

Exposures, Risk Factors, and Underlying Medical Conditions

This document was developed by the Data Standardization Work Group (DSWG) of the Council of State and Territorial Epidemiologists (CSTE) Surveillance Practice and Implementation Subcommittee (SPIS) as a part of a larger data harmonization effort. It will be published and maintained in a repository on the CSTE website.

More information about the DSWG and data harmonization can be found here: [link to DSWG Core Surveillance Project Brief (pending)].

Background

Exposures generally capture information on the specific or potential settings, events, or activities in which the individual may have acquired a condition, including the location and dates of exposure. Risk factors and underlying medical conditions both place an individual at higher risk for acquiring a condition or developing a more severe illness. Risk factors typically refer to factors that are behavioral, social, or situational, while underlying conditions are more specific to personal medical or health-related conditions.

Despite these general definitions, in practice the boundary lines between the three data classes are blurry. One contributing factor to the difficulty in distinguishing these is that their categorization in Message Mapping Guides (MMGs) is variable. Depending on the condition, it may be appropriate for the same value to be considered as an exposure in one condition, a risk factor in another, and perhaps even an underlying condition in yet another condition. A standardized definition for each concept would facilitate the application of the appropriate categorization to each data class. For example, the following elements are included in MMGs as different data classes for different conditions:

- Residence in a long-term care facility: Captured as a risk factor in the Tuberculosis (TB) MMG but as an exposure in the contact section of the Hepatitis B, acute MMG.
- Diabetes is captured as an underlying condition in the Respiratory and Invasive Bacterial Disease (RIBD) MMG, but as a risk factor in the COVID-19 MMG.

The message structure used to capture these data classes also varies by condition. Repeating groups were developed and incorporated into several MMGs to standardize the format of information transmission to CDC. One example is the Underlying Conditions (or “Underlying Causes”) Repeating Group, shown below in Figure 1 as used in the Carbon Monoxide MMGs. This repeating group structure allows for the same data element to be used in multiple MMGs with different value sets relevant to the specific conditions in the MMG are included in the associated value set.

Figure 1. Underlying Conditions Repeating Group data elements from the Carbon Monoxide (CO) MMG

Data Element (DE) Name	DE Identifier Sent in HL7 Message	Data Element Description	Data Type	Value Set Name (VADS Hyperlink)
START: Underlying Conditions Repeating Group				
Underlying Condition(s)	INV236	Select the patient's preexisting condition(s).	Coded	Underlying Conditions (CO)
Underlying Conditions Indicator	INV662	Indicator for underlying condition(s)	Coded	Yes No Unknown (YNU)
END: Underlying Conditions Repeating Group				

On the other hand, some MMGs use individual data elements to capture these data elements. For example, the Babesiosis MMG includes the elements in Figure 2 to collect various underlying conditions that are relevant to a case of babesiosis. Individual data elements are helpful in capturing additional information beyond a Yes/No/Unknown response, such as date or cause.

Figure 2. Individual data elements related to underlying conditions from the Babesiosis MMG.

Data Element (DE) Name	DE Identifier Sent in HL7 Message	Data Element Description	Data Type	Value Set Name (VADS Hyperlink)
Asplenic	300564004	Is the subject asplenic?	Coded	Yes No Unknown (YNU)
Splenectomy Date	82760-0	Date of subject's splenectomy	Date	N/A
Splenectomy Reason	82759-2	Why was the subject's spleen removed?	Text	N/A
History of Previous Illness	161413004	Did the subject have a history of babesiosis?	Coded	Yes No Unknown (YNU)
Patient Immunocompromised	370388006	At the time of diagnosis, was the subject immunocompromised?	Coded	Yes No Unknown (YNU)
Immunocompromised Associated Condition or Treatment	INV933	If the subject was immunocompromised, what was the associated condition or treatment?	Text	N/A

In some cases, such as risk factors in the TB MMG shown below, a combination of a repeating group and additional individual data elements is used to collect all of the desired information.

