Collaboration and Re-Use

Experiences with institutional data catalogs

Nicole Contaxis, MLIS
Lead, NYU Data Catalog

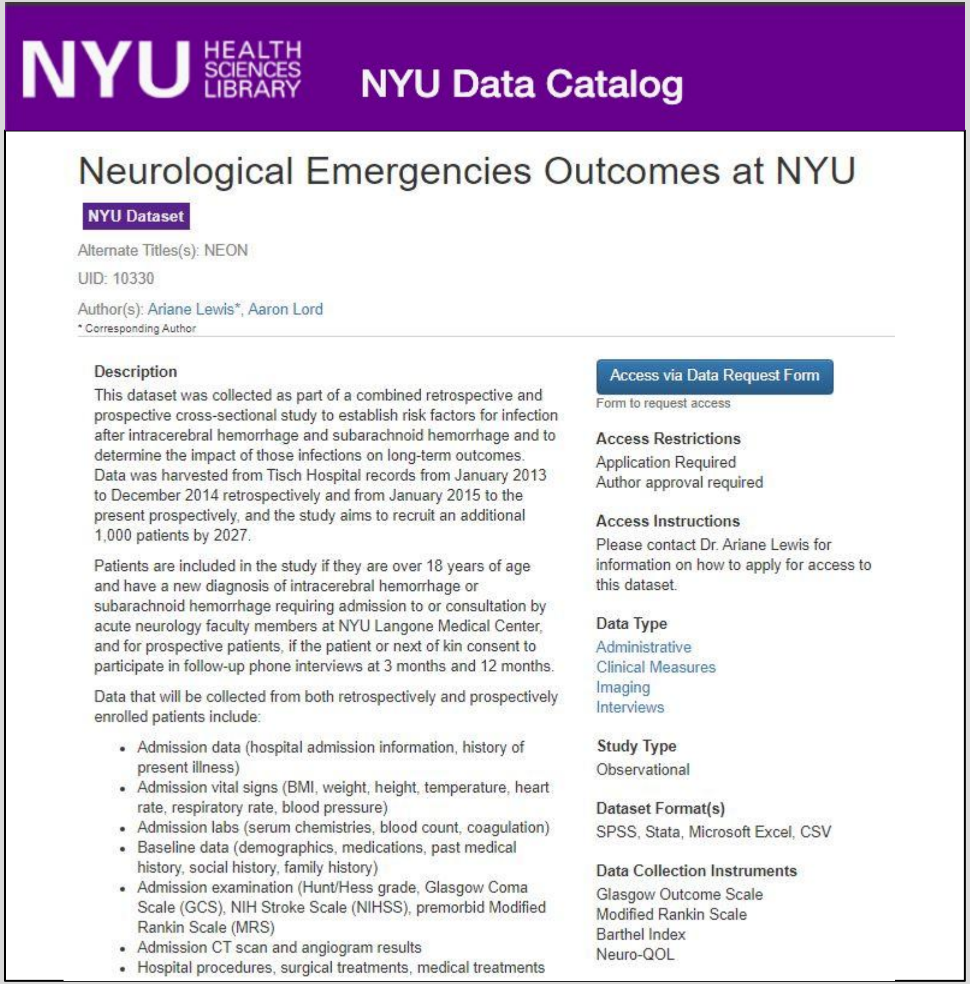
Agenda

1. Background on the NYU Data Catalog
2. Background on the Data Discovery Collaboration Project
3. Stories of Data Re-Use

Background on the NYU Data Catalog

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data



NYU Data Catalog

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Titles(s): NEON

UID: 10330

Author(s): Ariane Lewis*, Aaron Lord

* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin consent to participate in follow-up phone interviews at 3 months and 12 months.

Data that will be collected from both retrospectively and prospectively enrolled patients include:

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart rate, respiratory rate, blood pressure)
- Admission labs (serum chemistries, blood count, coagulation)
- Baseline data (demographics, medications, past medical history, social history, family history)
- Admission examination (Hunt/Hess grade, Glasgow Coma Scale (GCS), NIH Stroke Scale (NIHSS), premorbid Modified Rankin Scale (MRS)
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

Form to request access

Access Restrictions

Application Required
Author approval required

Access Instructions

Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type

Administrative
Clinical Measures
Imaging
Interviews

Study Type

Observational

Dataset Format(s)

