



National
COVID
Cohort
Collaborative

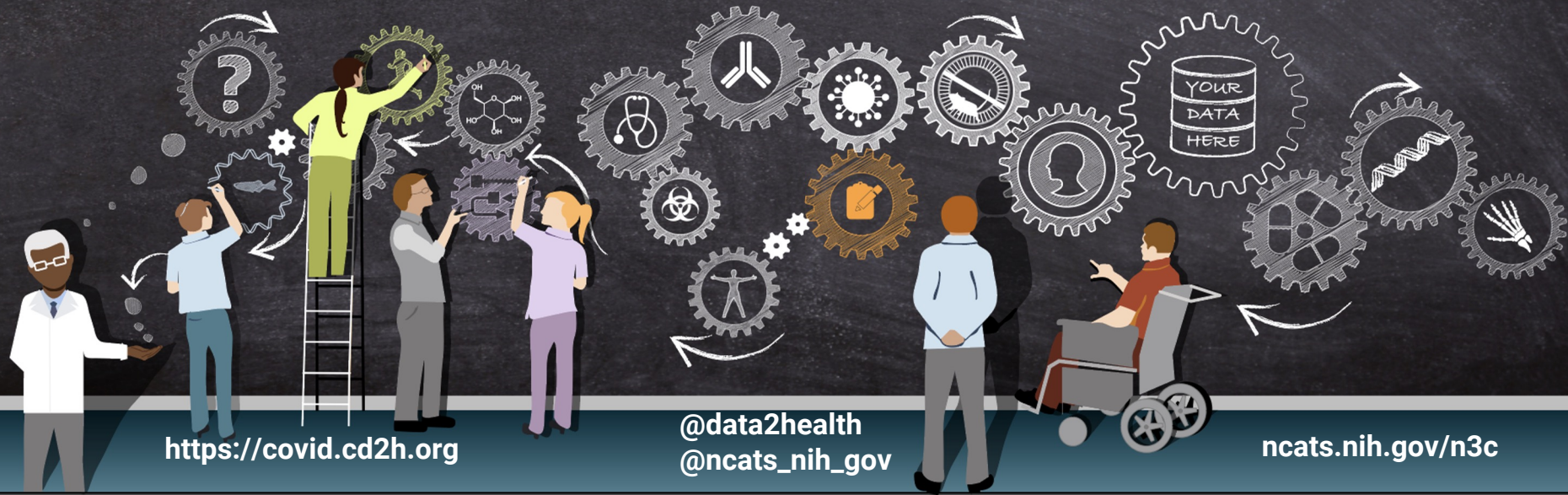
A program of NIH's National Center for Advancing
Translational Sciences

N3C Publication Committee Overview



NATIONAL CENTER
FOR DATA TO HEALTH

February 21, 2024



<https://covid.cd2h.org>

@data2health
@ncats_nih_gov

ncats.nih.gov/n3c



National
COVID
Cohort
Collaborative

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

Committee Members



NATIONAL CENTER
FOR DATA TO HEALTH



Swaroop Vedula
Machine learning, epidemiology



Jerrod Anzalone



Carolyn Bramante
*Clinician investigator
internal medicine*



Jeremy Harper
*Research
Biomedical Informatics*



Wendy Hernandez



Farrukh Koraishy
Data research design



Julie McMurry
Technical Leadership



Mary Emmett
*Laboratory medicine,
initiated research*



Amit Kumar Saha
*Data analytics,
statistical analysis*



Christine Suver
Governance Lead



Melissa Haendel
N3C Leadership



Chris Chute
N3C Leadership



Anita Walden
N3C Leadership



Tursynay Issabekova
Project Manager



Jin Ge



William Beasley



Andrew Barros



Ran Dai



Hemal Mehta



Jasvinder Singh



Meng-Hao Li



Nirup Menon



National
COVID
Cohort
Collaborative

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

N3C Publication Committee



NATIONAL CENTER
FOR DATA TO HEALTH

The N3C Publication Committee was organized to facilitate compliance with N3C requirements