Figure 4. Risk Factor Repeating Group and individual data elements related to risk factors from the Tuberculosis MMG

Data Element (DE) Name	DE Identifier Sent in HL7 Message	Data Element Description	Data Type	Value Set Name (VADS Hyperlink)
START: RVCT & TBLISS Risk Factor Repeating Group				
Patient Epidemiological Risk Factors	INV1117	Exposed risk factors for the patient - Please provide a response for all risk factors in the value set with an associated indicator.	Coded	Epidemiological Risk Factors (TB)
Patient Epidemiological Risk Factors Indicator	INV1118	Provide a response for each value in the patient epidemiological risk factors value set.	Coded	Yes No Unknown (YNU)
END: RVCT & TBLISS Risk Factor Repeating Group				
START: RVCT & TBLISS Additional TB Risk Factors Questions				
Type of Correctional Facility	INV1119	If patient was a Resident of Correctional Facility at Diagnostic Evaluation, indicate the type of correctional facility.	Coded	Type of Correctional Facility
Type of Long-Term Care Facility	INV1120	If patient was a Resident of Long Term Care Facility at Diagnostic Evaluation, indicate the type of long term care facility.	Coded	Type of Long Term Care Facility
Smoking Status	72166-2	What is the patient's current tobacco smoking status?	Coded	Smoking Status
Patient lived outside of US for more than 2 months	INV1121	Residence or Travel in countries other than the United States, Canada, Australia, New Zealand, or countries in northern or western Europe for >60 consecutive days at any point in the patient's lifetime.	Coded	Yes No Unknown (YNU)
END: RVCT & TBLISS Additional TB Risk Factors Questions				

The above examples highlight how the same data may be structured for collection and transmission in different ways, a variation that is seen with all three data classes. The lack of standardization in the definitions, value sets, and format of data elements results in jurisdictional difficulty related to data collection, data transmission, and implementation of MMGs.

In discussions surrounding these concepts, the DSWG considered the following questions:

- Are these three data concepts actually distinct? If so, how?
- Which of these concepts should be considered part of core surveillance for routine case notifications to CDC?
- What is the ideal message structure? What are the pros and cons of the repeating group structure when compared to individual data elements?

Core Surveillance Data on Exposures, Risk Factors, and Underlying Health Conditions in Routine Case Notifications

The workgroup recommends that the exposure, risk factor, and underlying health condition (ERU) data classes be considered part of core surveillance for case notifications to CDC. When polled, there was consensus that these data classes are three distinct concepts and should be defined and sent separately from one another. Table 1 lists the data concept definitions which were strongly supported by the group. The key distinctions between the data classes are as follows:

- Exposure includes a relevant time window, which can be different depending on the condition.
- Risk factors impact acquiring infection or developing disease.
- Underlying health conditions impact the course of infection or health condition.

While not every condition will require that data for all three data classes be sent, it was agreed that a core message format should include placeholders for each. Further details on eCR, ELR, and USCDI as input data sources on exposures, risk factors, and underlying medical conditions can be found in Appendix A.

Table 1. Data Class Definitions – Exposure, Risk Factor, Underlying Health Condition

Data Class Definitions	
Exposure	Within exposure window, either the person's contact with an entity (person, animal, or substance) or presence at a location where exposure to an agent could have occurred (i.e., where did pathogen come from)
Risk Factor	Characteristics, behaviors, or social/environmental factors that increase the likelihood of developing a disease or condition.
Underlying Health Condition	Pre-existing health conditions that may worsen the course or severity of the disease or health condition.

How individual data elements are categorized into the three data classes will be condition-specific and at the discretion of the CDC program areas. Jurisdictions will not be required to attempt to categorize data elements on their own for transmission to CDC. There may be cases when similar data elements are categorized into different data classes; however, these standard definitions should increase CDC program alignment and limit these occurrences.

