

Meeting: CSTE Data Standardization – Combined Exposures, Risks, and Underlying Conditions + Signs, Symptoms, & Clinical Manifestations -Data Element-Specific Team Meeting for Structure

Date: March 8, 2024

Attendees: 99

CSTE (2)	PHAs (42)	CDC (45)
Annie Fine	<u>DSWG Co-Chairs</u>	Allison Miller
Devra Barter	Laura Erhart AZ	Amanda Jones
JMC (7)	Molly Crockett MA	Andrea Steege
Erin Kough	<u>DEST Members</u>	Angie Agunloye
Grace Mallard	Aaron Bieringer MN	Antonio (Tony) Neri
Laura Lane	Andria Apostolou HIS	Carissa Rocheleau
Natalie Raketich	Anna Kocharian WI	Doris McGinness
Sam Spoto	Annie Bergman SD	Erin Miller
Sara Sanford	Apaegle UT	Erin Miller
Susan Downer	Beth Isaac NYC	Fidel Desir
Other/Unknown (3)	Cher Levenson WA	Gonza Namulanda
Amanda Markelz	Dimple Patel KY	Hailey Vest
Robbie Lojo	Ed Hartwick MI	Hang Nguyen
Yin	ej klingler NYC	John Gerstle
	Elizabeth Schiffman MD	Julie Vanderkolk
	Eman Addish Philadelphia	Katie Dodd
	Greg Fowler SC	Lauren Korhonen
	Hanna Chakoian MN	Laurie Barker
	Jesse Ellis MS	Lesliann Helmus
	Jordan McBride NH	Lizzi Torrone
	Karene Thomas-Watts CO	Luis Hernandez GDIT
	Kayleigh Nerhood GA	Marie Haring Sweeney
	Kyra Veatch IN	Mary Alao
	Lakin Mauch ND	Matt Stuckey
	Lea Moser NV	Meeyoung Park
	Li_Jiao HCPH	Melissa Pagaoa
		Monica Patton

	Luis Vela ID Lunden Espinosa IN Mario Beltran MD michael eberhart PA Misty Johnson WI Nwanne Onumah MS Rachel Klos WI Rob Arciuolo NYC Rob Laing OR Sam Groppell DC Sophia Cantor WA Steven "Ike" Eichner TX Sunanda Sarkar SC Teresa Schillo MN Theron Huntamer NV Tim Callahan GA Vajeera Dorabawila NY Yuritzy Gonzalez OR	Myrna del Mar González GDIT Nekole Locke Nicole Gallaway Nicole Lindsey Nicolette Bestul Noreen Kloc Peter Boersma Peter Daly Rachelle Oseni Renee Boehmert Sandy Roush Sara Johnston Sara Johnston Scott Grytdal Shawn Arshad GDIT Shelby Lyons Sonia Signh Veronica Burkel
--	--	---

**SOME INDIVIDUALS DIALED IN; THE NAMES ARE UNKNOWN, AND THEY ARE LEFT OFF OF THE ATTENDANCE*

HIGH-LEVEL MEETING NOTES (REFER TO MEETING SLIDES FOR MORE DETAIL AS NEEDED):

Data Class Structure in MMGs

DSWG Reminders

- Immunizations and Death recommendations are out with the DSWG for comment through 4/1.
- The DEST Sign-Up sheet has been updated to include future data classes.
- Sign up for DESTs here: <https://app.smartsheet.com/b/form/3da89d48bda94fc7a33c4a8a505fe54a>

Meeting Ground Rules

- We welcome everyone's opinions and active participation from jurisdictions, CDC programs, CSTE membership, etc.
- As we work through data elements/classes, we strive for 80/20 consensus.

In-Scope vs Out-of-Scope

- *Special notice for this call: Today's conversation is different in that it is not about whether a data element or class is "core," those conversations are happening separately, but today's focus is on the preferred structure to send groups of similar data elements in an HL7v2 case notification message.*
- Focus on core data for surveillance and transmission to the CDC within selected data classes.
- We will work to identify standard definitions and value sets within that definition of core data.
- This is not a discussion of how to collect this data, frame questions in surveys or interviews, or implement these data standards. This is also not meant to impose limitations on how jurisdictions collect data.

