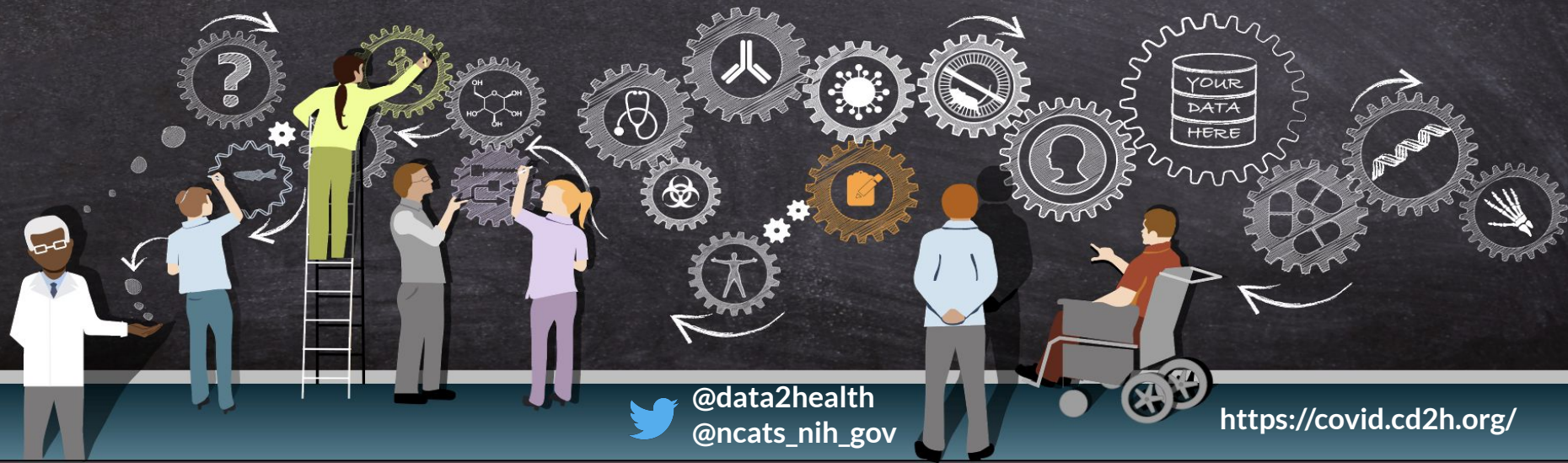


N3C Short Course



@data2health
@ncats_nih_gov

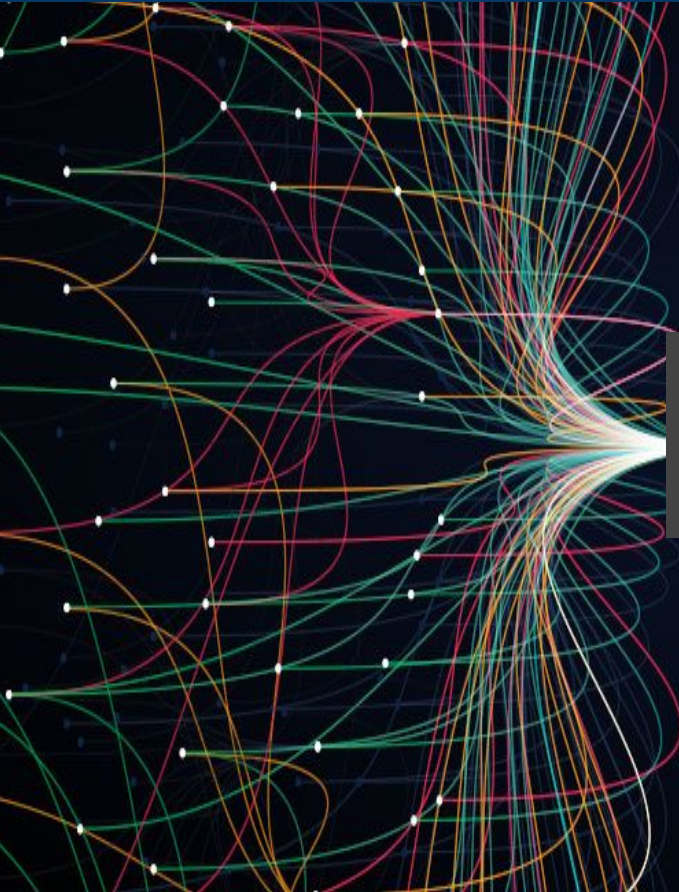
<https://covid.cd2h.org/>



National
COVID
Cohort
Collaborative

program of NIH's National Center for
Advancing Translational Sciences

Study Design & Project Planning



Harold Lehmann, MD, PhD
Johns Hopkins University

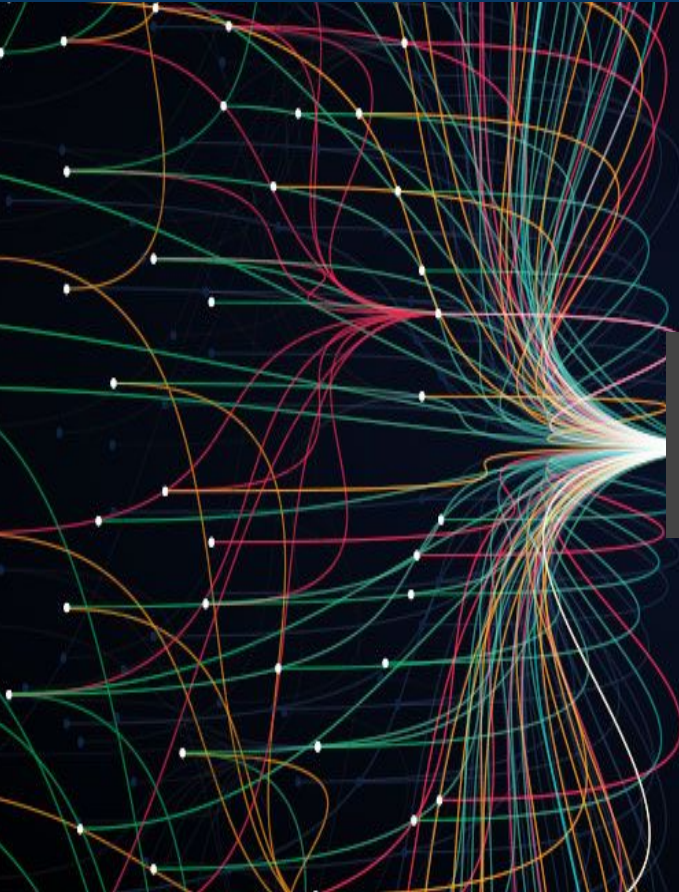




National
COVID
Cohort
Collaborative

program of NIH's National Center for
Advancing Translational Sciences

Or: How N3C Researchers (Try to) Organize Their Work



Harold Lehmann, MD, PhD
Johns Hopkins University



STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies

Shirley V Wang,¹ Simone Pinheiro,² Wei Hua,² Peter Arlett,^{3,4} Yoshiaki Uyama,⁵ Jesse A Berlin,⁶ Dorothee B Bartels,⁷ Kristijan H Kahler,⁹ Lily G Bessette,¹ Sebastian Schneeweiss¹

Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, Bartels DB, Kahler KH, Bessette LG, Schneeweiss S. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021 Jan 12;372:m4856. doi: 10.1136/bmj.m4856. PMID: 33436424.



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Study Design

See

<https://github.com/ohdsi-studies/TicagrelorVsClopidogrel>

Outline

Project Name

Milestones

☰ Milestones

10) Table of Contents

20) List of abbreviations

25) Research Question and Hypothesis

28) Study Administration

28.20) Study Team Composition

Team members by roles

Attributions

Funding and acknowledgments

40) Versions, stages, and progress

Protocol Versions and Amendments

☰ Protocol Versions and Amen...

