Background

We propose to include microbiology results in the OMOP-CDM model. Lack of microbiology data limits any meaningful studies of patient outcomes for infectious disease. Antibiotic resistance is a global crisis and US federal and international health agencies have been tasked with combating antibiotic resistance worldwide. While laboratory-based repositories for microbiology surveillance exist, these are rarely linked to patient level data. Benefits of incorporating microbiology data into OMOP will include the ability to:

I. Create accurate phenotypes for infectious disease syndromes for cohort selection (e.g. all patients with Staphylococcus *aureus* bacteremia)
II. Measure variations in antibiotic use for specific disease entities and organisms internationally
III. Enable comparative effectiveness studies of antibiotics for specific organisms and infectious disease syndromes

Uniqueness of microbiology results

As opposed to most measurements, microbiology tests are unique. Each specimen that is sent for a microbiology test - can be positive for MULTIPLE organisms. Each organism that is resulted is also related to a triplet of concepts - the organism, a lab test (e.g. antibiotic susceptibility), and its result (sensitive vs. resistant). The convention to represent this triplet has been discussed on the forums where the measurement, specimen, and observation tables are linked via the fact_relationship.

Let us illustrate this with a typical patient
On Feb 27, 2019 patient John Snow has a surgery - inguinal hernia repair.
On March 1, 2019 he has a fever, a small amount of pus is noted on physical exam at the surgery site. A diagnosis of cellulitis (wound infection is made). The patient is started on antibiotics and a wound culture (specimen) is sent from surgery site..
On March 4 2019, the wound culture specimen is positive for TWO organisms - Escherichia *coli* and Klebsiella *pneumoniae.*
Both of these organisms are tested against different sets of antibiotics and susceptibilities are available on March 5. One of the antibiotics tested (Cefazolin) shows different results:- i.e the E.*col*i is sensitive but the K.*pneumoniae* is resistant . They are both however sensitive to an antibiotic called Levofloxacin.
Patient John Snow is discharged home on March 6 on Levofloxacin.
Here a single specimen (the wound culture) is positive for two organisms and multiple related concepts e.g.

- E.coli - Cefazolin- Sensitive
- K.pneumoniae- Cefazolin- Resistant
- E.coli- Levofloxacin- Sensitive
- E.coli - Levofloxacin- Sensitive

What are some questions we need to answer effectively with microbiology data in the CDM

Clinical question 1 (a typical information required for clinicians before prescribing antibiotics)
What is the prevalence of a specific organism in blood-stream infections and what antibiotic is it sensitive to?

e.g. 20 % of all blood cultures at a specific center were positive for *Pseudomonas aeruginosa* and 70% of these were sensitive to the antibiotic Cefepime. This helps providers choose antibiotics for specific infections that are likely to succeed. We can also compare this across centers to minimize antibiotic overuse and design guidelines for antibiotic therapy.

Clinical question 2 (a retrospective cohort study requested by a funding agency)
In 2025, the World Health Organization wants to do an international study looking at microbiology of postoperative surgical site infection infections post inguinal hernia repair and whether these can be prevented using an antibiotic (Cefazolin) given just before the surgery. They want to focus on the years 2015-2025. We want to use the OHDSI data to answer this question.

1) The first step is to create a case definition for "infection after inguinal hernia repair". We use a procedure concept for inguinal hernia repair - and collect wound cultures in the 30 day period after the surgery.
2) The WHO wants to know if the use of Cefazolin could have prevented this infection. So we need to know if the two organisms in the wound culture were both susceptible to Cefazolin. Cefazolin sensitivity is tested separately for both organisms (which remember are obtained from the SAME specimen). **So antibiotic susceptibility has to be linked to an ORGANISM and not to the specimen**

See links below for example studies that we hope to replicate in OHDSI
https://www.ncbi.nlm.nih.gov/pubmed/16836755
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3346661/

Proposed Modification to the OMOP CDM

The proposed changes include:
1) Adding a new table called MEASUREMENT_ATTRIBUTE
2) Adding a new field called MEASUREMENT_ATTRIBUTE_ID to the measurement table. The new field will be a foreign key to the new MEASUREMENT_ATTRIBUTE table.