Background Review: MMG Structure for ERU and SSCM

ERU and SSCM Data Classes

- Last month, we met individually as groups to discuss data classes on exposures, risk factors, and underlying conditions as one team and signs, symptoms, and clinical manifestations as another team. We will still be working through the details of these data classes in April, but the groups have defined the four data classes as core.
- Each data class can be made up of many condition-specific data elements. For instance, the exposure data class may contain data elements asking about things the case may have been exposed to during the incubation period of the condition, such as "within an exposure period, did the person get a tattoo or consume a certain food." The risk factors data class may contain data elements that capture behaviors or circumstances that may make a person more likely to have the condition, such as if the person is a healthcare worker. The underlying conditions data class may capture information about the patient's relevant diagnoses before the condition is reported. Finally, the signs, symptoms, and clinical manifestations data class will capture information on the condition-specific signs, symptoms, and other clinical manifestations experienced by the case.
- As these data classes are all highly condition-specific, the individual data elements within them will not all be decided as a part of this standardization work. When implemented in the future, the data elements for a particular condition within each data class will be defined by the CDC, like how the CDC defines the values in current MMGs. These data classes being core means that the group agrees there should be a standard bucket to place this information in across all conditions. However, the bucket can be empty if the data class doesn't apply to a condition, and we will recommend a standard structure for these buckets.
- Today's conversation will be about this structure, not these definitions.

ERU and SSCM Data Classes *continued*

- Current message mapping guides contain data elements from these data classes using three separate structures.
 - They can be transmitted as individual Y/N/U questions, each representing one data element.
 - They could also be transmitted as a repeating element, allowing multiple data elements to be transmitted within one OBX segment. Think of this as one question where you can "select all that apply" for the response.

- Lastly, they can be transmitted as repeating groups, allowing many data elements in the same class to be transmitted as paired “question” and “answer” OBX segments for each data element.
- We’re going to dive into each of these more deeply next.

Individual Questions

- On this slide are some examples of exposure data elements transmitted as individual questions.
- Each of these data elements would be transmitted in its own OBX segment, with a unique identifier in OBX-3.1 and the Yes, No, or Unknown response in OBX-5.
- This response value set does not have to be limited to Yes/No/Unknown, but that is the most common method.
- This OBX segment cannot repeat, and there is no repeating of the OBX-5 response field.

Individual Questions *continued*

- This type of structure can be straightforward to map from most surveillance systems, although if a system is using a “select all that apply” structure on the front end then they may not have no or unknown values to map.
- There are still conversations happening about how to make determinations on how buckets like risk factors and underlying conditions are defined, so this structure also may make mapping of individual elements into a message more straightforward regardless of their data class bucket.
- On the CDC side, new data elements and revisions to existing data elements within MMGs must be cleared through an OMB process, which can take time, and a new or updated MMG version may then need to be onboarded, which can be a burdensome process for jurisdictions.

Repeating Element (aka “may repeat”)

- A repeating element will allow for multiple data elements to be sent in one OBX segment by repetition of the OBX-5 values. For instance, transmitted in the example here is the type of infection or clinical syndrome caused by the organism, with an answer value set that includes a list of infection types or syndromes.
- This example segment is transmitting values for both pneumonia and cellulitis.

Repeating Element (aka “may repeat”) *continued*

- Since the list of data elements, or in this example the list of types of infections, is contained within a value set, this value set is able to be updated without going through an OMB approval process, which can make updates timelier.
- Additionally, a change to a value set may not require jurisdictions to go through the MMG onboarding process.
- This structure also aligns with some jurisdictional systems that have elements structured this way on the front end.
- However, this structure only allows for transmission of a positive assertion of a value and lacks the ability to transmit No or Unknown responses.

Repeating Group

- The final structure we’re going to go over is the repeating group. This is where the data elements are transmitted using paired question and answer OBX segments, and the paired segments repeat as necessary to transmit responses for all values within the question value set.

- In this example, it may be small to see, but the “Question” OBX segment is a listing of underlying conditions or prior illnesses, and then “answer” OBX segment is an indicator using Y/N/U response values.