40.1) Study Stages and Outputs

40.1) Stage 1: Specify Protocol

40.2) Stage 2: Execute Protocol

40.3) Stage 3: Disseminate results

50) Background and Rationale

Specific Gap

60) Research questions and objectives

60.1) Research questions

Aims

70.3 Study Hypothesi(e)s

70) Methods

70.1) Participants

70.2) Study Design and Data So...

70.3 Study Hypothesis

80) Sample Size and Study Power

90) Data Analysis Plan

90.10) Causal Diagrams

90.20) Table shells

☰ Table 1. Patient characteri...

Mapping study variables to Enclav...

Potential data elements

OMOP Standard Tables

Enclave shared variables

Linking study variables and Enc...

☰ Concepts

Code for COVID + identification.

Code for COVID + identification v3

10) Ethical Statement and Protecti...

12) Plans for Publication

Prepare tables for Review

Publication Review

N3C Manuscript Concept Form

130) References

[Clinical Research Template](#)



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Outline

- What and Why a Lab Notebook
 - *Electronic Lab Notebook = ELN*
- Lab Notebooks outside the Enclave (Foundry)
- Lab Notebooks inside the Enclave
- Some future developments



Why

“ There seems to be a misconception in the computational biology community that a lab notebook is only useful for **recording experimental protocols and their results**. A lab notebook is much more than that. It is an **organizational tool** and **memory aid**, which serves as the **primary record** of scientific research and activity for all scientists. It also serves as a **legal record** of ownership of the ideas and results obtained by a scientist.

Why

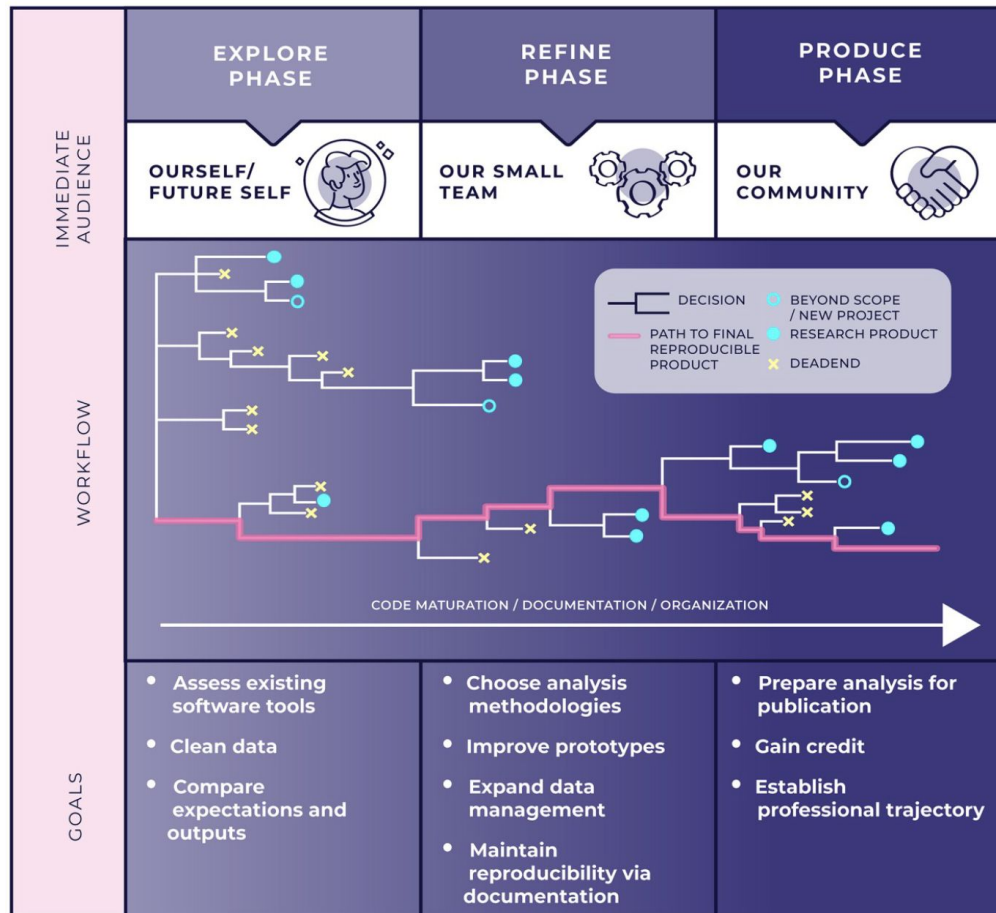
- Record original intent
- Preserve data and observations
- Assist others with understanding and reproducing your observations
- Support intellectual property claims
- Defend against false allegations of research misconduct
- Prime source for writing a dissertation or paper



