NCTN/NCORP Data Archive

Data from NCI-funded clinical trials

June 17, 2021

Patrick Schnell
College of Public Health

THE OHIO STATE UNIVERSITY
CENTER FOR CLINICAL AND TRANSLATIONAL SCIENCE

Today's goals

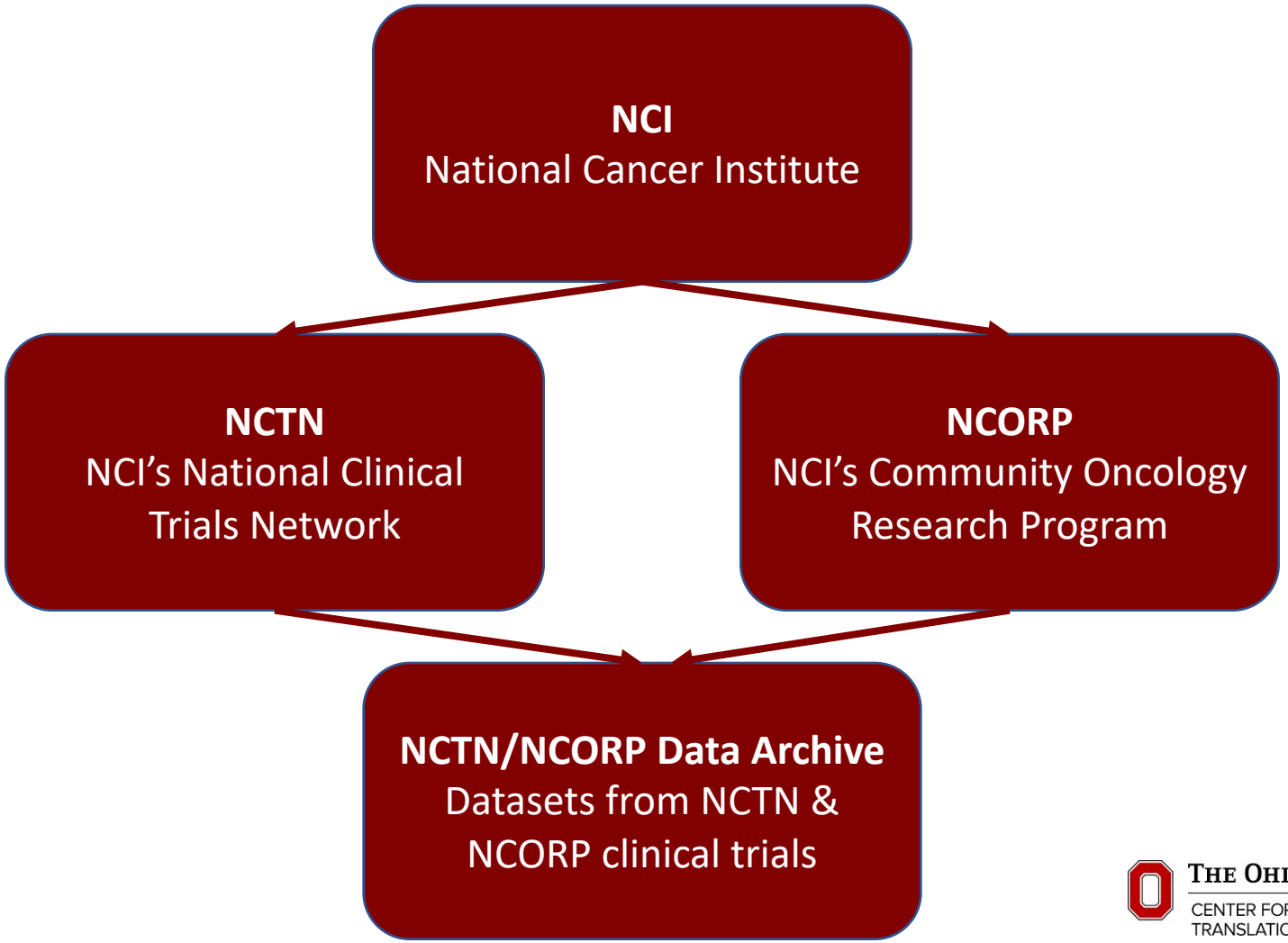
1. Become familiar with the NCI's **NCTN/NCORP Data Archive** datasets and how they can be used for your research.