Repeating Group (continued)

- Each pair of OBX segments is linked using the OBX-4 value.
- In the example here, you can see the OBX-3 indicator value is INV1117 for patient epidemiological risk factors in the first and third “question” OBX segments.
- The first OBX segment has an OBX4 value of “one”, which links it to the matching response OBX segment below it, which also has an OBX4 value of “one”.
- Looking at the OBX5 for these two segments, we can see that the first segment is conveying the question of “diabetic at diagnostic evaluation”, and the answer segment conveys the response of No. This case is not diabetic. Looking at the bottom two paired OBX segments, linked by the matching OBX4 value of “two”, the first one is conveying the question of “end stage renal disease”, and the answer segment conveys the response of yes. This case has end stage renal disease.

Repeating Group (continued)

- Since the list of data elements, or in the example from the last slide the list of patient epidemiological risk factors, is contained within a value set, this value set can be updated without going through an OMB approval process, which can make updates timelier. Additionally, a change to a value set may not require jurisdictions to go through the onboarding process for a new MMG.
- This structure also more easily allows for addition of another “answer” into the repeating block associated with all values. For instance, if it was important to capture the date that an exposure occurred, the information could be conveyed as a repeating triplet of OBX segments – one identifying the specific exposure, like consumption of eggs, one with the Y/N/U indicator, and one with the date that the case ate eggs. This triplet could be repeated for all exposure values.
- As with the individual questions, if a system is using a “select all that apply” structure on the front end then they may not have no or unknown values to map. There may be other jurisdictional surveillance system set ups that also make this structure tricky to map and require some informatics gymnastics.

Data Class Indicator

- A data class indicator is an individual data element that exists separate from the structures we just talked about. An example of this in the picture here shows a data class indicator of “Did the subject have any underlying causes or prior illnesses” with a responses limited to Y/N/U. This exists as its own OBX segment separate from the below repeating group for underlying conditions. An answer of “no” to this indicator value may remove the need to transmit the repeating group, or in the case of a repeating element question it can add value to explaining why the repeating element is absent. The implementation of this question varies across jurisdictions, and at the jurisdictional level may be independent of or in some cases calculated secondary to values in a repeating group or element.
- A “Symptoms present” indicator is present in the Minimum Dataset for Case Notification During an Emergency, developed by a separate CSTE working group. Whether or not a data class indicator should be transmitted as part of each of these data classes is likely data class dependent, as it may be more valuable for signs and symptoms compared to something like exposures. This group will not be discussing if a data class indicator is useful for a certain data class. However, we want to inform those data class specific conversations with information from this group

on the implementation of a question like this. After our structure conversations, we're going to talk about if and how a data class indicator affects the workload or ability to map data into the structure or structures we decide on.

Structure Discussion

The focus for today's discussion is on the message building that happens on the back end of the surveillance system, as much as it can be removed from that front end component.

Discussion

- The overarching question here is if we are considering exposure, risk factors, underlying conditions, and signs, symptoms, and clinical manifestations as separate data classes – what structure is best for all or each of these data classes? A discussion of the data indicators will be a secondary discussion.
- Would it be Individual questions, repeating elements, or repeating groups? This slide gives the barebones pros and cons of each structure type.
- Before heading into any further discussion, however, I want to open the floor for some open conversation about these structures – hopefully we have some people here who can speak to the advantages or difficulties that have come about at the jurisdictional level due to these types of structures.

DEST Discussion

- Ed Hartwick (MI): Noted that for Michigan they could implement any of these options and there are advantages and disadvantages to both options. The most advantageous options would be those with a focus that don't require additional onboarding with there is a change. Also, if OMB approval is required, we know those processes can take a long time and may be resource intensive.
- Misty Johnson (WI): Had a clarifying question regarding the discussion to confirm that this is focused more on the back end than on the user perspective.
 - Erin Kough (JMC): Commented that she understands these can be tied together, however, what we are trying to establish is whether there is a barrier in the back end on building an HL7 message from what jurisdictional systems can create.
- Misty Johnson (WI): As an informaticists, repeating elements and more particularly, repeating groups, takes a bit more coding especially if there is a parent element that is rarely in place, then the repeating group won't display. It does take a bit of skill to set up a repeating group on the back end. However, when you're onboarding messages all the time, you learn how to do it, so it becomes less of an issue. On the front end, the repeating element itself is a bit easier. She wanted to provide that clarity on the difference between front end and back-end implementation and that traditionally, for the back end, individual questions are easiest for informaticists, but this is shifting, and people are learning how to deal with the data from repeating elements/groups.
 - Erin Kough (JMC): Had a follow-up question, when you've completed one repeating group or element does it make it easier to do others?
 - Misty Johnson (WI): Yes, you learn tricks. She also likes the term "data class indicator" and noted that she liked the idea of adding that to a repeating group and you can capture if something was discussed with a patient.