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Data Project Life Cycle

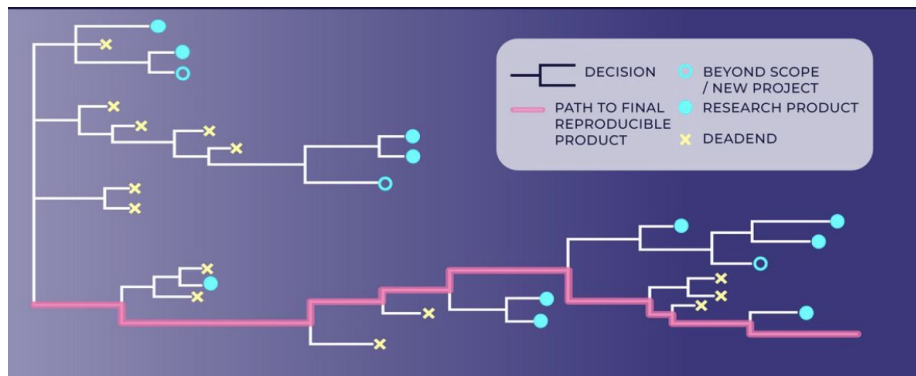


Stoudt S, Vásquez VN, Martinez CC. Principles for data analysis workflows. PLoS Comput Biol. 2021 Mar 18;17(3):e1008770. doi: 10.1371/journal.pcbi.1008770.



What Goes in an Electronic Lab Notebook (ELN)

- Admin issues
 - PIs, IRB#s, DURs, etc.
- High-level issues
 - Hypotheses (formal or not), plans
 - Intentions for participants, variables
 - Work plans, assignments
- Definitions
 - Results of data cleaning
 - Choices for operationalizing variables
 - Concept sets, Knowledge Store templates, code workbooks
 - Choices of analytic methods
 - Roads not taken
- Reality
 - Results of analyses
 - Sensitivity analyses
 - Conclusions
 - Limitations





National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

ELN Outside the Enclave

PROs

- Familiar
- Prior formats can be repurposed
- Accessible to folks without Enclave access
- Can link to resources inside the Enclave

CONs

- Resources within the Enclave are not accessible to users without DUR approval
- No analytic decision support



Media Outside the Enclave

- MS Word
 - Track changes
 - Outline view
 - Hyperlink to resources
- MS OneNote
- Google sheets (Excel)



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Google Doc Example

Stroke N3C Notes

(REMINDER: NO PATIENT DATA OUTSIDE ENCLAVE PRE-EXPORT APPROVAL)

Tools to make concept sets (groups of codes that make up a variable):

<http://atlas-covid19.ohdsi.org/#/home>

<https://athena.ohdsi.org/search-terms/start>

For N3C cohort phenotype see [Github](#).

HYPOTHESES

The association between ischemic stroke severity and concurrent COVID-19 persisted throughout the first year of the pandemic even after stroke systems of care normalized. We further hypothesize that a similar association existed between hemorrhagic stroke severity and COVID-19.