2. Learn about the process for accessing the data and publishing manuscripts.

Overview

1. What are NCTN and NCORP?
2. Data model
3. Searching and browsing datasets
4. Requesting datasets

Data Archive within NCI



NCTN/NCORP Data Archive

- https://nctn-data-archive.nci.nih.gov
- Central, controlled-access database
- Datasets generated from clinical trials of
  - National Clinical Trials Network (NCTN)
  - NCI Community Oncology Research Program (NCORP)
- Generally includes clinical data from
  - Primary publications of phase 3 trials published as of January 1, 2015
  - Non-primary publications of phase 3 trials published as of April 1, 2018

NCTN

- Over 2,200 sites in US, Canada, & elsewhere
- Provides infrastructure for NCI-funded treatment & primary advanced imaging trials
- Involved in
  - Setting standards of care
  - FDA approval
  - Testing new treatments
  - Validating new biomarkers

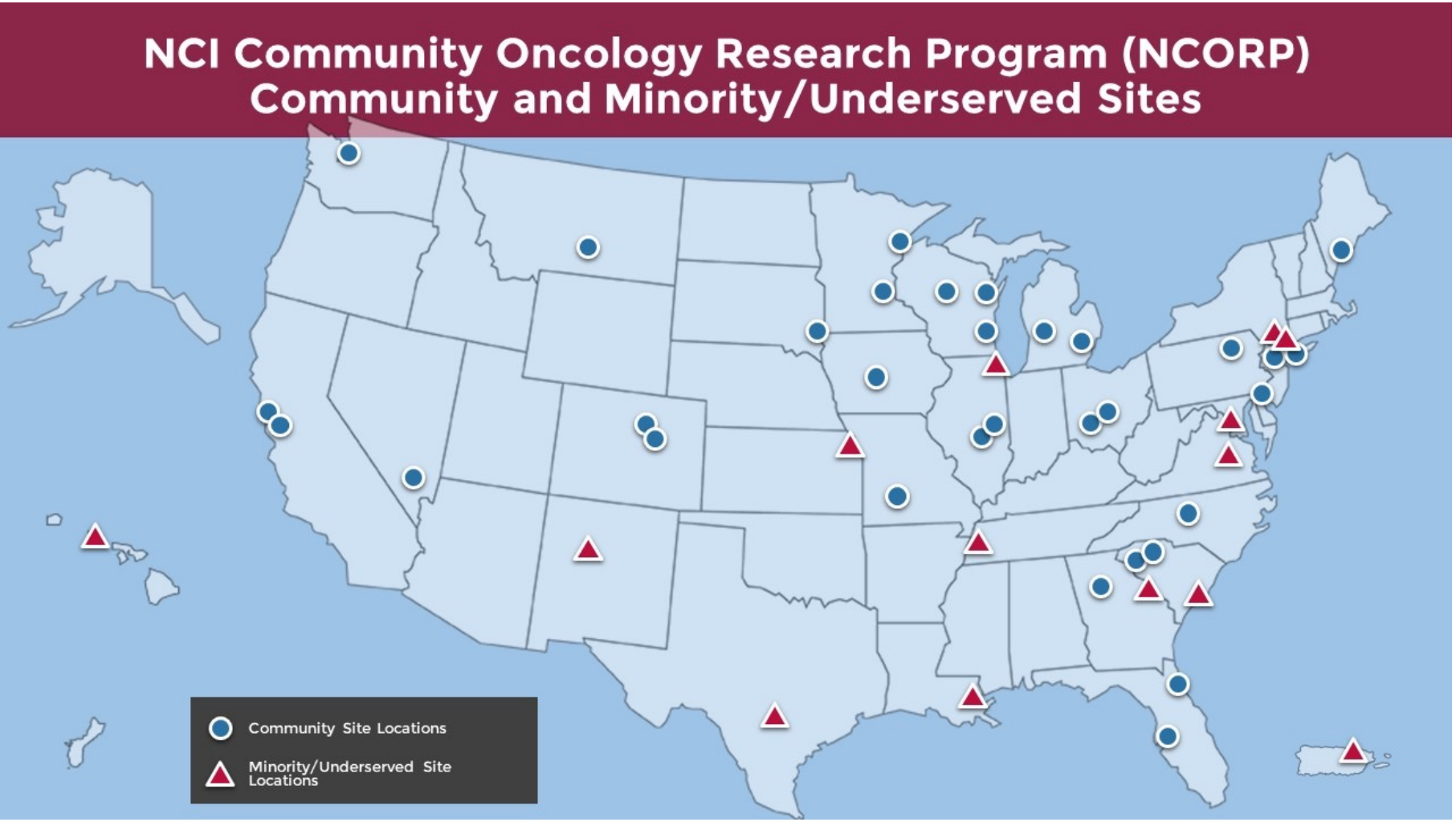
NCTN composition

- Alliance for Clinical Trials in Oncology
- ECOG-ACRIN Cancer Research Group
- NRG Oncology
- SWOG
- Children's Oncology Group (COG)
- Canadian Cancer Trials Group (CCTG)