- Teresa Schillo (MN): Agreed that the repeating elements and groups are more challenging than individual questions and that they always must go back to look at previous examples to understand how it was set up. It's challenging and can be time consuming, even if there is an example to use, but it can be done.
- Annie Fine (CSTE): wanted to pose a couple of other questions for consideration. Particularly for risk factors, it was felt that there will be associated questions that CDC will want to know. For example, what was the timing and did X occur within the incubation period. Those types of elements would be beneficial if they are contained within the repeating group. Another consideration is for a repeating group, if you are sending updates, how easy would it be to update that information within a repeating group? Would all the information need to be resent? How would that get incorporated on the receiving end if there is already existing data? This could add additional complexity to an already challenging concept. Is this something that CDC can process?
 - Ed Hartwich (MI): From his perspective it does not seem like it would be more work.
 - Misty Johnson (WI): Commented that from a jurisdictional perspective, within message mapping you are given a vocabulary set that is used. If there are 20 different exposures it would be a maximum of twenty iterations. It doesn't matter if it's right the first time around, if a person answered 10 questions and the second time there was an error, and they could delete that out and send only one set. So, if an error is made within a repeating group, you can delete what is not correct and send only what is correct because it is a finite set.
 - Sara Johnston (CDC): From the CDC side, we have snapshot processes within the HL7 messages. That means you send everything that is current as of the time you send it and then if you update the record, you send everything. So, when CDC receives the information in MVPS, the most current valid record is the one that is used, regardless of what was sent in previous records. It does not try to reconcile data that was sent in previous records, it is only what is contained in the most current valid record.
 - Annie Fine (CSTE): Was curious if that meant that it deletes all the previous data that is not present in the current data?
 - Sara Johnston (CDC): The information is still available in the database but if you are looking for the current data within a case, it would not be in that record because it is only looking at the most recent record.
- Greg Fowler (SC): Noted that he concurs with many of Ed's comments. Once you've created something like this, repeating it is not a big issue to resend the data. It feels straight forward from his perspective.
- Teresa Schillo (MN): Added that as a comment to one of the questions about the ease of adding a repeating element or group that the way their system is set up in MN, setting up individual questions within a repeating element is going to take a bit of work. It requires adding new fields and changing database structure. It's not a simple implementation regardless of what structure is chosen.
- Sara Johnston (CDC): Wanted to point out that as we consider ways to "future proof" these processes, is that we want to aim for more flexibility for emergency response or other situations is how to best make modifications without having it be a major issue within jurisdiction's systems to complete that type of modification. It is more complicated to set up the repeating elements/groups, but it gives the most flexibility in terms of the ability to make updates to individual risk factors within a value set. The ability to expand beyond a single concept such as was the risk factor present or not and have other concepts associated with it like timing or other indicators within a repeating group, allowing for greater flexibility as it doesn't limit relationships in the same way as a repeating element or individual question.