- (1)** Facilitate inclusive and appropriate attribution
- (2)** Troubleshoot authorship questions regarding N3C publications
- (3)** Track the output of the N3C Community

Attribution and Publication principles posted on Zenodo:

<https://zenodo.org/record/7787523>



Committee's Core Responsibilities



The guiding principle is to be stewards of critical components of the N3C:

(1) Data partners and individual patients: Assure that the final form of the data prepared for publication is:

- a) Consistent with the data use that was approved
- b) Consistent with policies to protect the identity of individuals, data partners, and marginalized groups

(2) Individual & consortial contributors: Assure inclusive, commensurate, and appropriate attribution

(3) N3C collaborative: Assure that information about the N3C is represented accurately
Continually evaluate procedures.

(4) Authors: Serve as a resource for authors





Committee's Core Responsibilities



(1) Data partners and individual patients: Assure that the final form of the data prepared for publication is:

- a) Consistent with the data use that was approved
- b) Consistent with policies to protect the identity of individuals and marginalized groups

Examples:

- **Does the work submitted fit within one or more approved data use requests?**
 - Or the authors on the work are not on the DUR
- **Are all the cell sizes ≥ 20 ? (patient subgroup numbers in tables figures)**
 - Cell sizes if less than 20 must be obscured to "<20"
 - Cell sizes must not be back-calculated to less than 20 unless a marginal group such as 'missing, unknown, or other'
 - Note, authors can appeal to NCATS to be authorized to show cell size <10 if there is a legitimate scientific reason to do so
- **Native populations check**
 - If native populations are specifically referenced, is it done in a way that acknowledges the NIH tribal council's role in the representation of the data.
- **Does this work identify specific data partners?**
- **Is the work submitted focused on COVID-19?**





Committee's Core Responsibilities



(2) Individual & consortial contributors:

Assure inclusive, commensurate, and appropriate attribution.

- Acknowledge the enormous amount of work to make this data available.
 - This includes codeset/artifact creators and domain team contributors.

Authors may not be familiar with the requirements for acknowledgment and attribution.

- We review whether or not they have the required components.
- We guide them to the appropriate acknowledgements and attribution.

New Step: attribution for projects funded by RECOVER.





Committee's Core Responsibilities



(3) N3C collaborative: Assure that information about the N3C is represented accurately

We review for:

- Accurate description of the N3C collaborative
- Especially if it is about N3C
- Acknowledge funding





Committee's Core Responsibilities



(4) Authors: Serve as a resource for authors

Overlap with other work being done at N3C

- Hard to address at the point of finished manuscript drafts.
- Duplication, replication of research findings is key to science
- We may connect authors with others for collaboration if it makes sense
- We *will* do a deep dive into disparate findings if they relate to nuances of the N3C dataset (not differences in analytic approach)

We are a resource to authors if there is a concern about inappropriate use, attribution of findings or data artifacts.