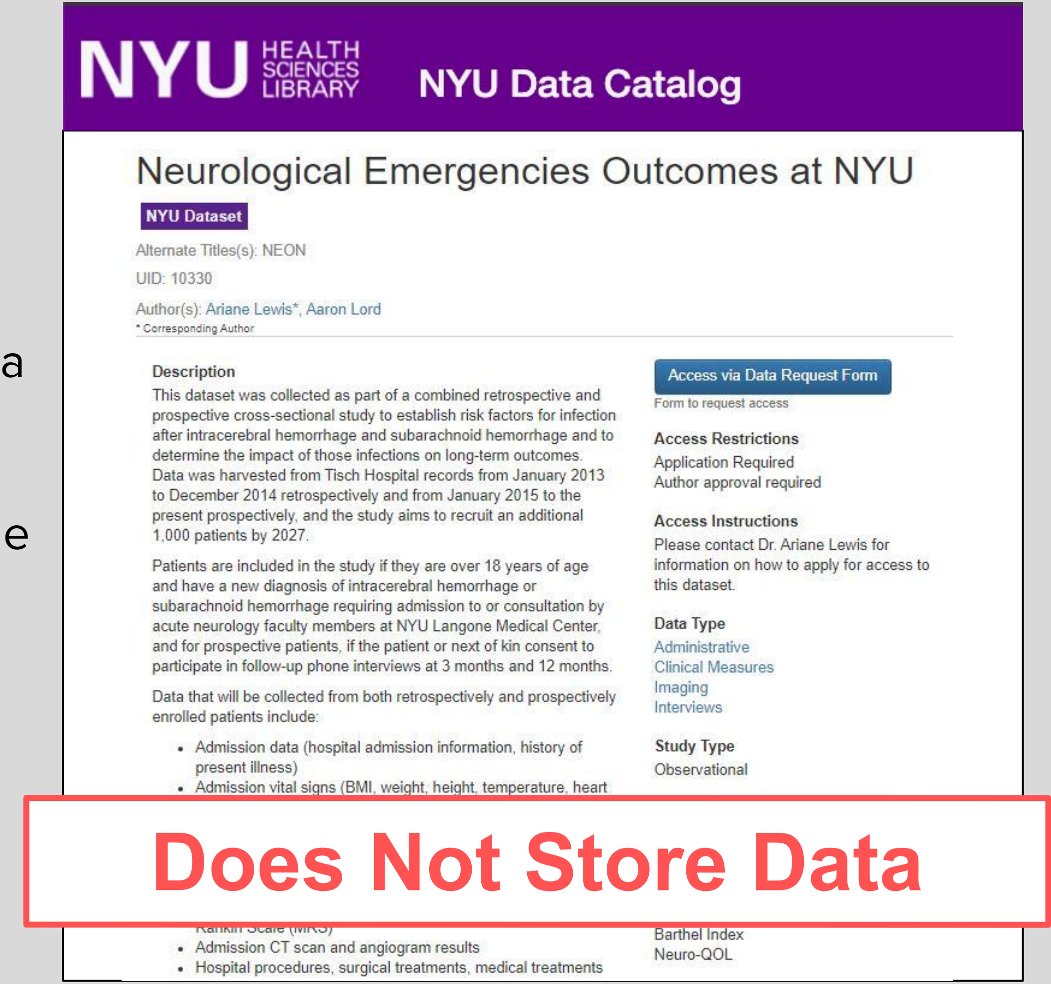
SPSS, Stata, Microsoft Excel, CSV

Data Collection Instruments

Glasgow Outcome Scale
Modified Rankin Scale
Barthel Index
Neuro-QOL

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data
- Makes research data discoverable regardless of where it is stored



NYU HEALTH SCIENCES LIBRARY NYU Data Catalog

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Titles(s): NEON

UID: 10330

Author(s): Ariane Lewis*, Aaron Lord

* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin consent to participate in follow-up phone interviews at 3 months and 12 months.

Data that will be collected from both retrospectively and prospectively enrolled patients include:

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart

Rankin Scale (MRS)

- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

Access via Data Request Form

Form to request access

Access Restrictions

Application Required

Author approval required

Access Instructions

Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type

Administrative

Clinical Measures

Imaging

Interviews

Study Type

Observational

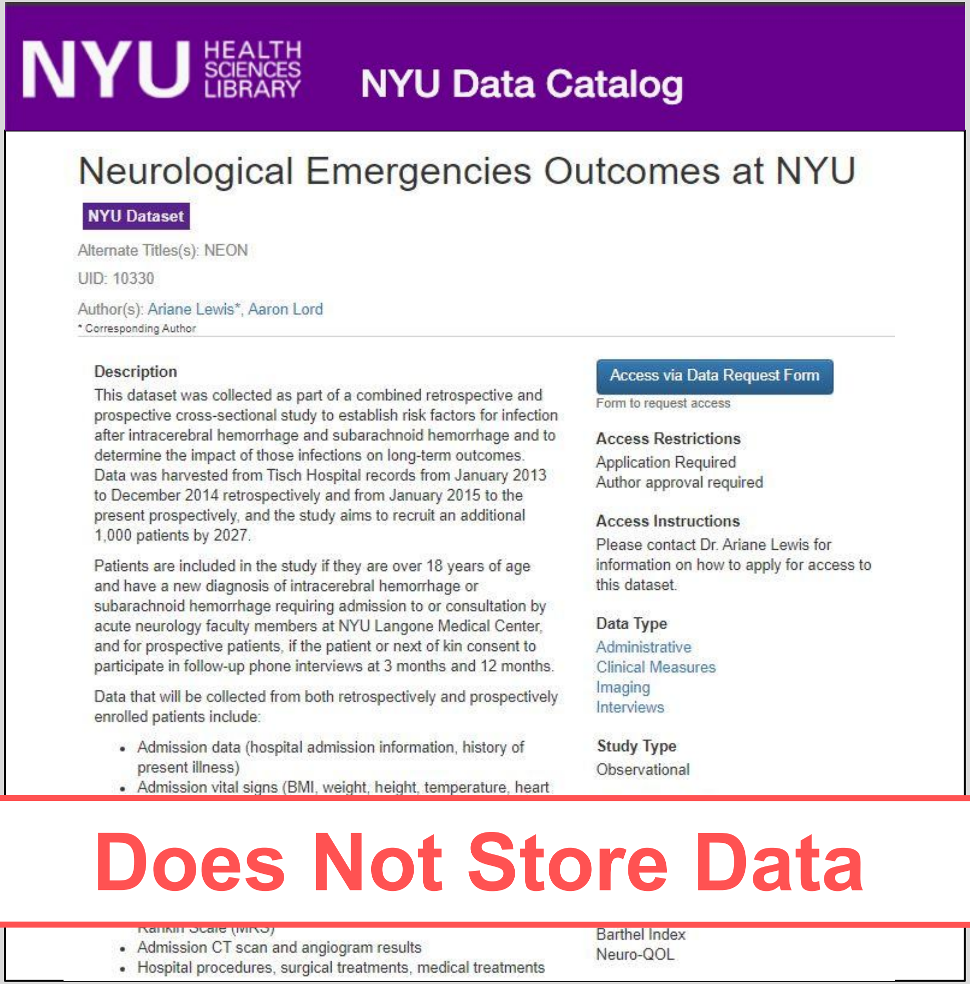
Barthel Index

Neuro-QOL

Does Not Store Data

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data
- Makes research data discoverable regardless of where it is stored
- Open source
  - Code on GitHub
  - Documentation on OSF



NYU HEALTH SCIENCES LIBRARY NYU Data Catalog

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Titles(s): NEON

UID: 10330

Author(s): Ariane Lewis*, Aaron Lord

* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin consent to participate in follow-up phone interviews at 3 months and 12 months.

Data that will be collected from both retrospectively and prospectively enrolled patients include:

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart
- Rankin Scale (MRS)
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

