National COVID Cohort Collaborative

Introduction to Analyzing Real-World Data Using the National COVID Cohort Collaborative (N3C)

Spring 2024 - N3C Education and Training Domain Team

Instructor: Jerrod Anzalone

Module 5 In-Class Exercises and Homework Instructions

Step 1. Pre-Class Requirements

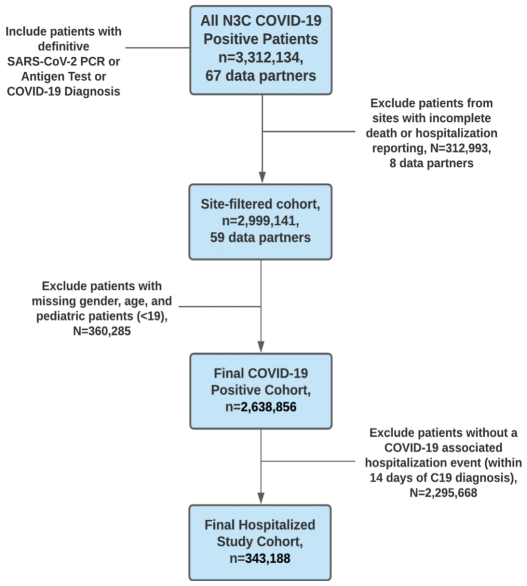
- Run LL COVID+ Template, adding concept set for 'Malnutrition (primary protein deficiency)'
- If you weren't able to do that, no worries! We have a pre-processed version available for use today. Review the logic here: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.d4bfd46c-6422-420c-9bb5-a64d516acabc?branch=master

Step 2. Site Level Data Quality Assessment

- We have three different approaches we can use:
  - Systematic Missingness by Site and Study Variable
  - Fact Density by Site
  - Domain Density by Site (this is the one we're using today)
- Follow along here: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.089dbc6c-04df-4bf9-916a-d3646e4c37fc?branch=master

Step 3. Inclusion and Exclusion Criteria

- We want to be able to recreate this diagram and apply the study inclusion/exclusion criteria:



- Follow along here: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.da1c122d-fafb-4d1f-9518-6e7fdf61cb4e?branch=master
- Code for individual transforms available here on GitHub.
- Create a new code workbook in your folder and call it 'Step 3: Inclusion and Exclusion Criteria.'
- Start by importing Logic_Liaison_Covid_19_patient_summary_fact_table_De_Id and Logic_Liaison_Covid_19_patients_visit_day_facts_table_De_Id.
- Change profile to 'profile-high-memory.'
- Create a new SQL transform and name it analytic_ll_person_table. Add Logic_Liaison_Covid_19_patient_summary_fact_table_De_Id as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it analytic_ll_visit_table. Add Logic_Liaison_Covid_19_patients_visit_day_facts_table_De_Id and analytic_ll_person_table as inputs.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!

Step 4. Add data elements for the study

- Now we need to add the data elements required for this study, which are detailed in the slide deck for this week.
- Follow along here: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.fac49b87-11b6-4acc-85d5-75066bb32767?branch=master
- Code for individual transforms available here on GitHub.
- Create a new code workbook in your folder and call it 'Step 4: Add data elements for study.'
- Start by importing Analytic_ll_person_table and Analytic_ll_visit_table.
- Change profile to 'profile-high-memory.'
- Create a new SQL transform and name it state_cleaning. Add Analytic_ll_person_table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it census_divisions. Add state_cleaning table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it cci_calculation. Add Analytic_ll_person_table table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it race_ethnicity. Add Analytic_ll_person_table table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it covid_hospitalizations_within_14_days. Add Analytic_ll_visit_table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!

- Create a new SQL transform and name it covid_death_within_30_days. Add Analytic_ll_visit_table as input.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it malnutrition_exposure. Add Analytic_ll_visit_table and Analytic_ll_person_table as inputs.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!
- Create a new SQL transform and name it analytic_dataset. Add all the transforms you just made in step 4 as inputs.
- Copy the code from the workbook or GitHub. Toggle 'Save as dataset' and click run!

Step 5. Statistical analysis using R

- Now that we have a clean dataset filtered to what we need, we are ready to analyze it!
- Follow along here: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.1643e613-b276-4fa3-8baa-e291b98b72e7?branch=master
- Code for individual transforms available here on GitHub.
- Create a new code workbook in your folder and call it 'Step 5: Data Analysis.'
- Start by importing Analytic_dataset.
- Copy this code into the global code window:

```
to_rdata <- function(data) {
    output <- new.output()
    output_fs <- output$fileSystem()
    saveRDS(data, output_fs$get_path("data.rds", 'w'))
}

from_rdata <- function(data) {
    fs <- data$fileSystem()
    path <- fs$get_path("data.rds", 'r')
    rds <- readRDS(path)
    return(rds)
}
```