N3C Publications Committee Review Process

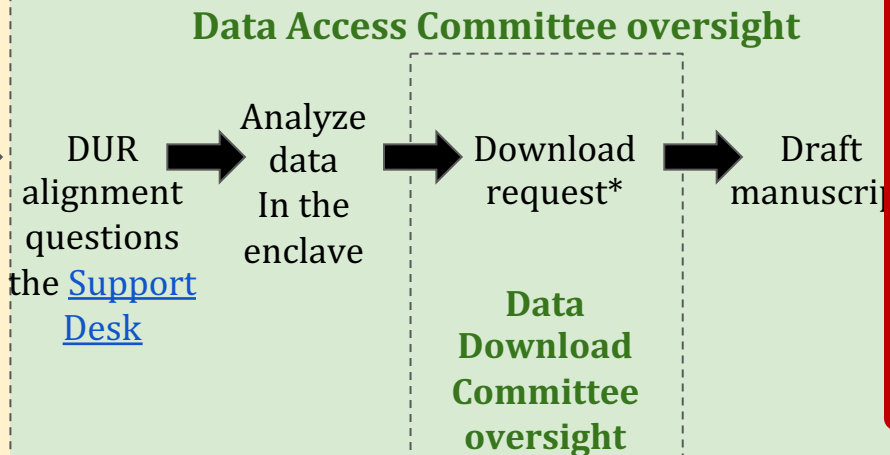


Authors (working within Domain Teams is highly encouraged)
covid.cd2h.org/teams

Assemble
team

Authors should first
consider DUR overlap,
manuscript overlap

covid.cd2h.org/dashboard



Publication Committee

**[N3C
Publication
intent form](#)**

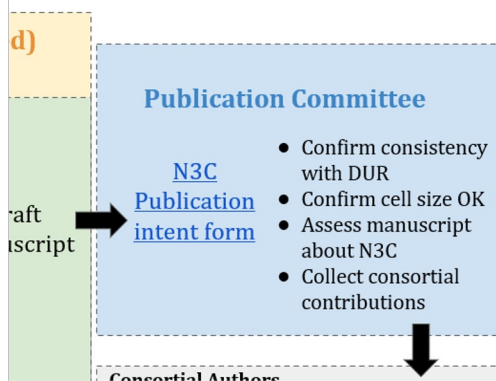
- Confirm consistency with DUR
- Confirm cell size OK
- Assess manuscript about N3C
- Collect consortial contributions

Consortial Authors

Read the paper; assess their contribution; and attest to being listed as a consortial author.



N3C Publications Committee Review Process



- Publication Intent Form workflow emails the N3C committee when there is a new form. Papers are submitted for review throughout the week.
- A primary and secondary reviewers are designated for each week. These reviewers + committee are notified to start reviewing and recording comments before the committee meeting.
- Reviewers review the paper using the review rubric and look for N3C policy violations, issues, etc., and record comments in the review rubric table.



N3C Publications Committee Review Process



Review Rubric

- This table is set up for all new papers. See the next slide for purposes of each column.
- It is included in the meeting [agenda](#) for each new paper for review

MSID:1235.678 Corresponding Author Name Manuscript Title										
Reviewers	About N3C?	N3C is consortial	DUR OK	Cell size OK	Priority (1=highest, 4 = lowest)	Overlap	Native pop OK	Acknowledgements OK	RECOVER-funded ?	Reviewer comments
Carolyn Bramante										
Amit Saha										
Mary Emmett										
Jeremy Harper										
Wenndy Hernandez										
Farrukh Koraishy										
Swaroop Vedula										
Jerrold Anzalone										

Summarized comments, decision recorded here and sent to the authors



N3C Publications Committee Review Process



The review rubric/table columns purposes: (some of these columns have a description in the submission form)

	What is this asking	Desired results
About N3C?	Is this manuscript about the N3C infrastructure, processes, governance, etc. Meaning it is not original research made possible BY the enclave / consortium. It is ABOUT the enclave / consortium itself	Depends, but will be no most of the time. Most papers are original research made possible by the N3C and will not be about N3C
N3C is consortia	Is "N3C Consortium" listed in the author byline?	Should be yes all the time. Any paper using N3C data would need to list "N3C Consortium" in the byline. Papers that also use data artifacts (most papers) should also invite those artifact builders to be included in the "N3C consortium" (indexed on pubmed)-- this might change
DUR OK	Does the submitted work fit under the purview of the DUR	Must be yes all the time. Before submission to a journal, all work should have download approval under the appropriate DUR
Cell size <u>OK</u>	Are the cell sizes all ≥ 20	Must be yes all the time. All numbers of patients < 20 are obscured to " < 20 " If the categories in a table section are mutually exclusive and there is only one " < 20 " cell, that constitutes a back-calculable result, thus additional steps are required.