- Erin Kough (JMC): Noted this as a good point and recalled that during her time in MN during COVID, it would have been nice to have some of these elements in place for sending data.
- Hannah Chakoian (WI): Added to the chat that some jurisdictions might not have the people or resources to figure out the complexities related to the repeating blocks, and this might be a barrier to MMG onboarding. Urged the group to keep this in mind for the discussions.
 - Ed Hartwick (MI): Agrees with this as well and feels fortunate to have the resources to manage this. Also wanted to comment that complexities wise from a set up and go, almost think that a repeating element would be a preferred element. It would be great to have checkboxes with symptoms, as an example. Having this inside a single element might be easier to expand or contract, and it would take more work within a repeating group, but the repeating group is the preference.
- Teresa Schillo (MN): Wanted to make a correction to a comment she gave earlier. As people have been talking about the repeating element, she noted it would be easier to change if there was a new value that needed to be sent as opposed to the individual questions and repeating groups that would require a lot more overhead with the system they have in place.
- Annie Fine (CSTE): One item of note from her recollection of working at a jurisdiction, is that sending data for things like risk factors or symptoms, a no or unknown can be hard to distinguish unless you are performing an epi study or a case control study. “Yes” data was always more meaningful than “No” responses because you don’t know if it is truly a no response or if the case just didn’t answer that question.
- Lesliann Helmus (CDC): Agreed with Annie’s point of being able to send true a true “no” or keeping the data as unknown. She also wanted to return to the question of what if additional information needed to be added to a repeating element, such as timeframe or other data. How complicated would that be if it needed to be added to a repeating element? Or is it easier to focus on repeating groups and we have that flexibility when needed?
 - Ed Hartwick (MI): Replied with a comment that usually it depends on how well the information is going to be received. If you are adding an element to a repeating group like a date value that is new, how is that handled within the message mapping guide. How new is it? Is it part of a repeating group or is it a new variable that describes something related to it? If it is the latter, we run into problems because it becomes an MMG generating problem. If it can be sent without a new MMB onboarding and CDC can receive the uptick in data, then it would not make a lot of difference on his side.
 - Misty Johnson (WI): In response to Lesliann’s question, she noted that if you did have repeating element and then, we wanted to add the date of that risk factor, you'd have to remove that element from your form, and then replace it with a repeating group. So for us, it would mean trouble at the back end with our database, because if we have all these records where we've captured, risk, for example, as a repeating element, it's a totally different structure. So, if you ever think in the future that you're going to want, more than one piece of information for each little snippet, then go with a repeating group, even if it only has two elements instead of a repeating element.

Discussion Points

The next question related to this is does it make sense to use the same structure across the data classes? For example, would it work to have the same structure for Signs/Symptoms and Clinical Manifestations and Exposures? And we don't need to think about whether these are core, we just need to think about structure. Would you rather keep things as simple as possible for one group or does it make sense to keep them the same?

DEST Discussion

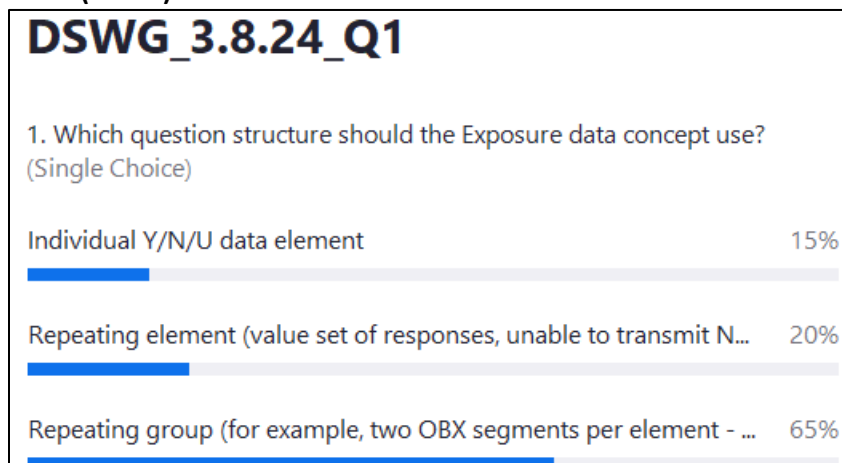
- Misty Johnson (WI): Even though Ed and I had said that a repeating group is very flexible, however, it would be challenging to see an MMG where everything is a repeating group. But this is where I would like to hear from Epis. From previous experience in putting MMG's together, sometimes they want you to just check a box and select apply that apply for say cough, fevers, chills, etc. Others have wanted to know information for each symptom, such as when did you experience fever or how severe it was. Likely this is just going to be a progression and we'll need knowledge for how to put this together, but she requested more input from Epis for what kinds of information they need.
 - Erin Kough (JMC): Confirmed this is something that we are asking this group to consider as we look forward to standardization.
 - Theresa Schillo (MN): Agreed with Misty's comments. She initially wondered why we are trying to force the same structure again and that we should instead be doing what is best for each one of these classes.
 - Ed Hartwick (MI): Agreed in the chat as well.
- Lesliann Helmus (CDC): Questioned whether the comments on flexibility were focused on if signs and symptoms should be done one way versus risk and exposures or if this is from guide to guide? What degree of flexibility is being asked for here. Is the request to leave it open and do whatever jurisdictions want? She noted there are some things that will need to be standardized.
 - Teresa Schillo (MN): Noted that she is not coming from an Epi perspective and was considering this as a question for each data class. If this is something that is core, does it mean it's the same structure? For example, would exposures be the same structure for each disease?
 - Erin Kough (JMC): Commented that moving forward the goal is to have say exposures structured in one way and that it would be the same for different conditions. Within each data class it would be the same. The question we're posing today is, should those structures be the same across each of the data classes or do they need to have their own structure for exposures, risk factors, signs, etc.
 - Ed Hartwick (MI): Wanted to better understand if this is by condition or by variable. He would propose that by variable you would want consistency across variables as much as possible. He felt we should not be imposing the structure for signs and symptoms for onset date, as an example. Signs and symptoms might have limited variability across conditions and act in the same way. That would help from an information systems standpoint on both sides. For the jurisdiction it helps us understand what is expected and for CDC it would be helpful to know what we will be sending.
 - Erin Kough (JMC): Took a moment to call back to a previous slide for repeating group examples, that the idea would be that for signs and symptoms there would be a repeating group that would be used across flu or foodborne or any condition.
 - Ed Hartwick (MI): Noted that he liked the idea of the recommendation because that does allow for flexibility.