In addition to defining the three data classes, the workgroup also recommends the use of summative data class indicators for each data class. These data elements would indicate if the person had a positive answer for any of the listed exposures, risk factors, or underlying health conditions. This type of data element may enable some jurisdictions to simplify how they build the case notification messages, as well as providing a simple Yes/No/Unknown data element for each of the three data classes which could simplify data analysis for many users. These data elements are summarized in Tables 2-4.

Table 2. Core Surveillance Data Element for Routine CDC Case Notification – Summative Exposure Indicator

Exposure Data Class Indicator	
Definition	Indicator that the person experienced any listed exposure associated with the condition being reported
Standard Code for Transmission to CDC	TBD
Value Set or Coding System for Transmission to CDC	Yes/No/Unknown
Abstracted/ Derived/ Created	Derived
Derived from (examples)	If a person has any listed exposures associated with the condition being reported, the response would be “Yes”. If the person does not have any of the listed exposures, this response is “No”. If this is not known, the response is “Unknown”.
Recommendations for Input Data Source(s)	Case Investigation eCR
Notes	This data element is a summative indicator that communicates if the person has any exposure associated with the condition as listed in the Exposure value set.
Data Source Details	
eCR CDA	N/A
eCR FHIR	N/A
ELR	N/A
USCDI v4	N/A
USCDI+ Public Health: Case Reporting Use Case	N/A

Table 3. Core Surveillance Data Element for Routine CDC Case Notification – Summative Risk Factor Indicator

Risk Factor Data Class Indicator	
Definition	Indicator that the person has any listed risk factor associated with the condition being reported.
Standard Code for Transmission to CDC	TBD
Value Set or Coding System for Transmission to CDC	Yes/No/Unknown
Abstracted/ Derived/ Created	Derived
Derived from (examples)	If a person has any listed risk factors associated with the condition being reported, the response would be “Yes”. If the person does not have any of the listed risk factors, this response is “No”. If this is not known, the response is “Unknown”.
Recommendations for Input Data Source(s)	Case Investigation eCR
Notes	This data element is a summative indicator that communicates if the person has any risk factor associated with the condition as listed in the Risk Factor value set.
Data Source Details	
eCR CDA	N/A
eCR FHIR	N/A
ELR	N/A
USCDI v4	N/A
USCDI+ Public Health: Case Reporting Use Case	N/A

Table 4. Core Surveillance Data Element for Routine CDC Case Notification – Summative Underlying Medical Condition Indicator

Underlying Health Condition Data Class Indicator	
Definition	Indicator that the person has any listed underlying health condition associated with the condition being reported.
Standard Code for Transmission to CDC	TBD
Value Set or Coding System for Transmission to CDC	Yes/No/Unknown
Abstracted/ Derived/ Created	Derived
Derived from (examples)	If a person has any listed underlying conditions associated with the condition being reported, the response would be “Yes”. If the person does not have any of the listed underlying conditions, this response is “No”. If this is not known, the response is “Unknown”.
Recommendations for Input Data Source(s)	Case Investigation eCR
Notes	This data element is a summative indicator that communicates if the person has any underlying health condition associated with the condition as listed in the Underlying Health Condition value set.
Data Source Details	
eCR CDA	N/A
eCR FHIR	N/A
ELR	N/A
USCDI v4	N/A
USCDI+ Public Health: Case Reporting Use Case	N/A

Data Class Structure in Message Mapping Guides

The preferred structure of the ERU data classes within a potential message mapping guide was discussed with a group of jurisdictional informaticists specifically convened for this purpose. Current message mapping guides have ERU data concepts in one of three formats: individual questions, repeating elements, or repeating groups (also referred to as repeating blocks).