Access via Data Request Form

Form to request access

Access Restrictions

Application Required
Author approval required

Access Instructions

Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type

Administrative
Clinical Measures
Imaging
Interviews

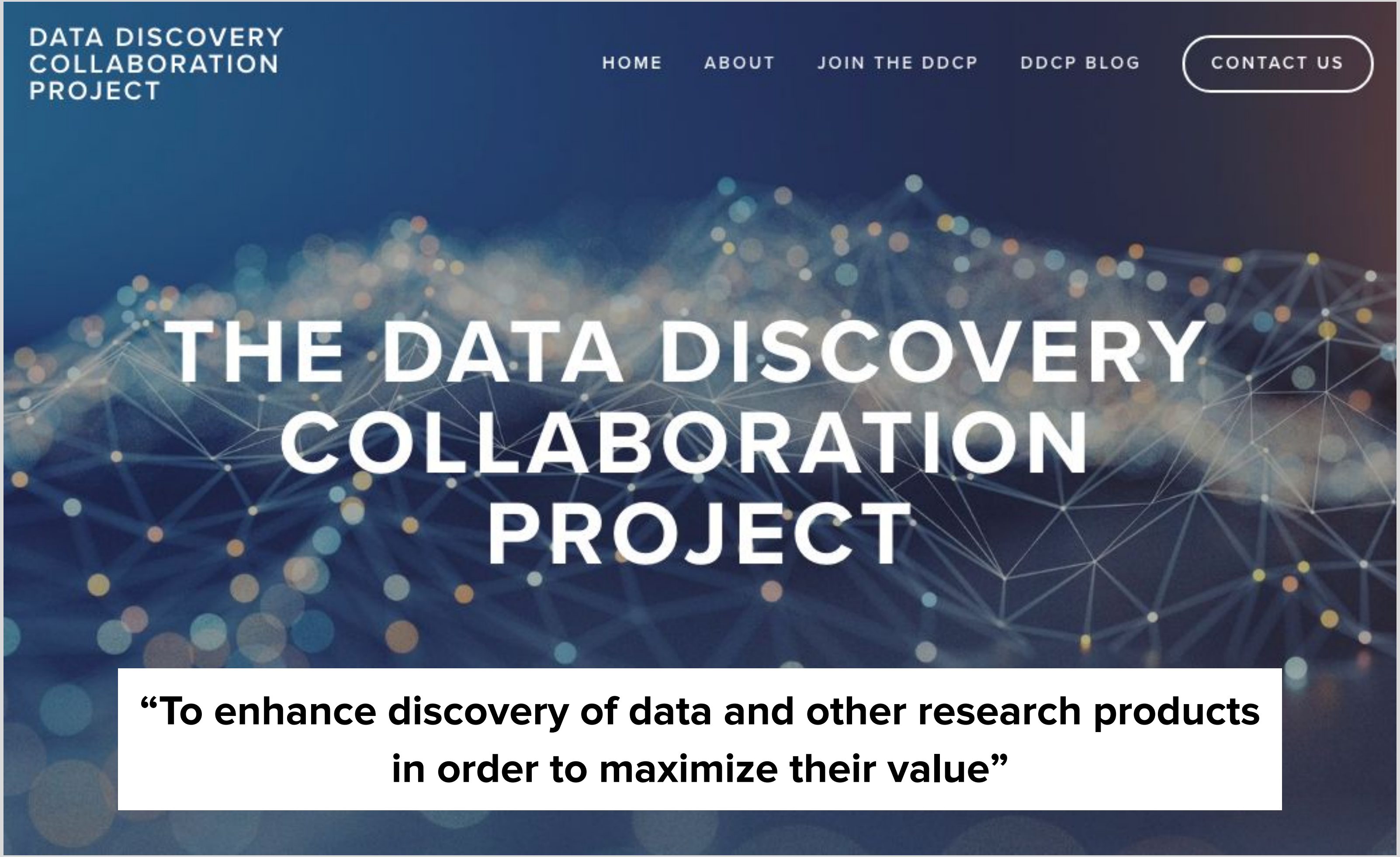
Study Type

Observational

Barthel Index
Neuro-QOL

Does Not Store Data

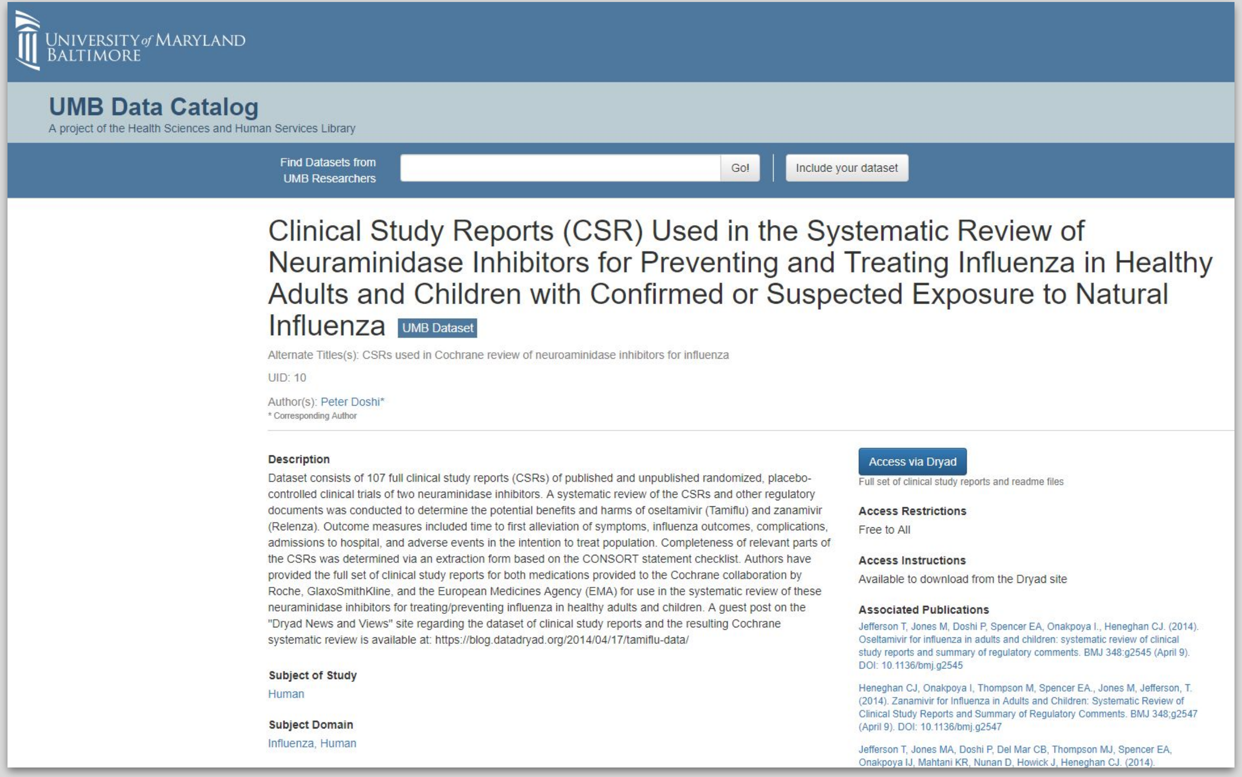
Background on the Data Discovery Collaboration Project



