N3C Short Course
Session 3

Jerrod Anzalone
Department of Neurological Sciences

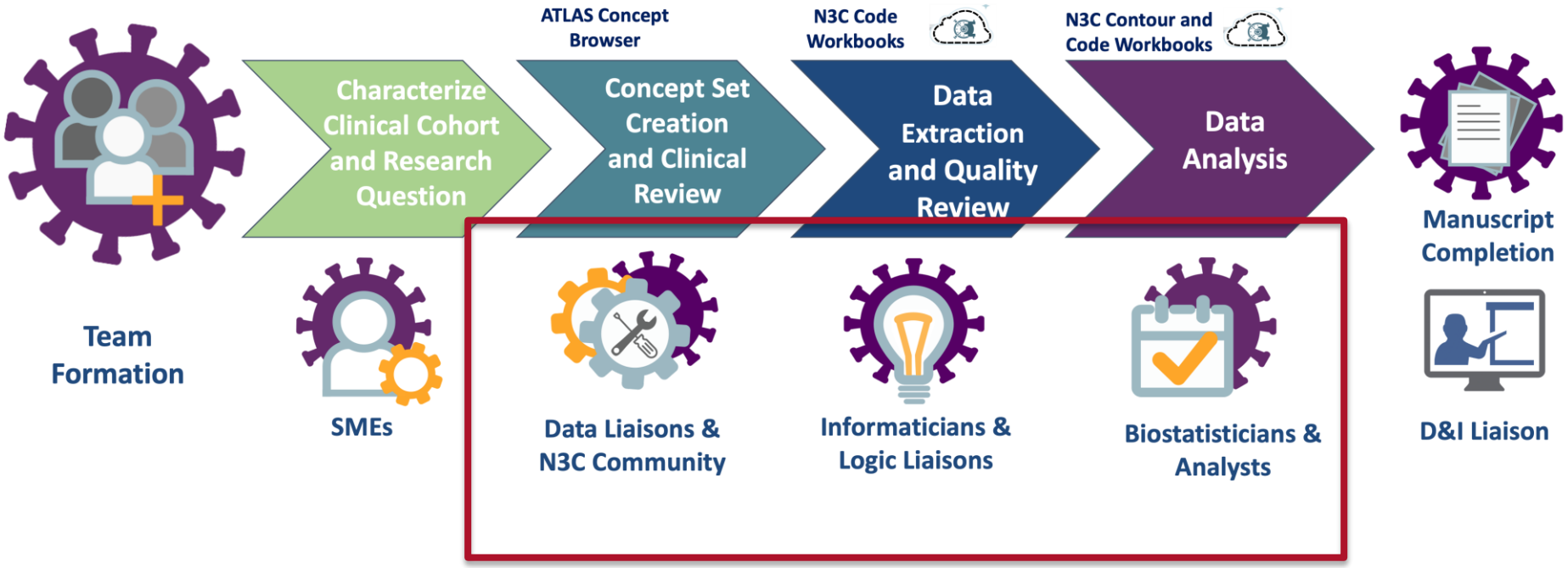
UNMC COLLEGE OF MEDICINE

Overview

- N3C roles and responsibilities
- Computable Phenotype
  - Analytic Plan
  - Concept Set Curation
  - Domain Requirements
- Logic Liaison Template
  - Data Extraction

N3C Role and Responsibilities

Domain Team Organization



Team Composition

SMEs

Jana Ponce, PhD, RD
Corrine Hansen, PhD, RD
Kristina Bailey, MD
Megan Timmerman, MPA, RD
Mariah Jackson, MMN, RD

Informaticians & Logic Liaisons

Jerrod Anzalone, MS

Data Liaisons & N3C Community

James McClay, MD, MS
Greater N3C Community (Davera Gabriel, etc.)