N3C Publications Committee Review Process



The review rubric/table columns purposes: (some of these columns have a description in the submission form) - continued

Priority (1=highest, 4 = lowest)	Marker of quality; could lead to invitation to present at n3c forum or in very special cases might warrant additional centralized resources.	Varies
Overlap	Overlap of main results relative to prior N3C papers	Should be no, ideally all the time
Native pop OK	Is the representation of native populations concordant with policy.	Must be yes all the time.
Acknowledgements OK	Are the acknowledgements present as stated in covid.cd2h.org/acknowledgements; note that abbreviated version ok for slides and posters	Must be yes all the time.
About RECOVER?	Was the work funded by the RECOVER contract? It in certain cases could be about long covid and NOT funded by RECOVER, however if RECOVER-funded authors are on it, it would look super bad not to credit.	Varies. If it is a RECOVER funded project we should refer them to the RECOVER intent process and PIs.



N3C Publications Committee Review Process



- **What the committee is looking for and common issues seen:**
 - **Cell Size:**
 - numbers of patients that are below the required threshold of 20
 - numbers of patients that from context can be derived as below the threshold of 20
 - With a prior approval from the Download Committee to use a cell size of less than 20 patients, the Publications Committee requests that authors include the following text in manuscript footnotes.
 - “The N3C Download committee reviewed and granted an appeal for this manuscript to report cell patient numbers as low as 10 (instead of the normal threshold of 20)”



N3C Publications Committee Review Process



	ALL Manuscript drafts <ul style="list-style-type: none">- about N3C- including any N3C data- for submission to journals, pre-prints- Form required BEFORE submission.	ALL Conference Submissions (podium presentations and posters) <ul style="list-style-type: none">- about N3C- including any N3C data.- Form required BEFORE submission.
Authors	<ul style="list-style-type: none">● Concordance with DUR● Adherence to download policy● Consortial author attribution● Acknowledgments● Overlap checks	
Committee	<ul style="list-style-type: none">● Confirm concordance with DUR● Check approval of data download● Review cell sizes● Verify adherence to Attribution and Publication Principles● Facilitate attribution (and for manuscripts, consortial authorship)● Promote and track published manuscripts	



N3C Publications Committee Review Process



- **What the committee is looking for and common issues seen:**

- **Cell Size Issue** – these options are offered to the author help remedy the small cell size issue

- *Small cell size and re-computation from marginals**

To protect person privacy N3C policy asserts that cell sizes less than 20 should be suppressed, unless it is actually 0. The question arises that cell sizes could be computed from the marginal totals in cases where there is only one small cell in a row or column. Correcting this circumstance can distort the underlying data, and impact the intended science. This is even more true for large, complex tables. To accommodate both person privacy and team science, the N3C download policy states:

- Percentages in the text that enable the calculation of a cell size <20 are disallowed, and should be replaced with a statement that the proportion is too small to quantitatively report. Below are some approaches that could be taken.
- For single cell sizes with five or more rows or columns:
 - Permits a single cell size <20 when there are five or more elements in a row or column. This makes rapid mental calculation more challenging, though it is still possible. Displaying the cell content as <20 in that circumstance constitutes due diligence.
- For single cell sizes with four or fewer rows or columns, the following options are possible:
 - Delete the marginal total for the row or column having four or less elements and retain the small cell size indicator
 - Collapse the row or column with another category to increase cell size
 - Revise the cut points for categories to increase cell size
 - Skewing the counts of the Total number by +-5 and the next closest sized column along with the <20 cell
 - (at least 2 cells should always be skewed)
- In all circumstances, an appeal process is available