THE DATA DISCOVERY COLLABORATION PROJECT

“To enhance discovery of data and other research products in order to maximize their value”

Background: Example Data Catalog Record



UMB Data Catalog

A project of the Health Sciences and Human Services Library

Clinical Study Reports (CSR) Used in the Systematic Review of Neuraminidase Inhibitors for Preventing and Treating Influenza in Healthy Adults and Children with Confirmed or Suspected Exposure to Natural Influenza UMB Dataset

Alternate Titles(s): CSRs used in Cochrane review of neuroaminidase inhibitors for influenza

UID: 10

Author(s): Peter Doshi*

* Corresponding Author

Description

Dataset consists of 107 full clinical study reports (CSRs) of published and unpublished randomized, placebo-controlled clinical trials of two neuraminidase inhibitors. A systematic review of the CSRs and other regulatory documents was conducted to determine the potential benefits and harms of oseltamivir (Tamiflu) and zanamivir (Relenza). Outcome measures included time to first alleviation of symptoms, influenza outcomes, complications, admissions to hospital, and adverse events in the intention to treat population. Completeness of relevant parts of the CSRs was determined via an extraction form based on the CONSORT statement checklist. Authors have provided the full set of clinical study reports for both medications provided to the Cochrane collaboration by Roche, GlaxoSmithKline, and the European Medicines Agency (EMA) for use in the systematic review of these neuraminidase inhibitors for treating/preventing influenza in healthy adults and children. A guest post on the "Dryad News and Views" site regarding the dataset of clinical study reports and the resulting Cochrane systematic review is available at: https://blog.datadryad.org/2014/04/17/tamiflu-data/