Biostatisticians & Analysts

Harlan Sayles, MS

Analytic Plan

Primary Exposure: Malnutrition

COVID and No Hx of Malnutrition

COVID and Hx of Malnutrition

COVID and Hospital-Acquired Malnutrition

Covariates

- Age
- Gender
- Race/Ethnicity
- Charlson Comorbidity Index
- Smoking Status
- Region

Outcomes

- Death or Transfer to Hospice (primary)
- Invasive Mechanical Ventilation
- Oxygen Support
- Acute Respiratory Distress Syndrome (ARDS)
- ECMO
- Hospital-acquired pressure injuries

Inclusion Criteria

- Age >18
- Positive COVID diagnosis or lab test (PCR, Ag)

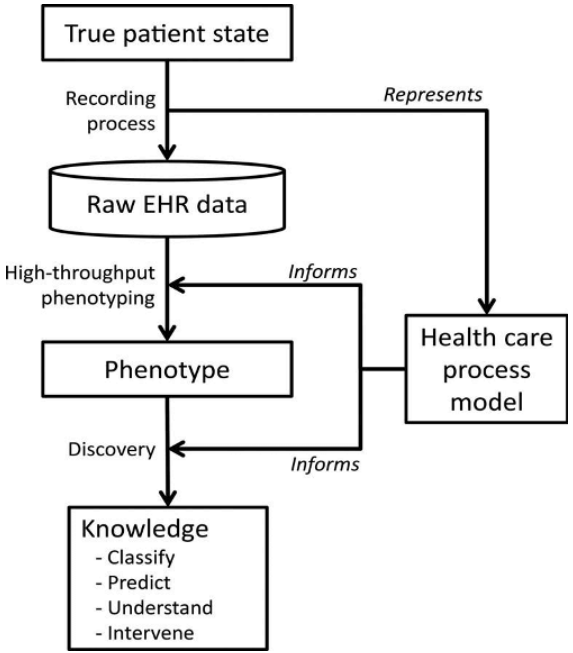
Exclusion Criteria

- Missing age, gender
- Data partners with low death reporting

Computable Phenotyping

Computable Phenotype

"A **computable phenotype** refers to a set of clinical conditions and characteristics that can be evaluated via a computerized query to an EHR system or clinical data repository using a defined set of data elements and logical expressions"

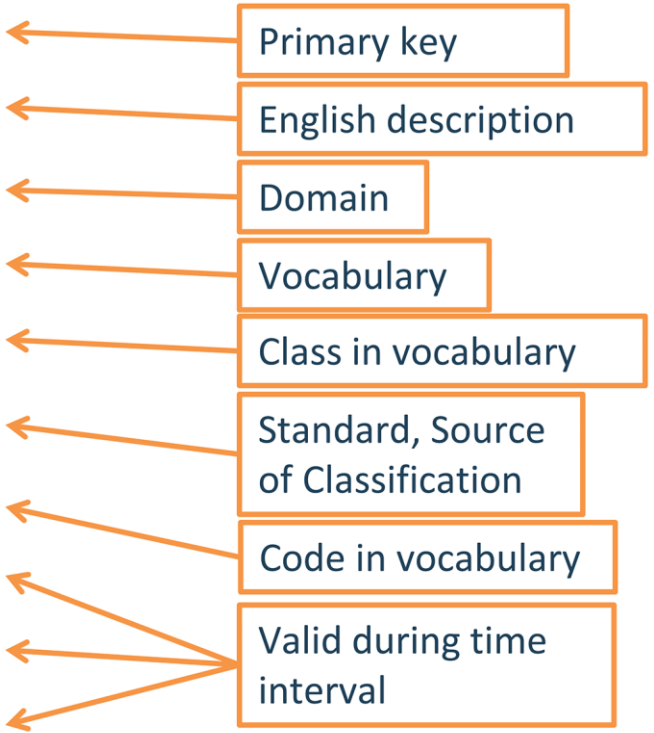


Richesson RL, Hammond WE, Nahm M, et al. 2013. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc. 20:e226-e231.