NCORP

- NCI's Community Oncology Research Program
- Designs & conducts clinical trials in:
  - Prevention
  - Screening
  - Supportive care & symptom management
  - Surveillance
  - Health-related quality of life
  - Cancer care delivery

NCORP sites



Research bases:

- Alliance
- Children's (COG)
- ECOG
- NRG Oncology
- SWOG
- University of Rochester
- Wake Forest

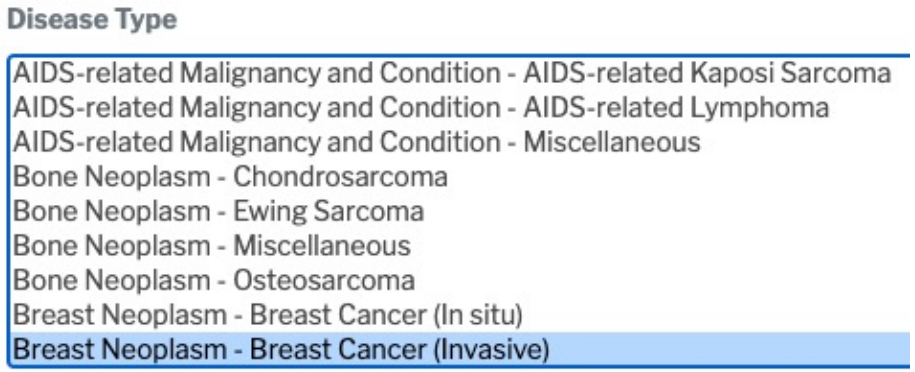
Hubs for design & implementation of multi-center trials

What is included with a dataset?

- For each trial, Data Archive entry includes:
  - Dataset
  - Data dictionary
  - Limited metadata fields
- Datasets:
  - Are patient-level
  - Are de-identified (so not human subjects research; can get IRB exemption)
  - Have values for all variables used in published analyses
- Some trials have imaging data available via The Cancer Image Archive

Searching the Data Archive

- If you already have a specific trial in mind, can find by:
  - **NCT Trial Number**, e.g., NCT01598298 (most reliable)
  - **Study ID**, e.g., S1202
  - **Trial Title**, e.g., A Randomized Placebo-Controlled Phase III Study of Duloxetine for Treatment of Aromatase Inhibitor-Associated Musculoskeletal Symptoms in Women With Early Stage Breast Cancer
  - **PubMed ID**, e.g., 29136387
- Search by disease type



Viewing data schema

Datasets Linked to Trial

Title	Description
NCT01598298-D3	This dataset will allow users to reproduce the results reported in the Journal of Clinical Oncology (Henry et al., JCO, 2017), which contains patient-level data (one row per patient), including eligibility, evaluability, treatment, demographic, and baseline information.
NCT01598298-D2	This dataset will allow users to reproduce the results reported in the Journal of Clinical Oncology (Henry et al., JCO, 2017), which contains event-level information one row per patient per adverse event experienced on trial).
NCT01598298-D1	This dataset will allow users to reproduce the results reported in the Journal of Clinical Oncology (Henry et al., JCO, 2017), which contains assessment-level data (one row per patient per planned assessment time), including scores for each of the patient-completed questionnaires, days from registration to time of assessment, and whether the questionnaires were completed within the protocol-mandated time frame.