Subject of Study

Human

Subject Domain

Influenza, Human

Full set of clinical study reports and readme files

Access Restrictions

Free to All

Access Instructions

Available to download from the Dryad site

Associated Publications

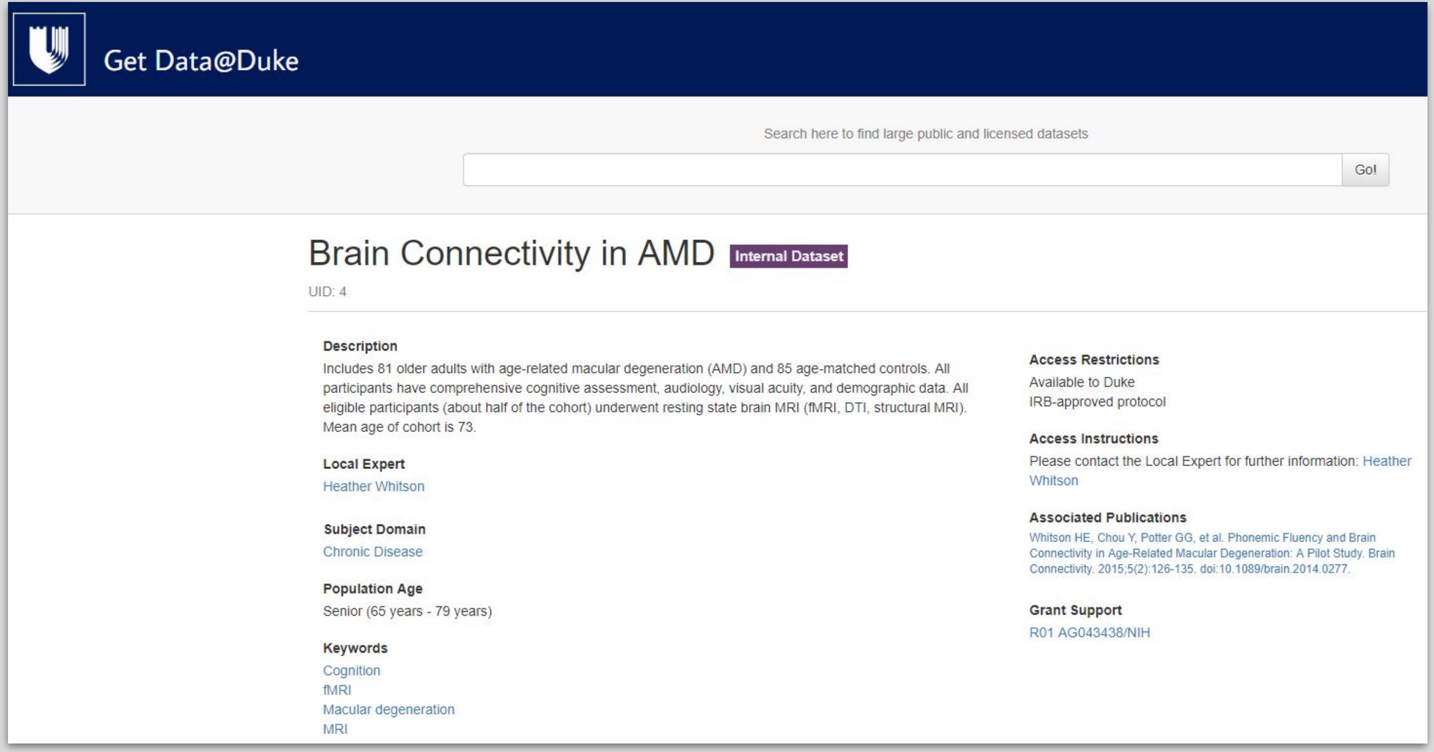
Jefferson T, Jones M, Doshi P, Spencer EA, Onakpoya I., Heneghan CJ. (2014). Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments. BMJ 348:g2545 (April 9). DOI: 10.1136/bmj.g2545

Heneghan CJ, Onakpoya I, Thompson M, Spencer EA., Jones M, Jefferson, T. (2014). Zanamivir for Influenza in Adults and Children: Systematic Review of Clinical Study Reports and Summary of Regulatory Comments. BMJ 348;g2547 (April 9). DOI: 10.1136/bmj.g2547

Jefferson T, Jones MA, Doshi P, Del Mar CB, Thompson MJ, Spencer EA, Onakpoya IJ, Mahtani KR, Nunan D, Howick J, Heneghan CJ. (2014).

External Repository

Background: Example Data Catalog Record



Brain Connectivity in AMD

Internal Dataset

UID: 4

Description

Includes 81 older adults with age-related macular degeneration (AMD) and 85 age-matched controls. All participants have comprehensive cognitive assessment, audiology, visual acuity, and demographic data. All eligible participants (about half of the cohort) underwent resting state brain MRI (fMRI, DTI, structural MRI). Mean age of cohort is 73.

Local Expert

Heather Whitson

Subject Domain

Chronic Disease

Population Age

Senior (65 years - 79 years)

