# N3C Logic Liaison Templates: Transforming Complex Health Data into Analytics Ready Tables

Andrea G. Zhou, M.E.[1], Evan T. French, M.S.[2], Richard A. Moffitt, Ph.D.[3], Johanna J. Loomba, M.E.[1]

[1]integrated Translational Health Institute of Virginia (iTHRIV), University of Virginia, Charlottesville, VA, USA;
[2]Virginia Commonwealth University, Richmond, VA, USA; [3]Emory University, Atlanta, GA, USA

UNIVERSITY of VIRGINIA · EMORY UNIVERSITY SCHOOL OF MEDICINE · iTHRIV · VCU

## Background

National COVID Cohort Collaborative

N3C is one of the largest collections of clinical data for COVID-19 research in the United States [1].

> 20 million

## Problem

- Limited toolset available for transforming billions of rows of data

- Significant missing data elements without tools to characterize and clean data

## Goal

### N3C Logic Liaisons [2]

- Provide rapid ways to transform raw information into usable facts

- Help researchers easily evaluate quality and filter data

## Community Fact Table Template

- Generates > 120 commonly used variables

- Provides method for quickly adding custom elements

- Utilizes specific Concept Sets (lists of key variables from standard vocabularies)



Messy Data → N3C Harmonized OMOP Data, External Public Datasets, Data Partner Characteristics → Data Selection, Cleaning, and Other Transformations → Cohort Identification with Base Characteristics ← customizable concept set input → Facts of Interest from Each Domain → Organized, Multiple Rows per Person Data → Date-level Fact Table → Organized, One Row per Person Data → Person-level Fact Table

## Interoperable Set of Logic Liaison Templates



**Overall Quality Templates**
- Data Density by Site and Domain
- Whitelist Filtering

**Fact Table Templates**
- Confirmed Covid Positive
- All Patients
- SDoH

**Ancillary Fact Templates**
- Combined Variables
- CCI
- Study-specific Fact Indexing

**Ancillary Quality Templates**
- Systematic Missingness
- Fact Density by Study Variable

Data Catalog Resources → Select Sites → Filtered Resource Datasets → Derive Variables → Base Fact Tables → Create Complex Variables → Augmented Fact Tables → Select Data for Question → Question 1 Fact Tables → Apply Quality Filters to Fact Tables → Question 1 Analysis Dataframe → Analyze → Question 1 Result

### Logic Liaison Template Pipeline for N3C Research

After being documented and peer-reviewed for quality, templates are published to the N3C Knowledge store where they can be accessed by the entire N3C community along with ancillary templates to augment, examine, and clean the data prior to analysis. The application of these templates throughout the research process are illustrated in the figure above.

## Impact

**Confirmed COVID+ Patients Fact Table Template**
(released in October 2021)

65 teams → 315 uses

**All Patients Fact Table Template**
(released May 2022)

49 teams → 110 uses

## Discussion

- Eliminated redundant efforts and accelerated project timelines with flexibility to fit custom needs.

- Aligned key definitions of COVID-associated hospitalization, SARS-COV2 vaccination, COVID-19 reinfection, etc.

## Future Direction

- Expansion to include more sources such as linked Medicare and Medicaid claims data.

- Produce disease agnostic version of templates for use in the N3C Clinical Pilot

References:
1. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, et al. The National Covid Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. Journal of the American Medical Informatics Association. 2020;28(3):427-43.
2. Data Liaisons and Logic Liaisons Services [Internet]. N3C. 2022 [cited 2023Feb24]. Available from: https://covid.cd2h.org/liaisons

NATIONAL CENTER FOR DATA TO HEALTH · OHDSI Observational Health Data Sciences and Informatics · pcornet The National Patient-Centered Clinical Research Network · TriNetX · ACT The ACT Network