Individual questions are built as a series of discrete data elements, each with a Yes/No/Unknown (Y/N/U) value “answer”. For example, “Did person attend a county fair?”, paired with a Y/N/U answer. An ERU data class built with individual questions will have several data elements, each addressing a specific exposure, risk factor, or underlying condition separately. Repeating elements use the data class label as the “question” and include a group of values for a select-all-that-apply “answer”. An example is a data element question of “What underlying health conditions does the person have?”, with a value set including “Diabetes mellitus”, “COPD”, and “bipolar disorder”. Repeating groups involve sending one instance of a repeating block for each value. Each instance will have a parent “question” to denote the value (such as “COPD”) and a corresponding child Y/N/U indicator to denote the presence or absence of that particular condition. This structure repeats as needed for all variables in the list. Examples of repeating groups can be seen in Figures 1 and 2 in the Background section of this document. Note that although the data elements for any of the three structures may be phrased as questions here, PHAs should have the flexibility to collect the information in a way that appropriately matches the data source (including a case interview, for example), and need not use the exact question from the MMG.

Participants from CDC strongly preferred the repeating group structure, as modifications to value sets do not require the lengthy review process that new questions do. In addition, value set changes can generally be incorporated into MMGs in a way that prevents jurisdictions from having to re-onboard the MMG in the event of such changes. Jurisdictional participants’ structure preferences were split between repeating group, repeating element, and individual questions. This split likely reflects differences in surveillance systems used across the US; however, it was agreed that consistency in structure across conditions for each data class is more important than prescribing the exact structure.

One potential benefit of the repeating group structure is that CDC could add more “child” data elements to a data class relatively easily. However, it was noted in the detailed structure conversation that a repeating group dependency structure of parent/child/.../child is more easily implemented than a parent/child/grandchild structure. To demonstrate with an example: a parent element in a repeating group is to select an exposure (e.g., consumed raw milk), and the child element indicates yes/no/unknown for whether or not the person consumed raw milk within the exposure window. Adding a “date of consumption” variable is simpler if it is not dependent on the answer of “yes” (making it a grandchild) and is instead an additional answer to “consumed raw milk” (making it a second child).

When polled, the majority of the workgroup (71%) agreed that it was very or moderately important, from an epidemiological perspective, to be able to send “No” and “Unknown” answers for all ERU data classes. This rules out the repeating element structure as an option for all three data classes. Combined with the benefits of the repeating group, this suggests that the repeating group is likely the most advantageous structure for all three ERU data classes.

Appendix A. Input Data Sources for Exposure, Risk Factor, and Underlying Health Conditions Data

Electronic Case Reporting (eCR)

Because of the broad nature of the type of data that can be included in the ERU data concepts, such data may be captured in several locations within an eCR transmission. The most common templates used to transmit this data in an electronic initial case report (eICR) are as follows.

- The **Problem Observation** template reflects a discrete observation about a patient's problem using SNOMED-CT or ICD-10-CM codes. This template is used across multiple sections within the eICR, including Past Medical History, Problem Concern Act, and Encounter Diagnosis.
- The **Exposure/Contact Information Observation** template can be used in either the Social History or Travel History sections. It represents a person's contact with another person, animal, or substance, or the person's presence at a location where exposure to an agent could have occurred; therefore, this template might capture exposures from the environment, at a specific event or location, with an animal, or with another person.
- The **Social History Observation** template is only used within the Social History section to represent a patient's occupation(s), lifestyle, and environmental health risk factors. Data exchanged using this template can include environmental exposures, risk factors and health-related behaviors such as exercise, tobacco use and exposure, and alcohol intake.

Electronic Laboratory Reporting (ELR)

Information on ERU data is typically not collected or transmitted in ELR. Although relevant content may be found in or derived from the Reason for Study field (OBR-31), caution must be taken if using this data to populate any ERU data elements. The specific SNOMED-CT or ICD-10-CM codes used in this field would likely require manual review to determine which of the three concepts apply.

United States Core Data for Interoperability (USCDI) and USCDI+

Table 5 below contains five data elements from the Exposure/Contact Information data class in USCDI that capture exposure-related information.