Example schema (patients)

Data Dictionary for SWOG S1202 Patient Level (PATIENT) Dataset

Column	Variable Name	Type	Variable Description	Code	Code Definition
1	ALTPATID	Char	Patient identifier		Text
2	ELIG	Char	Patient was eligible	Y	Yes
				N	No
3	ELIGSP	Char	Reason patient was not eligible	AI	AI started > 36 months prior
				BASEPAIN	Baseline average pain score < 4
				CRCL	Baseline CrCl level too low
				PREMEN	Not postmenopausal
				VENLA	Prior venlafaxine for pain
				WRCANC	Noninvasive breast cancer
				NA	NA – patient was eligible
4	EVAL	Char	Patient was evaluable for primary endpoint	Y	Yes
				N	No
5	EVALSP	Char	Reason patient was not evaluable for primary endpoint	INEL	Patient was not eligible
				NOIW	No in-window BPI at week 2, 6, or 12
				NA	NA – patient was evaluable

Example schema (outcomes)

Data Dictionary for SWOG S1202 Outcomes Dataset

Column	Variable Name	Type	Variable Description	Code	Code Definition
1	ALTPATID	Char	Patient identifier		Text
2	S1202TIM	Num	Data collection time point	1	Pre-registration
				15	Week 2
				43	Week 6
				85	Week 12
				168	Week 24
3	BPI_WEEKS	Num	Number of weeks after registration that BPI-SF questionnaire was completed		Number
				‘ ’	Missing data
4	BPI_WINDOW	Char	Was the BPI-SF completed within the protocol-specified window	Y	Yes
				N	No
				‘ ’	Missing data
5	BPI_WINDOW14	Char	Was the BPI-SF completed within 14 weeks after registration	Y	Yes
				N	No
				‘ ’	Missing data
6	AVERAGE	Num	BPI-SF score for average pain		Number
				‘ ’	Missing data
7	WORST	Num	BPI-SF score for worst pain		Number
				‘ ’	Missing data
8	INT_SCORE	Num	BPI-SF score for pain interference		Number
				‘ ’	Missing data

Example schema (adverse events)

Data Dictionary for SWOG S1202 Adverse Events (AES) Dataset

Column	Variable Name	Type	Variable Description	Code	Code Definition
1	ALTPATID*	Char	Patient identifier		Text
2	TOXCODE	Char	CTCAE 4.0 code		Text
3	TOXLABEL	Char	CTCAE term		Text
4	GRADE	Char	CTCAE grade	Grade 1	Grade 1
				Grade 2	Grade 2
				Grade 3	Grade 3
*Patients who did not experience any adverse events on study will not appear in the AES dataset					

Requesting data

- Create an account with institutional email (osu.edu or osumc.edu)

- Request form with:
  - List of trials you are requesting data from
  - Brief research plan
    - Mission / legal / administrative review only; **no scientific review**
  - Data Use Agreement (DUA)

- Took me a month to get DUA signatures from OSU side; completed request took less than a week for NCI to process.

Research plan example (one paragraph)

Our overall goal is to determine how adverse events (AEs) relate to patient perceptions of the effect of duloxetine for treating aromatase inhibitor (AI)–associated musculoskeletal symptoms (AIMSS). **We will identify the effect of duloxetine on patient-perceived benefit in patient subgroups defined by post-randomization AEs and reductions in pain using principal stratification**, as well as the relative sizes of those subgroups. Subgroups of special interest are patients for whom duloxetine caused: reduction in average pain but no AEs of any grade; one or more AEs but not a reduction in pain; and both a reduction in pain and one or more AEs.

Data use agreement

- Entering investigator info, research plan into Data Archive request web form automatically generates unsigned DUA
- Need signatures from
  - Institution representative
    - OSU: Technology Commercialization Office
    - OSUMC: ???
    - **Need individual email address: name.123@osu.edu, not contracts@osu.edu**
  - Investigator
- Must list all people who will have access to data; additions must be made through DUA amendment approved by NCI

Stipulations

- Pre-publication review by NCI and any involved NCI collaborators (e.g., original trial sponsor)
- Include acknowledgment of NCI/NCTN Data Archive and specific trials and datasets in any publications
- Send PubMed ID of any publications to Data Archive within 60 days of publication
- Data Use Agreement expires after 3 years; renew or destroy individual data

For more information and BERD support:

- NCTN/NCORP Data Archive Website:
  - https://nctn-data-archive.nci.nih.gov
- To request BERD support on a paper:
  - https://myccts.osu.edu/
  - Alternatively, submit a ticket to the Center for Biostatistics:
    - https://medicine.osu.edu/departments/biostatistics/service-request-form

Thank you!

THE OHIO STATE UNIVERSITY
CENTER FOR CLINICAL AND TRANSLATIONAL SCIENCE

rethink reuse research