Keywords

Cognition
fMRI
Macular degeneration
MRI

Access Restrictions

Available to Duke
IRB-approved protocol

Access Instructions

Please contact the Local Expert for further information: Heather Whitson

Associated Publications

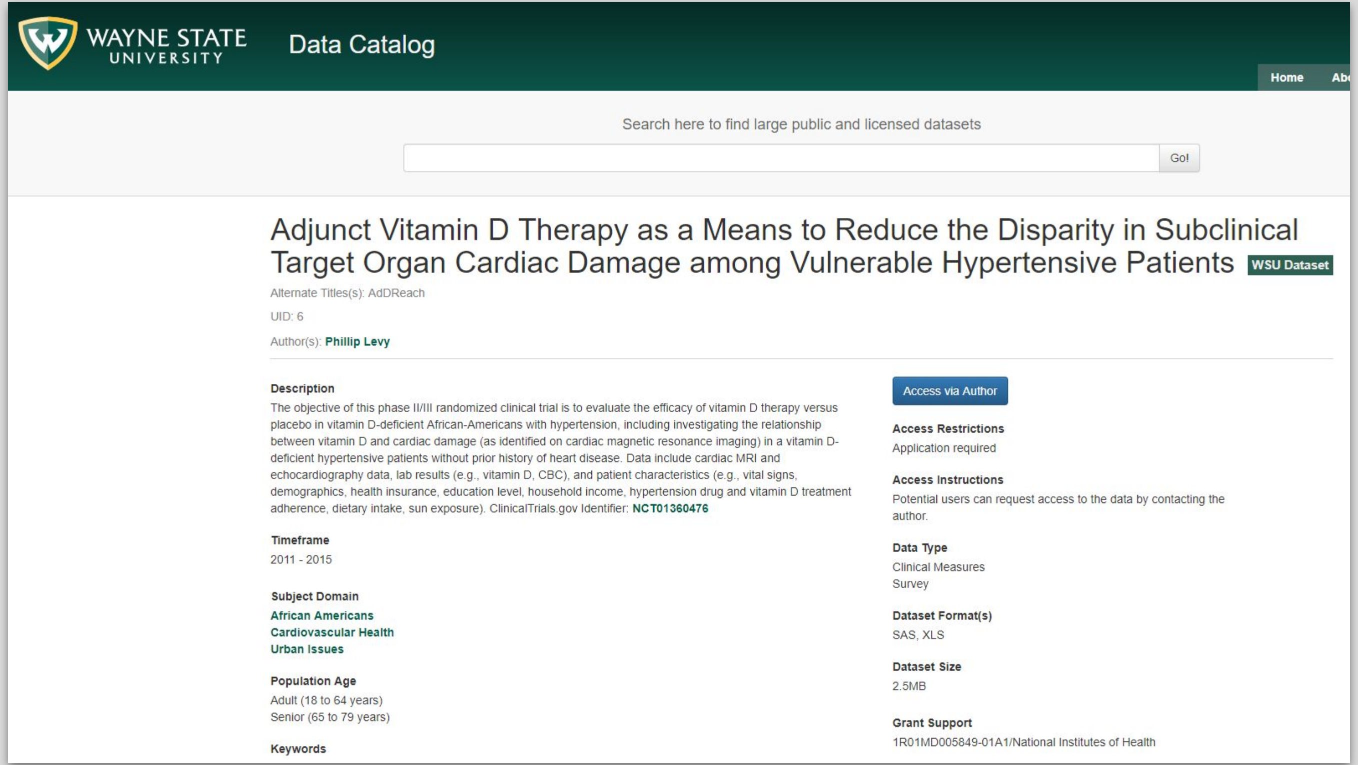
Whitson HE, Chou Y, Potter GG, et al. Phonemic Fluency and Brain Connectivity in Age-Related Macular Degeneration: A Pilot Study. Brain Connectivity. 2015;5(2):126-135. doi:10.1089/brain.2014.0277.

Grant Support

R01 AG043438/NIH

Restricted Access Data

Background: Example Data Catalog Record



Data Catalog

Search here to find large public and licensed datasets

Adjunct Vitamin D Therapy as a Means to Reduce the Disparity in Subclinical Target Organ Cardiac Damage among Vulnerable Hypertensive Patients

Alternate Titles(s): AdDReach

UID: 6

Author(s): Phillip Levy

Description

The objective of this phase II/III randomized clinical trial is to evaluate the efficacy of vitamin D therapy versus placebo in vitamin D-deficient African-Americans with hypertension, including investigating the relationship between vitamin D and cardiac damage (as identified on cardiac magnetic resonance imaging) in a vitamin D-deficient hypertensive patients without prior history of heart disease. Data include cardiac MRI and echocardiography data, lab results (e.g., vitamin D, CBC), and patient characteristics (e.g., vital signs, demographics, health insurance, education level, household income, hypertension drug and vitamin D treatment adherence, dietary intake, sun exposure). ClinicalTrials.gov Identifier: NCT01360476

Timeframe

2011 - 2015

Subject Domain

African Americans
Cardiovascular Health
Urban Issues

Population Age

Adult (18 to 64 years)
Senior (65 to 79 years)

Keywords

Access Restrictions

Application required

Access Instructions

Potential users can request access to the data by contacting the author.

Data Type

Clinical Measures
Survey

Dataset Format(s)