Handling small cell counts

Original (fictional) data - Unmasked

	Total	5000	1500	3500
Age: Scalar data not subject to masking	Age, Years, median	46	45	48
	Age, Years, IQR	31-60	30-60	35-59
Gender: Discrete, non-overlapping categories (<=4 categories). If <20 corresponds to missing or unknown, no secondary masking needed	Gender			
	Female	2050	600	1435
	Male	2942	898	2059
	Missing or Unknown	8	2	6
Race/Ethnicity: Discrete, non-overlapping categories but 5 or more categories	Race / Ethnicity			
	black hispanic	82	10	72
	black non-hispanic	830	210	620
	white hispanic	270	145	125
	white non-hispanic	2838	750	2088
	AI/AN	14	6	8
	others	966	379	587
Medical Comorbidity: Possibly overlapping categories, ergo strict back calculation by column is not possible across rows. However, where multiple columns can be added, additional masking in the row is needed.	Medical comorbidity			
	Diabetes Mellitus	1000	297	703
	Heart failure	460	138	322
	Dementia	6	0	6
	HIV	32	5	27
	Liver disease	620	186	434
	Smoker	840	635	205
Severity: No counts under 20, thus not subject to masking	Severity of Infection			
	Hospitalization	1550	459	1680
	Invasive Vent	106	78	28

Insufficient masking

	Total	5000	1500	3500
Age: Scalar data not subject to masking	Age, Years, median	46	45	48
	Age, Years, IQR	31-60	30-60	35-59
Gender: Discrete, non-overlapping categories (<=4 categories). If <20 corresponds to missing or unknown, no secondary masking needed	Gender			
	Female	2050	600	1435
	Male	2942	898	2059
	Missing or Unknown	<20	<20	<20
Race/Ethnicity: Discrete, non-overlapping categories but 5 or more categories	Race / Ethnicity			
	black hispanic	82	<20	72
	black non-hispanic	830	210	620
	white hispanic	270	145	125
	white non-hispanic	2838	750	2088
	AI/AN	<20	<20	<20
	others	966	379	587
Medical Comorbidity: Possibly overlapping categories, ergo strict back calculation by column is not possible across rows. However, where multiple columns can be added, additional masking in the row is needed.	Medical comorbidity			
	Diabetes Mellitus	1000	297	703
	Heart failure	460	138	322
	Dementia	<20	0	<20
	HIV	32	<20	27
	Liver disease	620	186	434
	Smoker	840	635	205
Severity: No counts under 20, thus not subject to masking	Severity of Infection			
	Hospitalization	1550	459	1680
	Invasive Vent	106	78	28

Sufficient masking

	Total	5000	1500	3500
Age: Scalar data not subject to masking	Age, Years, median	46	45	48
	Age, Years, IQR	31-60	30-60	35-59
Gender: Discrete, non-overlapping categories (<=4 categories). If <20 corresponds to missing or unknown, no secondary masking needed	Gender			
	Female	2050	600	1435
	Male	2942	898	2059
	Missing or Unknown	<20	<20	<20
Race/Ethnicity: Discrete, non-overlapping categories but 5 or more categories	Race / Ethnicity			
	black hispanic	<85	<20	<75
	black non-hispanic	830	210	620
	white hispanic	270	145	125
	white non-hispanic	2838	750	2088
	AI/AN	<20	<20	<20
	others	966	379	587
Medical Comorbidity: Possibly overlapping categories, ergo strict back calculation by column is not possible across rows. However, where multiple columns can be added, additional masking in the row is needed.	Medical comorbidity			
	Diabetes Mellitus	1000	297	703
	Heart failure	460	138	322
	Dementia	<20	0	<20
	HIV	<35	<20	27
	Liver disease	620	186	434
	Smoker	840	635	205
Severity: No counts under 20, thus not subject to masking	Severity of Infection			
	Hospitalization	1550	459	1680
	Invasive Vent	106	78	28

82	=	72	<20
----	---	----	-----

‡ To comply with N3C policy, counts below 20 are displayed as <20 and additional values were skewed by up to 5 in order to render it impossible to back-calculate precise counts in the 'other gender' and HIV categories.