Hripcsak G, Albers DJ. 2013. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 20:117-121. doi:10.1136/amiajnl-2012-001145.

Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	

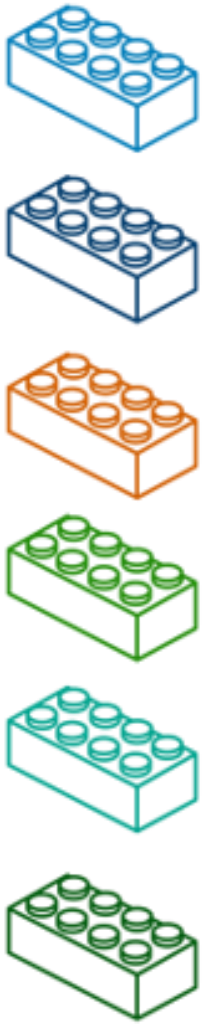


Source: *Book of OHDSI*

Concept Set

Concept Set: A data-agnostic expression that defines one or more Standard Concepts encompassing the clinical entity of interest.

Concept sets are the building blocks of a cohort definition.



Conditions

Drugs

Procedures

Measurements

Observations

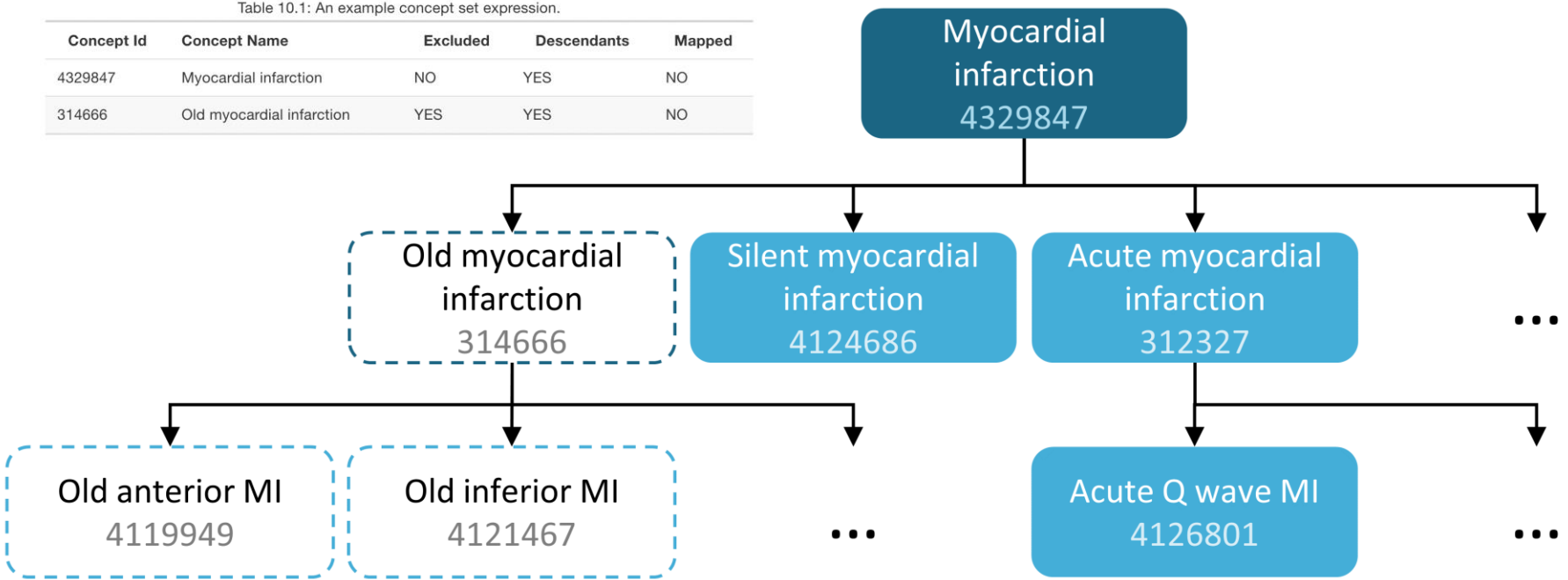
Visits

Source: *Book of OHDSI*

Concept Set Example

Table 10.1: An example concept set expression.