- Change profile to ‘publications-and-presentations’ and leave as is or ‘r-high-driver-memory’ if you don’t have the ‘publications-and-presentations’ profile. If you’re using ‘r-high-driver-memory’ you need to add the ‘r-gtsummary’ package to the profile for this exercise.
- Create a new R transform and call it local_overall.
- Copy the code from the workbook or GitHub. Toggle ‘Save as dataset’ and click run!
- Create a new R transform and call it table_1_overall.
- Copy the code from the workbook or GitHub. Toggle ‘Save as dataset’ and click run!
- Create a new R transform and call it death_glm_overall.
- Copy the code from the workbook or GitHub. Toggle ‘Save as dataset’ and click run!
- Create a new SQL transform and call it table_1_overall_clean. Add table_1_overall as input.
- Copy the code from the workbook or GitHub. Toggle ‘Save as dataset’ and click run!
- Create a new SQL transform and call it death_glm_overall_clean. Add death_glm_overall as input.
- Copy the code from the workbook or GitHub. Toggle ‘Save as dataset’ and click run!
- DONE!

Homework Assignment (due February 22, 2024)

We have submitted our paper and received requests for major revisions! One reviewer leaves the following feedback:

> *"The authors examine the associations between malnutrition and several acute COVID-19 outcomes. However, they lump together many disparate conditions related to malnutrition, limiting their findings' clinical interpretability. I'd recommend looking at malnutrition severity among those with a history of malnutrition to identify potential interventions necessary among those presenting with a history of malnutrition (see Hypothetical Study #1, Hypothetical Study #2, Hypothetical Study #3 for previously used classification methods). I don't see how this contributes to our existing understanding of medical nutrition therapy without…"*

There's always one difficult reviewer… to address their concerns, we devise a new classification scheme for a series of sensitivity analyses to appease their concerns. Extending our original concept set, we review the literature and come up with this classification scheme for our malnutrition concepts:

Severity	Concept ID	Condition Concept Name
Mild Malnutrition	437832	Malnutrition of mild degree (Gomez: 75 percent to less than 90 percent of standard weight)
	4096196	Mild protein-calorie malnutrition (weight for age 75-89 percent of standard)
Moderate Malnutrition	436078	Malnutrition of moderate degree (Gomez: 60 percent to less than 75 percent of standard weight)
	4101278	Moderate protein energy malnutrition
	4098458	Moderate protein-calorie malnutrition (weight for age 60-74 percent of standard)
Severe Malnutrition	4233565	Severe protein-calorie malnutrition (Gomez: less than 60 percent of standard weight)
	4123542	Wasting disease
	432593	Kwashiorkor
	4029268	Marasmic kwashiorkor
General Malnutrition	4156515	Malnutrition (calorie)
	4276360	Undernutrition
	435227	Nutritional deficiency disorder
	433163	Deficiency of macronutrients
	4028220	Malnutrition following gastrointestinal surgery
Starvation and Related Conditions	443082	Starvation
	4337279	Semi-starvation

It's your job to extend our existing work to include these sensitivity analyses. You are on the study team and were assigned to examine one of the following cohorts:

1. The multilevel severity measure to compare them collective versus no hx of malnutrition (level 1)
2. One of these severity types to compare individually to those with no hx of malnutrition (level 2)
3. Stratified among only those with malnutrition to compare relative differences among those with malnutrition (level 3)

Requirements:

- If your last name is between A and F, you are group A (level 1)
- If your last name is between G and P, you are group B (level 2)
    - G and H: Mild Malnutrition
    - I and J: Moderate Malnutrition
    - K and L: Severe Malnutrition
    - M and N: General Malnutrition
    - O and P: Starvation and Related Conditions
- If your last name is between Q and Z, you are group C (level 3)

Based on your group assignment:

- Create a Table 1 for your cohort (e.g., descriptive statistics) with an overall column and comparison groups. Include *p* values using the gtsummary package. Be sure to include additional code to censor small counts.
- Using the approach we covered in class, run crude and adjusted logistic regression for 30-day all-cause mortality using the exposure group you were assigned above. If you include event rates, include an additional step to censor small cell counts.
- Please replicate the analyses for your cohort following the example in this code workbook: https://unite.nih.gov/workspace/vector/view/ri.vector.main.workbook.1643e613-b276-4fa3-8baa-e291b98b72e7?branch=master
- At the end of this exercise, you should have two ready to include in your updated analysis.

Note:

- For those in level 2, you will need to subset this to those without a hx of malnutrition and *your assigned group above* (e.g., you will need to remove those patients with a hx of malnutrition in a group that is not part of your cohort). I'd recommend doing this as follows using dplyr (use your assigned group for the second item in this list):

```
filter(malnutrition_type %in% c("No Malnutrition Documented", "Mild Malnutrition"))
```

- For those in level 3, you must first subset to those with some documented hx of malnutrition (e.g., remove those with "No Malnutrition Documented"). Please set your factor reference to have 'Mild Malnutrition' as the reference.