MEASUREMENT_ATTRIBUTE Table

Field *Proposed_revision*	Required	Type	Description
measurement_attribute_id	Yes	Integer	Unique identifier for each completed sub-measurement (will be connected to the measurement table)

measurement_id	Yes	Integer	A foreign key that refers to a Concept identifier in the Standardized Vocabularies
measurement_attribute_concept_id	Yes	Integer	This would be a standard concept id for an Organism (or Panel test)
specimen_id	No	Integer	A foreign key to the specimen table
measurement_attribute_source_value	No	Varchar	The organism value as it appears in the source.

Using the example from above, the below tables represent a patient that has E.Coli and is sensitive to Levofloxacin. New fields are highlighted in orange.

Measurement Table

measurement_id	123
person_id	1
measurement_concept_id	3009401
measurement_date	03/04/2019
measurement_datetime	03/04/2019 2:30
measurement_type_concept_id	
operator_concept_id	
value_as_number	
value_as_concept_id	36210475
unit_concept_id	
range_low	
range_high	
provider_id	

visit_occurrence_id	
visit_detail_id	
measurement_source_value	20629-2
measurement_source_concept_id	3009401
unit_source_value	
value_source_value	LA24225-7
measurement_attribute_id	1

Specimen Table

specimen_id	12345
person_id	1
specimen_concept_id	4001225
specimen_type_concept_id	581378
specimen_date	03/04/2019
specimen_datetime	03/04/2019 12:30
quantity	
unit_concept_id	
anatomic_site_concept_id	4274743
disease_status_concept_id	
specimen_source_id	
specimen_source_value	
unit_source_value	
anatomic_site_source_value	
disease_status_source_value	

Measurement_Attribute Table

measurement_attribute_id	6789

measurement_id	123
measurement_attribute_concept_id	4011683
specimen_id	12345
measurement_attribute_source_value	E.Coli

How will this schema help answer the two clinical questions posed in the first section

1. From the specimen table extract all specimen_ids for a specific source (blood). These specimen_ids are linked to the organism (Pseudomonas) via measurement_attribute_concept_id in the measure_attribute table. We link the required antibiotic susceptibility measurements (antibiotic Cefepime) in the measurement_table using the measurement_attribute_id. Then create a frequency table of susceptible vs. resistant.
2. First identify patients undergoing the specific surgery from the procedure table. From the specimen table extract all specimen_ids for a specific source (wound) for these patients using a time window of 30-days post date of surgery. For this group of specimens, then identify all positive organisms from the measure_attribute table . Last, extract antibiotic susceptibilities (cefazolin) from the measurement table using the measurement_attribute_id.

Generalization of Use Case

Though the proposed changes to the CDM was motivated by microbiology results, it is flexible to handle any measurement (e.g. see **Table 1**) **where a single laboratory specimen will have a multiple measurements and related attributes (>1)** (e.g. in this case a single wound culture has two specific organisms, each of which will have further antibiotic susceptibility testing). A related example would be laboratory tests which use a nucleic acid based test (multiplex PCR) to identify multiple viruses at the same time. These are done on a single specimen (nasal swab) and can be positive for multiple viruses (e.g. influenza or RSV). Each of these viruses can have other related measurements (e.g. copy number).

Additionally, this table will allow for the CDM to track the individual lab results from a panel test. For example, in a basic metabolic panel, there are seven lab tests performed (BUN, creatinine, carbon dioxide, glucose, serum chloride, serum potassium, serum sodium). Each of these lab test is captured as one row in the measurement table; however, there is no clear way to link these 7 results together without the fact-relationship table. With the proposed measurement_attribute table, the 7 rows in the measurement table can reference a single row in the measurement_attribute table where the measurement_attribute_concept_id=3022035.

Another example where this one-to-many relationship exists is with ECG results. The proposed ECG mappings from this forum post could populate individual rows in the measurement table, and each row can be linked to the originating ECG study via the measurement_attribute, where the measurement_attribute_concept_id would contain the concept ID for ECG Study (301512) .

Table 1.

Test	Specimen	Level 1- Measurement	Level 2- Measurement
Wound culture	Wound culture	Specific organism: Organism 1 Organism 2	Antibiotic susceptibilities for each organism
Viral Multiplex PCR panels (used to test for multiple viruses from a single specimen)	Nasal Swab Blood PCR test	Specific virus Influenza HIV	Copy number for each virus Resistance to antiviral
Autoimmune encephalitis screening panel http://ltd.aruplab.com/tests/pub/2013601	Spinal fluid antibody panel	Specific antibody (e.g.NMDA, GAD)	Level of each antibody