STUDY COHORT

Population for all analyses will be limited to patients in N3C dataset who were admitted for stroke from Mar 2020 - Feb 2021.

CONCEPT SETS

[Stroke Project Data Dictionary \(with code sets\)](#)

STUDY DESIGN and PLANNED FIGURES

- 1) Table 1: Stats on four sub-groups of Hospitalized Stroke Patients (Rows of Percentages: Race/Ethnicity, Average Age, Gender, CoMorbidity, NIHSS upon admission, Death, IMV/ECMO, LOS, all variables). TABLE COLUMNS:
 - a) Hemorrhagic Stroke (COVID+)
 - b) Hemorrhagic Stroke (COVID-)
 - c) Ischemic Stroke (COVID+)
 - d) Ischemic Stroke (COVID-)
- 2) Figure 1: TIME RELATIONSHIP between COVID diagnosis and stroke (Charts of PREMATCHED concurrent stroke patients - stroke 1 week prior to 3 months after COVID+)
 - a) Hemorrhagic Stroke (COVID+)
 - b) Ischemic Stroke (COVID+)
- 3) Figure 2: Bar graph - NIHSS split up by 2 month blocks. 4 bars in each 2 month block
 - a) Hemorrhagic Stroke (COVID+)
 - b) Hemorrhagic Stroke (COVID-)
 - c) Ischemic Stroke (COVID+)
 - d) Ischemic Stroke (COVID-)

Johanna Loomba. Stroke Domain Team.



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Google Sheets Example

SDoH Concept set mapping and inventory ☆ ⓘ ☁

File Edit View Insert Format Data Tools Extensions Help Last edit was made on 10 May by Umit Topaloglu

100% \$ % .0 .00 123 Arial 10 B I S A

	A	B	C
1		PLEASE NOTE THAT CONTRIBUTIONS TO THIS PAGE ENDED IN SUMMER OF 2020...PLEASE LOOK IN THE ENCLAVE FOR FURTHER WORK	
2	Status	Domain	Variable
3		SDOH	
4	Has prior development	Income	Disability income
5	Has prior development	Income	Median household income
6	NEED	Income	Individual income
7	NEED	Income	Percent of population less than 100% Federal Poverty Level
8	Has prior development	Employment	Percent unemployed
9	Has prior development	Employment	Employment status
10	NEED	Employment	Percent of essential service employees
11	Has prior development	Employment	Current occupation type
12	Has prior development	Housing Insecurity / Instability / Homelessness	Housing insecurity (broad)
13	Has prior development	Housing Insecurity / Instability / Homelessness	Homelessness (broad)
14	Has prior development	Housing Insecurity / Instability / Homelessness	Current housing status
15	Has prior development	Housing Insecurity / Instability / Homelessness	Length of time homeless
16	NEED	Housing Quality	Percent population living in renter occupied and crowded housing units
17	NEED	Housing Quality	Percent single-parent households with dependents
18	NEED	Housing Quality	Percent multigenerational households
19	Has prior development	Food insecurity	Food insecurity
20	NEED	Transportation	Percent of population with no vehicle
21	Has prior development	Health Care / Medicine Access & Affordability	Health Care / Medicine Access & Affordability
22	NEED	Health Care / Medicine Access & Affordability	Has health insurance

+ ☰ Concept set READMEs [SELECT REFERENCE COLUMNS] N3C Public Concept set ⓘ Explore

Jimmy Phuong. SDOH Domain Team. July, 2020.



