**Methods for Conveying Unknown or Missing Values in Case Notification Messages**

**November 8, 2016**

**Background**

During the Message Mapping Guides (MMGs) development process, the CDC Programs identified a specific need to differentiate between a “missing” value (no value is recorded for the variable in an observation) and a value of “unknown” (a value is applicable but unknown) for some key numeric, date, and coded data elements in case notification messages. To accommodate this request, the NNDSS and CDC Message and Vocabulary (M&V) teams initiated a review of possible technical solutions for those jurisdictions that can already capture the concept of “unknown.” Although an empty data field effectively communicates that information is “missing”, a designated code or value is still needed to allow differentiation of “unknown” at the data entry, data transmission and data provisioning levels.

During the requirements gathering phase, CDC program subject matter experts (SMEs) will identify any key data elements in their MMGs for which they request that the health jurisdictions differentiate “unknown” from “missing” values in case notification messages. The CDC programs will be responsible for vetting these choices with the appropriate data collection personnel at the jurisdictional level (in advance of the external comment period) as part of the requirements gathering phase and vocabulary development for the MMG. This vetting process will help to minimize the rework that might be needed if the feedback is received as part of the 6-week external comment period on the MMG.

In general, reporting jurisdictions should plan to send all the data elements from the MMGs that they currently collect within their surveillance systems, and they are encouraged to consider adding all of the requested elements to their systems as resources become available. If information is not available for a data element for a specific case, no value would be transmitted (i.e., the field would be empty) in the HL7 message or, if the data element is not required, the OBX segment does not need to be sent in the message. However, for data elements where the CDC Program has requested that “unknown” values be distinguished from those that are simply “missing”, the following guidelines should be used.

Please note that fields that are “required” in the message may not be left empty and may not be populated as an unknown value. If the jurisdiction captures only a “yes” reply and will never be able to differentiate between “no” or “unknown” values for a particular data element of interest to the CDC Program, jurisdictions will need to document that they do not have the capacity to make this distinction for the data element in question while either pilot testing the MMG, or during the onboarding process.

**Coded Data Elements**

For coded data elements, where the CDC Program requests that “unknown” be distinguished from “missing”, CDC requests that the submitting jurisdiction add “unknown” and other null flavors (see Appendix 1 for complete null flavor value set) to the coded value sets. CDC programs will have the discretion to request the use of any subset of these null flavors for the identified value sets. This will ensure the ability to differentiate between “unknown” and “missing” for coded values.

Example of a coded data element:

|  |  |
| --- | --- |
| **Data Type** | **Option for Conveying Unknown in Observations using the Literal UNK^UNKNOWN^NULLFL** |
| **Coded** | **For Coded Data Types:** For coded observations that are unknown, many value sets already contain an explicit “unknown”; Unknown and Other will be added to value sets, as appropriate.  **Unknown**: OBX|4|CWE|INV152^Case Disease Imported Code^PHINQUESTION||UNK^UNKNOWN^NULLFL||||||F  **Empty/Missing**: OBX|4|CWE|INV152^Case Disease Imported Code^PHINQUESTION||||||||F |

**Numeric and Date Fields**

If a jurisdiction’s surveillance system does not have a mechanism to capture “unknown” vs. “missing” information for numeric and date fields, the data elements sent from the jurisdiction to CDC will appear as “missing/empty” and the CDC programs will not be able to differentiate “unknown” from "missing". For situations where the program requests an unknown indicator and the jurisdiction’s surveillance system is able to distinguish between “unknown” and “missing”, CDC recommends the following:

To indicate that the value for the numeric (NM) or structured numeric (SN) data element is unknown, the sender should send a series of “9s” outside the normal range for the data element as instructed in the MMG implementation notes for that variable (see Appendix 2).

For date fields, the guidance for conveying unknowns is to also send a series of “9s” outside the normal range for the particular data element of interest outlined in the MMG. Please refer to Appendices 2-4 for examples of how indicator values will be displayed in various contexts.

**In developing these recommendations, CDC recognizes that jurisdictions could incur significant financial costs if changes to their surveillance systems were required in order to collect this information, as well as from additional time spent in data entry.  Therefore, these guidelines address only the methods for data transmission in HL7 messages when the information is collected in the surveillance system, and are not a requirement for the collection of the data.**

**The NNDSS Program recommends that CDC Programs be judicious in requesting that submitters distinguish between “unknown” and “missing” values in HL7 messages so as to minimize the burden on the jurisdictions.**

