not hold a valid SSN, a pseudo SSN number is created. The pseudo SPONSOR ID may begin with 800 or 900, followed by a date. If the number is already assigned to another patient, it could begin with 801 or 901. Also, records for quality assurance/control tests conducted in the laboratory may use 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, QC, Quality Control, PSR, CAP, or Non-human (NH, #). |

| Field Name | Input Method | Field Format | Values | Percentage of Missing Values % (N) | Notes |
|---|---|---|---|---|---|
| TEST NAME | CHCS host or MTF may offer drop-down lists or lookup tables; staff can edit if desired | Text | | 0% (N = 10,533) | The TEST NAME is a text field showing which test was used per the samples provided. This value is usually selected from a pull-down listing from the TEST ORDERED field. This field will never have missing values because TEST NAME is automated by the regional CHCS system. The TEST NAME includes entries such as tests to be performed, quality controls, temperature, and status of culture growth. When TEST NAME contains the name of an organism, this indicates that the test had a positive result for that organism. Quality control tests may also be viewed in this field and will often have "ZZZ" prior to the actual test name.<br><br>The variance between test names suggests the fields are automated by a regional CHCS system, not the main location. A test procedure can be specific or general. A test name can refer to the procedure type, such as convalescent, cultures, or antibody testing. Therefore, a test name could reflect what an expected result should be. For example, results for a titer should be numeric, while results for organism-specific tests can be positive or negative or have sub-typing. |
| TEST ORDERED | CHCS host or MTF may offer drop-down lists or lookup tables; staff can edit if desired | Text | | 0% (N = 10,538) | The TEST ORDERED 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 CHCS regions. 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 regional CHCS location or lab equipment OR entered manually | Alphanumeric | | 27% (N = 36,649,636) | TEST RESULT is an alphanumeric field which shows the results of a test. The TEST RESULT field can have positive or negative results, control values, dates, reorders, references to comment fields, growth status, or that the test was not performed due to inadequate results or insufficient quality. In 2018, 27% of records had blank values for TEST RESULT. A blank entry is typically observed when either a result or test identification was seen in the adjacent TEST NAME field. There are a wide variety of values in this field, including misspellings and slang language, indicating that the results are either automotive per each regional CHCS location or entered manually. Many of these variations show the same result formatted differently, such as Positive, POSITIVE, POSTIVE, POS, and so forth. Currently, CHCS is in the process of regulating the regional CHCS locations to create one specific text for each TEST NAME outcome. This would limit the variation of TEST RESULT significantly. |

## Abbreviations

| | |
|---|---|
| AMA | American Medical Association |
| CBC | Complete Blood Count |
| CFU | Colony Forming Units |
| CHCS | Composite Healthcare System |
| CPT | Current Procedural Technology |
| DHSS | Defense Health Services System |
| DMIS | Defense Medical Information System |
| DHA | Defense Health Agency |
| DHA-SDD | Defense Health Agency Solutions Delivery 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, 9th Edition, Clinical Modification |
| MDRO | Multi-drug resistant organism |
| MEPRS | Medical Expense Performance Reporting System |
| MID | Management Information Department |
| MHS | Military Health System |
| MSG ID | Message Identification |
| MTF | Military Treatment Facility |
| NMCPHC | Navy and Marine Corps Public Health Center |
| PCR | Polymerase chain reaction |
| SSN | Social Security Number |
| USAFSAM | U.S. Air Force School of Aerospace Medicine |
| VA | Veterans Administration |

## Example Projects using HL 7 Microbiology Data

| Project Name/Description | Population of Interest | Purpose | Impact |
|---|---|---|---|
| **Daily Reporting** Identification of reportable multi-drug resistant organisms (MDROs) from microbiology data | DOD Beneficiaries | Identify antibiotic resistance among DoD beneficiares in near real-time | Identification of disease outbreaks among specific populations; identify disease trends; evaluate gaps in infectious disease reporting |
| **Antibiograms** Reporting of bacterial isolate susceptibilities to a variety of standard antibiotics | DOD Beneficiaries | Calculate antibiotic susceptibilities based on MTF testing results of bacterial isolates | Clinical tool used in MTFs to guide treatment decisions and to support stewardship |
| **Dashboard** Display of routine reporting results by MTF location for MDROs, antibiotic susceptibilities, and antibiotic consumption | DOD Beneficiaries | Improve ease and timeliness of data access for customers | Support antimicrobial stewardship efforts across MHS and remove barriers for MTFs in accessing timely data to inform prevention and treatment |
| **MTF-Specific Surveillance** Langley urine cultures, Carl A. Darnall MDRO validation, Portsmouth cystitis | DOD Beneficiaries | Investigate special requests for MTFs; validate data collection methods and reporting for the EDC | Reinforces customer confidence in data, helps customers understand infection burden, prescribing practices, and testing practices within their MTF |
| **Other Surveillance Projects and Taskers** Nitrofurantoin consumption, enteric bacteria surveillance and antibiogram | DON and/or DOD Beneficiaries | Describe special populations or practices of interest using HL 7 data; apply surveillance data to new areas of interest | Expand the use of HL 7 data for surveillance purposes and identify new strengths, weaknesses, uses, and future directions for data |