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

ELN Inside the Enclave

PROs

- Integrated with Enclave resources
- Potential for analytic decision support
- Can link to resources inside the Enclave

CONs

- Not accessible to users without DUR approval
- Options are not as functional as external tools



ELN Inside the Enclave: Fusion Sheet

Pro

- Looks like Google Sheet
- Can update datasets, and vice versa

Con

- Hyperlinking is not so great

Project - Coh... > analyses > ... > Codeset and alias table

File Help 1

Edit View Insert Table Data 0 Document

Spreadsheet view Read-only Share

Undo Redo Bold Italic Underline 12 Text Fill Grids Borders Background Color Font Color Data type Auto

This cell is read-only

	A	B	C	D	E	F	G	H
	Alias	codeset ID	ATC	Data type	Domain	Include in Cohort	MainFigureTrajectory	Link to workbook
1	Pneumonia	920959633		Bool	diagnosis	TRUE		
32	Bilirubin - (Unconjugated/Indirect), mg/dL	59108731		Qn	Lab	FALSE		https://unite.nih.gov/workspac
33	BNP (all forms), pg/mL	248128869		Qn	Lab	FALSE		https://unite.nih.gov/workspac
34	eGFR If NonAfricn Am, mL/min/1.73	784232790		Qn	Lab	FALSE		https://unite.nih.gov/workspac
35	MCV, fL	258288169		Qn	Lab	FALSE		https://unite.nih.gov/workspac
36	SvO2, %	207,642,726		Qn	Lab	FALSE		
37	Troponin T >0.01	458505731		Nom	Lab	FALSE		
38	Bilirubin (total), mg/dL	586434833		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
39	BNP, pg/mL	703853936		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
40	c-reactive protein CRP, mg/L	371622342		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
41	Creatinine, mg/dL	615348047		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
42	D-Dimer, mg/L FEU	475972797		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
43	Ferritin, ng/mL	317388455		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
44	Lactate, mM	400660753		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
45	White blood cell count, x10E3/uL	138,719,030		Qn	Lab	TRUE	2b	https://unite.nih.gov/workspac
46	Albumin (g/dL)	104464584		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
47	ALT (SGPT), IU/L	538737057		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
48	AST (SGOT), IU/L	248480621		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac
49	Bilirubin (Conjugated/Direct), mg/dL	50252641		Qn	Lab	TRUE	sup2	https://unite.nih.gov/workspac



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

ELN Inside the Enclave: Reports

Pro

- Editable by markdown
- Direct links to resources
- Default content can point to resources (analytic decision support)

Con

- Editable by markdown

Harold Lehmann, Eli Levitt.
For Applicable Data Methods
and Standards, 2021.

Research Template 2

2/2/2021 Forked from Kstore/Staging/Clinical Research Template to accommodate greater focus on in-Enclave work

- ☐ The following is a template for documenting your study. This outline follows one provided by Andrew Williams (10/28/2020) and is based on "Instructions" (more like suggestions) are provided in italics, like here.
- ☐ As much as possible, we have tailored this Template to direct you to appropriate Enclave tools and resources at the appropriate points in the research or analytic decision support: Providing support at the right place and time and to the right person.
- ☐ Contributors to this template:

Eli Levitt: Initial draft protocol

Andrew Williams: Initial content

Harold Lehmann: Initial abstraction of protocol into a template

Contributors: **<add your name here>**

Highlighted text is either a recent update or text that needs attention.

Project Name

- ☐ Use the name from the DUR and as approved by DAC. *My Projects* (Enclave) or *N3C Data Enclave Projects* (Public)

Short title:

- ☐ Your choice. If in a Domain Team, use the study name as articulated by the team.

Version summary:

- ☐ The summary here is for the entire document. You will be providing milestones for the project itself, below.

Outline

Project Name

Milestones

Milestones

10) Table of Contents

20) List of abbreviations

25) Research Question and Hypothesis

28) Study Administration

28.20) Study Team Composition

Team members by roles

Attributions

Funding and acknowledgments

40) Versions, stages, and progress

Protocol Versions and Amendments

Protocol Versions and Amen...