**Appendix 1. Null flavor value set for coded data elements**

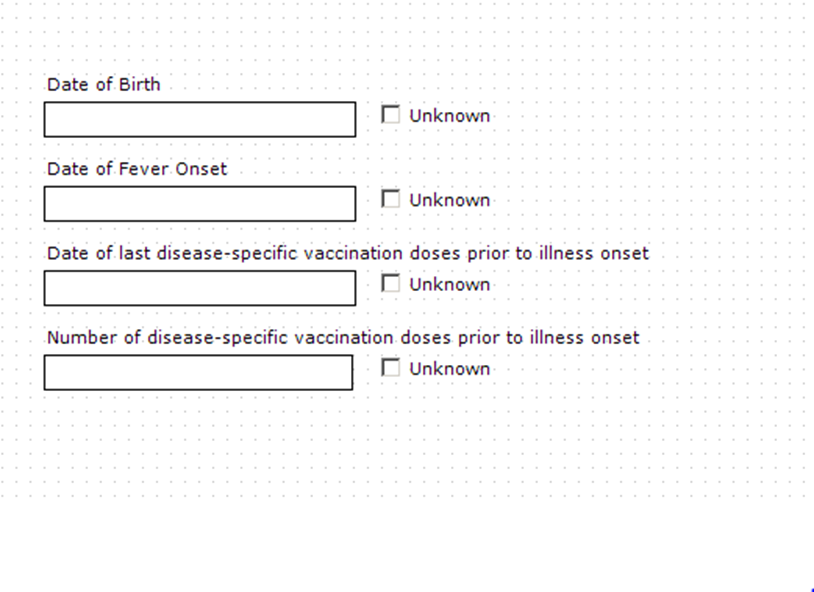
|  |  |  |
| --- | --- | --- |
| **Code** | **Name** | **Definition** |
| NI | No information | No information whatsoever can be inferred from this exceptional value. This is the most general exceptional value. It is also the default exceptional value. |
| OTH | other | The actual value is not an element in the value domain of a variable. (e.g., concept not provided by required code system). |
| UNK | unknown | A proper value is applicable but not known. |
| ASKU | asked but unknown | Information was sought but not found (e.g., patient was asked but didn’t know). |
| NAV | temporarily unavailable | Information is not available at this time but it is expected that it will be available later. |
| NASK | not asked | This information has not been sought (e.g., patient was not asked). |
| MSK | masked | There is information on this item available but it has not been provided by the sender due to security, privacy, or other reasons. There may be an alternate mechanism for gaining access to this information.  **Note:** Using this null flavor does provide information that may be a breach of confidentiality, even though no detail data are provided. Its primary purpose is for those circumstances where it is necessary to inform the receiver that the information does exist without providing any detail. |
| NA | not applicable | No proper value is applicable in this context (e.g., last menstrual period for a male). |
| NP | not present | Value is not present in a message. This is only defined in messages, never in application data! All values not present in the message must be replaced by the applicable default, or no-information (NI) as the default of all defaults. |

# Appendix 2: Examples of implementation notes for use with out-of-range values

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PHIN Variable ID** | **Code System Source** | **Label/Short Name** | **Description** | **Data Type** | **CDC Priority** | **May Repeat** | **Valid Values** | **HL7 Message Context** | **HL7 Data Type** | **HL7 Optionality** | **HL7 Implementation Notes** |
| DEM115 | PHINQuestions | Date of Birth (generic variable) | Date of Birth (generic variable) | Date | P | N |  | PID-7 Date of Birth | TS | O | For unknown date, PID-7 **MAY** be populated with '99999999' |
| VAC142 (proposed) | PHINQuestions | Date of last disease-specific vaccination doses prior to illness onset | Date of last disease-specific vaccination doses prior to illness onset | Date | P | N |  | Observation/OBX Segment with this variable ID and label | TS | RE | OBX-5 will contain '99999999' for unknown Date of last disease-specific vaccination doses prior to illness onset |
| VAC140 (proposed) | PHINQuestions | Number of disease-specific vaccination doses prior to illness onset | Number of disease-specific vaccination doses prior to illness onset | Numeric | O | N |  | Observation/OBX Segment with this variable ID and label | SN | O | \*OBX-5 will contain '^9999' for unknown Number of disease-specific vaccination doses prior to illness onset |

*\*Please note that Structured Numeric must have the ^ in the first component to be structurally valid.*

# Appendix 3: Example of how information on “unknown” status might be collected in a surveillance system:



# Appendix 4: Proposed CDCP Business Rules for processing out-of-range “unknown” indicators:

**If the field’s datatype is “TS” (timestamp), and the value is ‘99999999’, it does not fail. This applies to DOB (PID-7) and all the date observations in OBX-5.**

Example showing unknown DOB in PID segment:

PID|1||LocalPatID1DEM197^^^SendAppName&2.16.840.1.114222.GENv2&ISO

||~^^^^^^S||99999999|||||||||||||||

Examples showing unknown date in an observation/OBX segment:

OBX|53|ST|MTH153^Mother's Birthdate^PHINQUESTION||99999999||||||F

OBX|57|TS|75200-6^Date of first prenatal visit^LN||99999999||||||F

**If Value Type is NUM or SN, it may contain a series of 9s outside the normal range for the data element, and this will not fail.**

Examples showing unknown number in observations/OBX segments:

OBX|54|SN|75201-4^Number of previous pregnancies^LN||^9999||||||F

OBX|55|SN|75202-2^Number of live births (total)^LN||^9999||||||F