Concept Id	Concept Name	Excluded	Descendants	Mapped
4329847	Myocardial infarction	NO	YES	NO
314666	Old myocardial infarction	YES	YES	NO



Source: *Book of OHDSI*

Concept Sets Required

Cohorts:
- COVID-19 Positivity
- Malnutrition

Outcome Concept Sets:
- Hospitalization
- Invasive Mechanical Ventilation
- ECMO
- Death
- Pressor exposure
- Hospital-acquired events

Comorbid Conditions:
- Smoking
- Hypertension
- CCI Categories, including:
  - Heart failure, MI, CVD, dementia, diabetes, pulmonary disease, rheumatic disease, peptic ulcer disease, liver disease, renal disease, cancer, HIV

Medication exposure
- Remdesivir
- Steroids

OMOP Domains

Domain	Standard Vocabulary
Condition	**SNOMED**, ICDO3
Procedure	**SNOMED, CPT4, HCPCS**, ICD10PCS, ICD9Proc, OPCS4
Measurement	**SNOMED, LOINC**
Drug	**RxNorm**, RxNorm Extension, CVX
Device	**SNOMED**
Observation	**SNOMED**
Visit	CMS Place of Service, ABMT, NUCC

Source: *Book of OHDSI*

Conditions

- A diagnosis, a sign, or a symptom, which is either observed by a provider or reported by the patient.
- Concept set requirements for study:
  - COVID-19 Diagnosis
  - Malnutrition
  - Smoking status
  - Hypertension
  - Charlson Comorbidity Index Categories
  - Hospital-Acquired Pressure Injury
  - Supplemental Oxygen
  - Acute Respiratory Distress Syndrome
  - Invasive Mechanical Ventilation

Drugs

- A Drug is a biochemical substance formulated in such a way that when administered to a Person it will exert a certain physiological effect. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Radiological devices ingested or applied locally do not count as Drugs.
- Concept set requirements for study:
  - Pressor support
  - Remdesivir
  - Steroids
  - Supplemental Oxygen

Procedures

- Activity or process ordered by, or carried out by, a healthcare provider on the patient to have a diagnostic or therapeutic purpose.
- Concept set requirements for study:
  - Hospital-Acquired Pressure Injury
  - Supplemental Oxygen
  - Invasive Mechanical Ventilation
  - ECMO

Measurements

- A structured value (numerical or categorical) obtained through systematic and standardized examination or testing of a person or person’s sample.
- Concept set requirements for study:
  - COVID PCR or Antigen Test
  - Height
  - Weight

Observations

- A clinical fact about a Person obtained in the context of examination, questioning or a procedure.
- Concept set requirements for study:
  - Smoking status
  - Hospital-Acquired Pressure Injury
  - Supplemental Oxygen
  - Invasive Mechanical Ventilation
  - ECMO

Visits

- The span of time a person continuously receives medical services from one or more providers at a care site in a given setting within the health care system.
- Concept set requirements for study:
  - Hospitalization

Defining Malnutrition

- Step 1: do a literature search to see what's been used elsewhere
- Step 2 (may not be applicable): look at local data with more homogenous reporting to build a preliminary set of concepts
- Step 3: build preliminary concept set in Atlas; parent code:
  - Nutritional deficiency disorder (SNOMED 70241007)
- Step 4: refine concept set with SME. Refine some more.
- Step 5: deploy in N3C Concept Set Browser

Exercise 1 (follow along)

Using Atlas, build a starting concept set for malnutrition.