N3C Publications Committee Review Process



- **What the committee is looking for and common issues seen:**
Adherence to Attribution and Publication Principles
 - What **level of data** was used in the paper
 - Confirm the correct **DUR** was used in your intent form submission
 - Include in the **discussion limitations of missing data**
 - Ensure the paper is **about COVID-19** or **COVID-19 focused**



N3C Publications Committee Review Process



Policy:

- In compliance with the NCATS Data Transfer and Data Use Agreements that all N3C Data Enclave users agree to, under no circumstances are tables, figures, parameter estimates, or aggregated statistics to leave the Enclave until download (aka export) approval is obtained by the ResultDownload Committee.
- Prior approval is required for export regardless of the result format (e.g., tables, figures) and regardless of the target venue (manuscripts, posters, presentations, supplementary material, or even your own private hard drive).
- Prior approval is also required regardless of the export mechanism (screenshot, copy/paste, or download).



Result Download (a.k.a Export) Request Process



The following constraints to keep in mind:

- Constraints with aggregated data pertaining to small groups of patients (aka small 'cell sizes')
 - No table or figure may report cell results from fewer than 20 persons unless the value is zero. Cells with counts less than 20 will be identified with a symbol (often seen as "<20" or *).
 - **Resource to authors:** Appeal to reduce the reportable cell size threshold to 10 based on scientific interpretation. If appeal is granted authors must identify affected values.
- Constraints related to Native Populations
 - All zip codes of regions representing rural populations of 20,000 persons or less are fully obscured to '00000'
 - All zip codes of regions of more than 20,000 persons that overlap with Tribal communities or where the majority of residents identify as AI/AN are identified by the first 3 digits.
- Constraints related to Data Partner IDs
 - For any exports that include Data Partner IDs, these IDs must be masked before being published and a statement that these have been masked must also be included. You may consider assigning random labels to data partners, such as four-digit codes which do not occur in the partner IDs.



- **What the committee is looking for and common issues seen:**

Attribution

- [Acknowledgment statement](#) for all publications and preprints using relevant grant numbers and citing the [N3C marker paper](#)
- Statement about the research not being from NCATS or N3C
- Citing the Institutional Development Award (IDeA) Program grant
 - This is for ALL authors that reside in an IDeA state, regardless of their institutional affiliation

The IDeA States are composed of the Commonwealth of Puerto Rico and the following 23 states: Alaska, Arkansas, Delaware, Hawaii, Idaho, Kansas, Kentucky, Louisiana, Maine, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Oklahoma, Rhode Island, South Carolina, South Dakota, Vermont, West Virginia, Wyoming.

Attribution tiers	National COVID Cohort Collaborators		
	Tier 1	Tier 2	Tier 3
	Masthead Authorship	Consortial Collaborators	In Text Acknowledgements
Indexed in Medline	Yes	Yes	No
Nature of contribution	Substantial intellectually important contributions made directly to the manuscript	Less substantial but still intellectually important contributions made directly to the study/manuscript or to knowledge artifacts used directly by the study** (e.g. phenotype definition, knowledge store products, variable codeset)	Important contributions directly to the N3C infrastructure. Note that in text acknowledgement may occasionally overlap with masthead and / or consortial authors.
Who is invited	Exclusively at the discretion of contact & senior authors	by default: Core Authorship blocks as relevant: codeset and other artifacts creators/reviewers and selected domain teams ***	Core Authorship blocks
Publication Committee responsibilities	*	Pub committee notifies core contributors, Codeset or other core artifacts creators/reviewers & relevant domain teams of manuscripts. Pub committee alerts contact author of any responses to consortial-level contributions	Pub committee maintains the core contributors table and the default pastable acknowledgements text. Pub committee notifies core contributors of manuscripts. Pub committee alerts contact author of any opt-outs.
Contributor responsibilities	*	Contributors respond within 3 business days via contributor's form .	Contributors do not need to take action to be acknowledged; they may optionally opt-out within 3 business days via contributor's form .
Contact Author responsibilities	*	Contact author is provided with a spreadsheet of those that have responded to the contributions form. The spreadsheet will include Names, Emails, ORCIDs, and Contributions. Please upload this file with your submission. Do NOT publish email addresses.	If anyone opts out, contact author will be provided with the names of individuals to omit from acknowledgement, else use the default acknowledgements text. Contact author will paste the acknowledgements text in the appropriate section in the body of their manuscript.