Table 5. USCDI Data Elements Relevant to Exposure-Related Information

Data Element	Definition	Status	USCDI+
Exposure/Contact Source/Target Participant	The source or target of the exposure or contact. Could be person, animal, or location.	Level 2	Yes
Exposure/Contact Direction	Whether the direction of exposure/contact is acquisition (person is the target and another person, animal, location, etc. is the source) or transmission (patient is the source and another person, animal, location, etc. is the target).	Level 2	Yes
Exposure/Contact Type	The type of exposure/contact (environmental, activity, event, location, person, animal, etc.) to an agent.	Level 2	Yes
Exposure/Contact Agent	The possible agent of concern in the exposure/contact.	Level 2	Yes
Exposure/Contact Date	The date on or period in which possible exposure/contact occurred.	Level 2	Yes

The USCDI guidance indicates that the applicable value sets and codes for these elements are pulled from the HL7 FHIR US Public Health Exposure Contact Information Profile. These data elements are all at a USCDI maturity level 2, indicating that though they have not been approved for inclusion in USCDI, they have broad standardization and guidance for their collection, existing terminology standards or technical specifications for their representation, are already being captured and exchanged in multiple EHRs, and can be applied to most care settings.

Unlike with Exposures, there is no single data class that offers data elements useful to the Risk Factors concept, but information on specific risk factors is spread across several data classes.

Table 6. USCDI Data Elements Relevant to Risk Factor-Related Information

Data Class	Data Element	Definition	Status	USCDI+
Problems	Problems	Condition, diagnosis, or reason for seeking medical attention.	Included since Version 1	No
Problems	SDOH Problems / Health Concerns	Social Determinants of Health (SDOH)-related health concerns, conditions, or diagnoses. Examples include but are not limited to homelessness and food insecurity.	Included since Version 2	No
Health Status Assessments	Alcohol Use	Evaluation of a patient's consumption of alcohol. Examples include but are not limited to history of alcohol use, alcohol use disorder identification test and alcohol intake assessment.	Included since Version 4	No
Health Status Assessments	Disability Status	Assessments of a patient's physical, cognitive, intellectual, or psychiatric disabilities. (e.g., vision, hearing, memory, activities of daily living)	Included since Version 3	Yes – Case Reporting
Health Status Assessments	Mental/ Cognitive Status	Assessment or screening for the presence of a mental or behavioral problem. Examples include but are not limited to alertness, orientation, comprehension, concentration, and immediate memory for simple commands.	Included since Version 3	No
Health Status Assessments	Substance Use	Evaluation of a patient's reported use of drugs or other substances for non-medical purposes or in excess of a valid prescription. Examples include but are not limited to substance use disorder score, and substance use knowledge assessment.	Included since Version 4	No
Health Status Assessments	Smoking Status	Assessment of a patient's smoking behaviors.	Included since Version 3	No

The data elements listed below may also contain information pertaining to risk factors. However, being categorized as either Level 2 or Level 0, they have not yet been approved for inclusion in USCDI. Data elements marked with an asterisk (*) are currently included in the Case Reporting use case for USCDI+.

- Social History data class
 - Sexual Activity (Level 2)
 - Congregate Living (Level 2)*
 - Refugee Status (Level 2)
- Social Determinants of Health data class
 - Incarceration History (Level 0)
 - Housing Instability and Homelessness (Level 0)*

Two data elements within the Problems data class, Problems and SDOH Problems / Health Concerns (Table 6, Rows 1-2), could be applied to the Underlying Medical Conditions concept. Any codes from either the SNOMED-CT or ICD-10-CM vocabulary would be applicable to these data elements. The other two data elements within the Problems data class, Date of Diagnosis and Date of Resolution, should be referenced to ensure that the time period during which the condition, problem, or diagnosis is present is consistent with that of being considered underlying; that is, the condition should have been present prior to and/or at the time of the case being investigated.

DRAFT