- Sara Johnston (CDC): Given what was just said, she was curious how would we identify which structure makes the most sense for each data class if this is across conditions? Part of the challenge is that we've got so much overlap in how some of these categories could be conceptualized. It makes it challenging to consider where we would or would not want to use the same structure when there is so much similarity in how we treat those things. Also, we need to consider the analysis of the data. This discussion is not about the analysis but it's still something we would want to consider if we moved away from checkboxes to more informational data collection. It's hard to identify when it makes sense to have one thing versus the other and where the break points would be.
 - Erin Kough (JMC): Confirmed that in the specific discussions held for exposures, risk factors, and underlying conditions, there were a lot of discussions around this issue within the group. It is a challenging concept.

Polling

Today's polling is slightly different as it will not lead to recommendations however it will help to guide continued discussions.

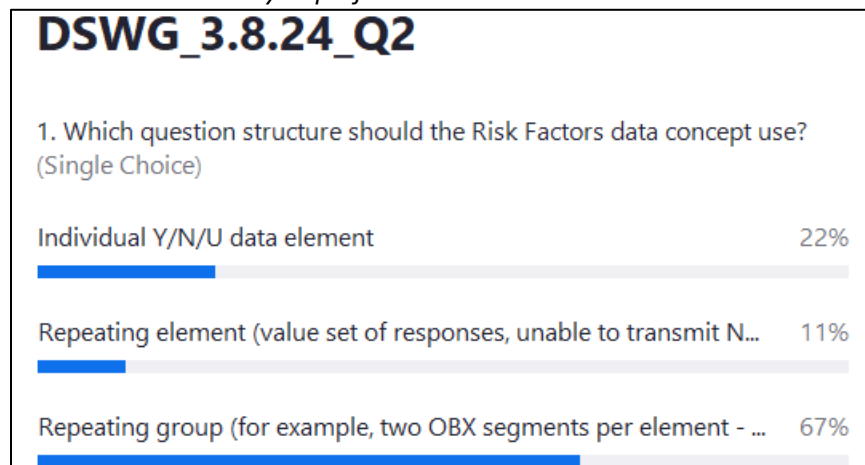
Poll 1 (n = 40)



Poll 2 (n =36)

Note that in the meeting the polling question was updated in the chat to reflect the following, "If a repeating element is not feasible, which of the other

two structures would you prefer the most?"



Next Steps

- There will be a second structure call on March 15.
- Exposures, Risks, Underlying Conditions (Call #2) scheduled for April 12
- Signs, Symptoms, Clinical Manifestations (Call #2) scheduled for 4/19.
- The next large DSWG call is scheduled for 4/26/24.