1. Go to Atlas: https://atlas-demo.ohdsi.org/
2. In the Search window, look up Nutritional deficiency disorder (SNOMED CT 70241007)
3. Include all descendant concepts and click “Add to Concept Set”
4. View concept set and extract it as a csv using the “Included Concepts” tab

Exercise 2

Using Atlas, refine previous concept set to include only Deficiency of Macronutrients.

1. Starting with Nutritional deficiency disorder (SNOMED CT 70241007) from the search window, click the concept and navigate to the ”Hierarchy” tab
2. Go to the descendant concept “Undernutrition” and navigate to the ”Hierarchy” tab
3. Create a new concept set using Deficiency of macronutrients (SNOMED CT 238107002) and add all descendant concepts
4. Extract concept set as a csv

Exercise 3 (follow along)

Single-Table Query

1. Create a new folder in your workspace and call it Session 3. Within this, create a new code workbook called “Malnutrition Conditions.”
2. Import the concept table from the data catalog:
    1. Click ”Import dataset”
    2. Navigate to the data catalog and find the folder “OMOP Concepts”
    3. Select concept and import into your code workbook
3. Select the concept table and select “new transform” then select “SQL code.” Name the transform exercise_3
4. Write a query to return the concept_id for Nutritional deficiency using the vocabulary_id ”SNOMED.” Save the results to a dataset by toggling “Save as dataset” and run the transform.

Exercise 4

Single-Table Query

1. Import the concept_ancestor table from the data catalog:
    1. Click ”Import dataset”
    2. Navigate to the data catalog and find the folder “OMOP Concepts”
    3. Select concept_ancestor and import into your code workbook
2. Select the concept_ancestor table and select “new transform” then select “SQL code.” Name the transform exercise_4
3. Write a query to return every descendant_concept_id for Nutritional deficiency using concept_id returned in exercise 3. Save the results to a dataset by toggling “Save as dataset” and run the transform.

Exercise 5 (follow along)

Multiple-Table Query

1. Select the resulting transformation from exercise 4 (should be called exercise_4) and select “new transform” then select “SQL code.” Name the transform exercise_5.
2. On the new exercise_5 transformation, add an additional input for the dataset from for the concept table used in exercise 3.
3. Write a query to return the resulting set from exercise 4 using a JOIN with the concept table and only include those concepts that are standard concepts using the standard_concept column (standard_concept = ‘S’). Save the results to a dataset by toggling “Save as dataset” and run the transform. (Note: don’t be surprised when this is the same result as exercise 4 since we are sticking within one vocabulary)

Exercise 6

Multiple-Table Query

1. Import the condition_occurrence table from the data catalog:
    1. Click ”Import dataset”
    2. Navigate to the data catalog and find the folder “De-Identified Data”
    3. Select condition_occurrence and import into your code workbook
2. Select the condition_occurrence table and select “new transform” then select “SQL code.” Name the transform exercise_6.
3. Add a second input on the new transformation using the resulting transform from exercise 5 (should be called exercise_5). This can be done by clicking the “+” icon in the SQL transform or going to the “Inputs” tab at the bottom of the transform.

Exercise 6

4. Select a data partner to be discussed in class and rite a query to return every row in the condition_occurrence table from that data partner with a matching condition_occurrence_concept_id from the resulting set of exercise 4. You will need to filter on data_partner_id (either in the WHERE clause or in the JOIN). Only include malnutrition records prior to the start of the COVID-19 pandemic in the US (hint: condition_occurrence_start_date prior to ‘2020-01-01’). Save the results to a dataset by toggling “Save as dataset” and run the transform.

Bonus Exercise (follow along)

We will do the same thing, but using the concept set browser.

HOMEWORK 1

Follow the steps above, but this time for the micronutrient deficiency SNOMED CT code you identified in exercise 2. SNOMED CT: 70241007. Save the results in a new workbook and name it ”Micronutrient Deficiency.”

HOMEWORK 2

Go through the entire process (exercise 1 through 6) with the parent code for rheumatoid arthritis (69896004): https://snomedbrowser.com/Codes/Details/69896004. Save the results in a new workbook and name it “Rheumatoid Arthritis.”