N3C Publications Committee Review Process



- **What the committee is looking for and common issues seen:**

- Attribution

- **Typographical issues** - the manuscript had spelling and grammatical errors, proofread before submitting for peer review
- N3C Publications Committee is here to serve the authors as a resource



N3C Publications Committee Review Process



- **During the weekly meeting (30 min meeting on Thursdays), the committee deliberates and decides:**
 - **Approve the reviewed item** and communicate the decision to the author - work is good and no violations.
 - On behalf of the author, committee invites the core N3C contributors to also review and declare contributions
 - Committee PM collects consortial contributions and shares with author to include in paper. See the slide with contribution tiers.
 - **Request changes** – if committee determines that there are issues with the paper that need to be addressed (see next slide for issues), committee lists what changes need to be made and asks the author to update and resend for review.
 - Review process restarts when changes are made and updated manuscripts is resubmitted.



Publication Intent Form



Available via N3C Enclave: New Publications Committee Members will be added to the intent form dashboard, communication to follow.

N3C Publication Intent Form

Home

Search... Ctrl + J

Notifications

Recent

Files

Applications

PLATFORM APPS

Workshop

PROMOTED APPS

Publication Intent Form

Publication Utilities

Guide for Authors

Form Homepage

The N3C Publication Intent Form is designed to facilitate compliance with N3C policies on data use (covid.cd2h.org/dur), attribution, and publication principles (doi.org/10.5281/zenodo.7126728) and inform the N3C Community about works in progress and published.

All works (manuscripts, conference manuscripts, abstracts, posters, presentations) that make use of N3C community resources must be reviewed and approved by the Publication Committee prior to their being sent onward to peer review or published in any form (such as preprint). Before you start, please consult the guidelines on the N3C website <https://covid.cd2h.org/publication-review> to be aware of the process.

This form cannot be edited once submitted; a separate form is available for revisions to the work. Please use <https://covid.cd2h.org/support> for any questions or concerns.

Thank you from the Publications Committee, and congratulations on the progress of your study!

Please click on 'Create New Publication Intent' or 'Continue Working on a Draft Submission' below.

Already submitted to the Publication Committee and have a manuscript ID (MSID)? Then instead go to the [N3C Publication Intent Revisions Form](#).

Create New Publication Intent

1

Manuscript overview

2

Add your declarations

3

N3C acknowledgements and consortial authorship

4

Manuscript links and metadata

5

Presented work

6

Target journal details

7

Comments

8

Review and Submit



National
COVID
Cohort
Collaborative

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

Important Resources



NATIONAL CENTER
FOR DATA TO HEALTH

- N3C Publication Presentations at N3Community Forum:
 - [Overview of Publications Process](#)
 - [Guide2N3C: Publications Committee Chapter](#)
 - [Publishing & Sharing Your Work](#) chapter in *[The Researcher's Guide to N3C](#)*



Example



- We'll go through an end-to-end example using our malnutrition study
- To date, we've completed our analysis and prepared it for download review. Our next steps include:
 - Submit a download review request (DRR)
 - Prepare our manuscript
 - Pay close attention to the attribution and acknowledgement policies here:
<https://covid.cd2h.org/acknowledgements/>
 - Submit a publication intent form
 - Required for abstracts, presentations, and manuscripts
 - Revise if necessary
 - Submit!