40.1) Study Stages and Outputs

40.1) Stage 1: Specify Protocol

40.2) Stage 2: Execute Protocol

40.3) Stage 3: Disseminate results

50) Background and Rationale

Specific Gap

60) Research questions and objectives

60.1) Research questions

Aims

70.3 Study Hypothesi(e)s

70) Methods

70.1) Participants/PICO

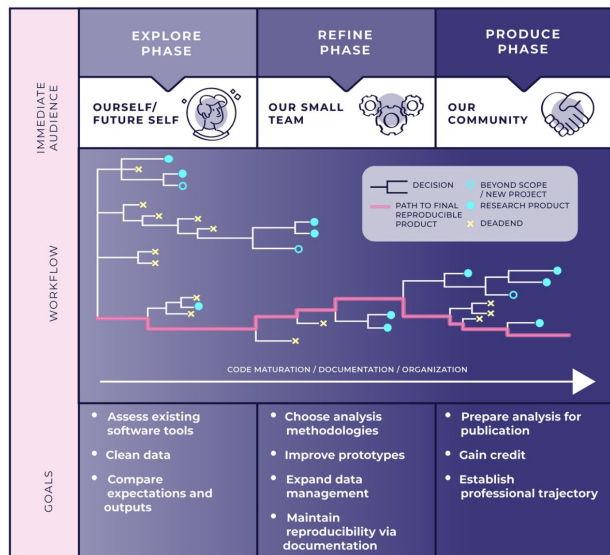
70.10) Causal Diagrams



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

Future: Computable Protocol



Stoudt S, Vásquez VN, Martinez CC. Principles for data analysis workflows. PLoS Comput Biol. 2021 Mar 18;17(3):e1008770. doi: 10.1371/journal.pcbi.1008770. PMID: 33735208; PMCID: PMC7971542.

Protocol Pad [Select Protocol or Create New Protocol](#) [View or Edit Selected Protocol](#) [Edit](#)

Welcome to Protocol Pad! This is a space for all your N3C protocols*, aimed at assisting manuscript production and reproducibility. You can design, develop and finalize protocols over any of your research projects, all from one location!

METRICS FOR YOUR PROTOCOLS

- 7 PROTOCOLS
- 5 RESEARCH PROJECTS CONTAIN YOUR PROTOCOLS
- 0 PROTOCOLS SHARED TO THE KNOWLEDGE STORE
- 0 SHARED PROTOCOLS ARE USED IN OTHER RESEARCH PROJECTS

Your Protocols

RESEARCH PROJECT NAME
Select...

PROTOCOL STATUS
Select...

(7) Created At

Retrofit: JAMA Internal Medicine: Immune Dysfunction

Protocol ID - 390e8cca-35e7-4962-9566-63b95df8bd5d
Created At - Apr 27, 2022, 12:26 AM
Link To DUR - DUR-140E37D
Research Project - RP-919BAE
Research Project - [\[RP-919BAE\] Immunosuppressed or -compromised \(ISO\)](#)
Protocol Status - Finished

Impact of Demographics and Socioeconomic Factors on COVID-19 Patient Intake, Follow-up and Retention; A Report from

[Create New Protocol](#)

Background Rationale

A large evidence gap exists for patients with immune dysfunction because they were largely excluded from SARS CoV- 2 vaccine clinical trials

Objectives

identify the incidence rate (IR) and incidence rate ratio (IRR) for COVID-19 breakthrough infection after SARS-CoV-2 vaccination among persons with or without immune dysfunction

Looking for help on protocols? [View Examples of Protocols in the Knowledge Store](#) [Provide Feedback or Ask Questions on Protocol Pad](#) [Read our Guide to Making a Protocol](#)

*Background on computable protocols: Lehmann, Manaa, Wilkins, Bradwell, Amor, Loomba, Levitt, Williams, ADMS, N3C. A research-protocol object to generate biomedical knowledge that is auditable and reproducible. MCBK. 2021.

Kate Bradwell, Harold Lehmann. Protocol Pad.



National
COVID
Cohort
Collaborative

A program of NIH's National Center for
Advancing Translational Sciences

lehmann@jhmi.edu