Logic Liaison Template

Rationale

- We just defined one concept, meaning we only 29 to go for this very basic study!
- The Logic Liaison Template seeks to streamline this process while enabling best practices in line with the literature (e.g., what constitutes a COVID hospitalization)
- Documentation available here: https://unite.nih.gov/workspace/module/view/latest/ri.workshop.main.module.3ab34203-d7f3-482e-adbd-f4113bfd1a2b

LL Templates

- [LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: COVID-19 Diagnosed or Lab Confirmed Patients
  - This LOGIC LIAISON template produces a table of commonly used variables for COVID-19 Positive Patients (diagnosed with U07.1 and/or rt-PCR/AG confirmed). Both level 2 and level 3 versions of the template are provided in this object, as well as the final tables they produce.

- [LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: All Patients
  - This set of templates (Level 2 and Level 3 data versions) provides sample code and summary datasets including a visit-level and a patient-level table. The visit-level "all facts" table has a single row for each patient and each visit day (where any of the facts searched are found to be present). The patient-level table contains one row for each patient and a number of commonly referenced facts and indicators derived from the N3C datasets
- [Logic Liaison Template] Data Density by Site and Domain
  - This template calculates the Standardized Density, Median Absolute Deviation (MAD), and Directional Median Deviations (DMD) with respect to the number of unique patient/concept/days for each of the major OMOP tables (i.e. condition_occurrence, drug_exposure, etc) and uses them to create a heatmap displaying how many MADs each site is from the median for each OMOP table. The template also scores the site's date shifting practices.
- [Logic Liaison Template] Fact Density by Site Visualization
  - This template calculates the Standardized Density, Median Absolute Deviation (MAD), and Directional Median Deviations (DMD) with respect to the numerical values in each column of the input table (any non-numerical field is converted to a binary value using the isNotNull() function) and creates heatmaps to visualize the metrics.

Exercise 7 (follow along)

Deploy the Logic Liaison Template in your student folder!

- Create a new code workbook and name it Logic Liaison COVID Positive
- Skip the step to import a dataset

Importing into a Code Workbook

- Open a new code workbook and select “Skip this Step” under "Import Dataset".
- Make sure you are using the "default" or "high-memory" environment (selected under the environment menu towards the top middle of the window: Customize Spark environment → profile-high-memory on the left panel of pop-up → Update Spark environment).
- Click “New Transform”, select “Templates”, search for the “[LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: COVID-19 Diagnosed or Lab Confirmed Patients" and import into the workbook. When you press the ‘apply transformation’ button, a box pops up that says that there are resources within the template that are not within the scope of the project. You can agree to add these resources.

Required Modifications for Malnutrition Study

- Update concept sets in use
  - Concepts in Use: https://github.com/National-COVID-Cohort-Collaborative/short-course-2022-june/blob/main/lessons/session-3/malnutrition_concept_sets.csv
  - Using the “Manual entry” option in a code workbook, paste the list from GitHub and name the transformation ll_input.
  - Toggle “Save as dataset” and run the transform
- Modifications
  - In the cohort template, change proportion_of_patients_to_use to 0.10.
  - Point all templates to the newly created ll_input under customize_concept_sets.
  - In visits_of_interest template, toggle off the covid_associated_ED_or_hosp_requires_lab_AND_diagnosis, change num_days_before_index to 3 and num_days_after_index to 14.
  - Save all transformation as datasets and click ”Run all saved datasets” from the cog icon at the top of the code workbook

HOMEWORK 3

1. Create a new code workbook and name it Logic Liaison All Patients
2. Skip the step to import a dataset
3. Import the template: [LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: All Patients
4. Keep the template as is, but toggle all datasets to save
5. Run all transforms

UNMC[SM] | COLLEGE OF MEDICINE

University of Nebraska Medical Center