SAS, XLS

Dataset Size

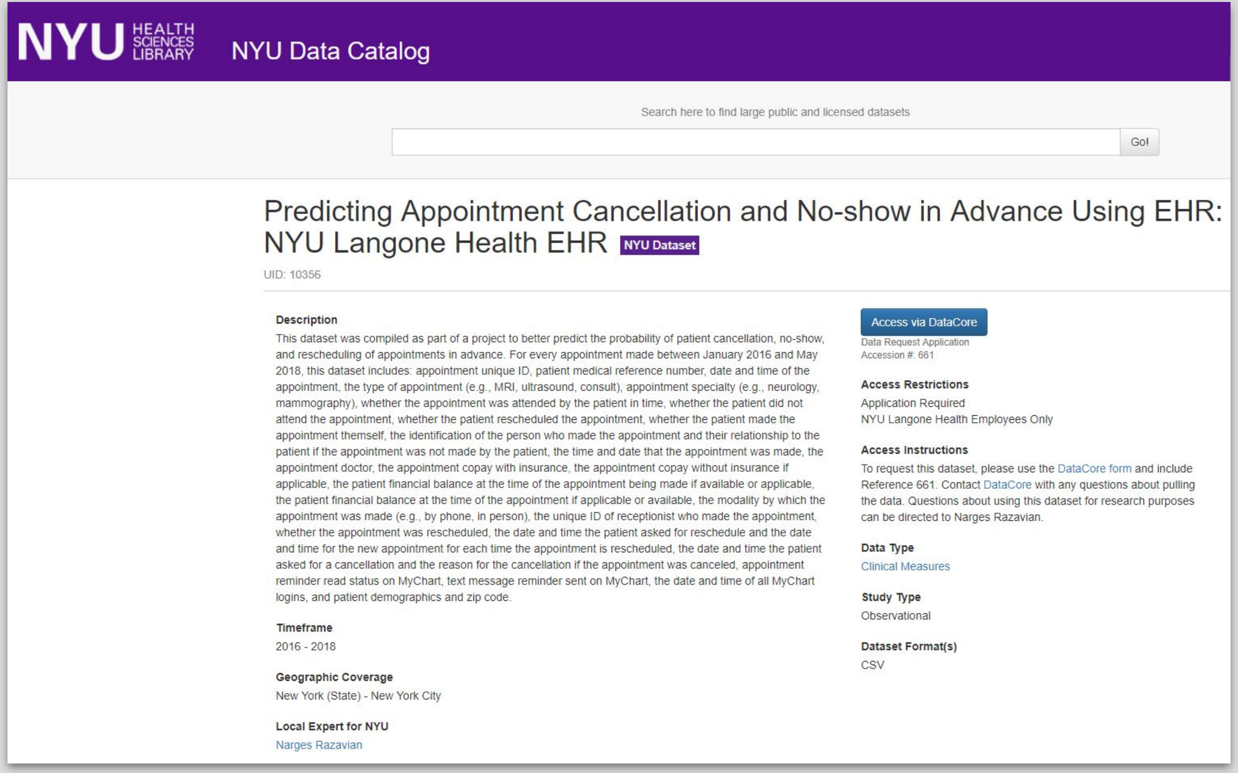
2.5MB

Grant Support

1R01MD005849-01A1/National Institutes of Health

Datasets Via Author Only

Background: Example Data Catalog Record



NYU Data Catalog

Search here to find large public and licensed datasets

Predicting Appointment Cancellation and No-show in Advance Using EHR: NYU Langone Health EHR NYU Dataset

UID: 10356

Description

This dataset was compiled as part of a project to better predict the probability of patient cancellation, no-show, and rescheduling of appointments in advance. For every appointment made between January 2016 and May 2018, this dataset includes: appointment unique ID, patient medical reference number, date and time of the appointment, the type of appointment (e.g., MRI, ultrasound, consult), appointment specialty (e.g., neurology, mammography), whether the appointment was attended by the patient in time, whether the patient did not attend the appointment, whether the patient rescheduled the appointment, whether the patient made the appointment themself, the identification of the person who made the appointment and their relationship to the patient if the appointment was not made by the patient, the time and date that the appointment was made, the appointment doctor, the appointment copay with insurance, the appointment copay without insurance if applicable, the patient financial balance at the time of the appointment being made if available or applicable, the patient financial balance at the time of the appointment if applicable or available, the modality by which the appointment was made (e.g., by phone, in person), the unique ID of receptionist who made the appointment, whether the appointment was rescheduled, the date and time the patient asked for reschedule and the date and time for the new appointment for each time the appointment is rescheduled, the date and time the patient asked for a cancellation and the reason for the cancellation if the appointment was canceled, appointment reminder read status on MyChart, text message reminder sent on MyChart, the date and time of all MyChart logins, and patient demographics and zip code.

Timeframe

2016 - 2018

Geographic Coverage

New York (State) - New York City

Local Expert for NYU

Narges Razavian

Access via DataCore

Data Request Application
Accession #: 661

Access Restrictions

Application Required
NYU Langone Health Employees Only

Access Instructions

To request this dataset, please use the DataCore form and include Reference 661. Contact DataCore with any questions about pulling the data. Questions about using this dataset for research purposes can be directed to Narges Razavian.

Data Type

Clinical Measures

Study Type

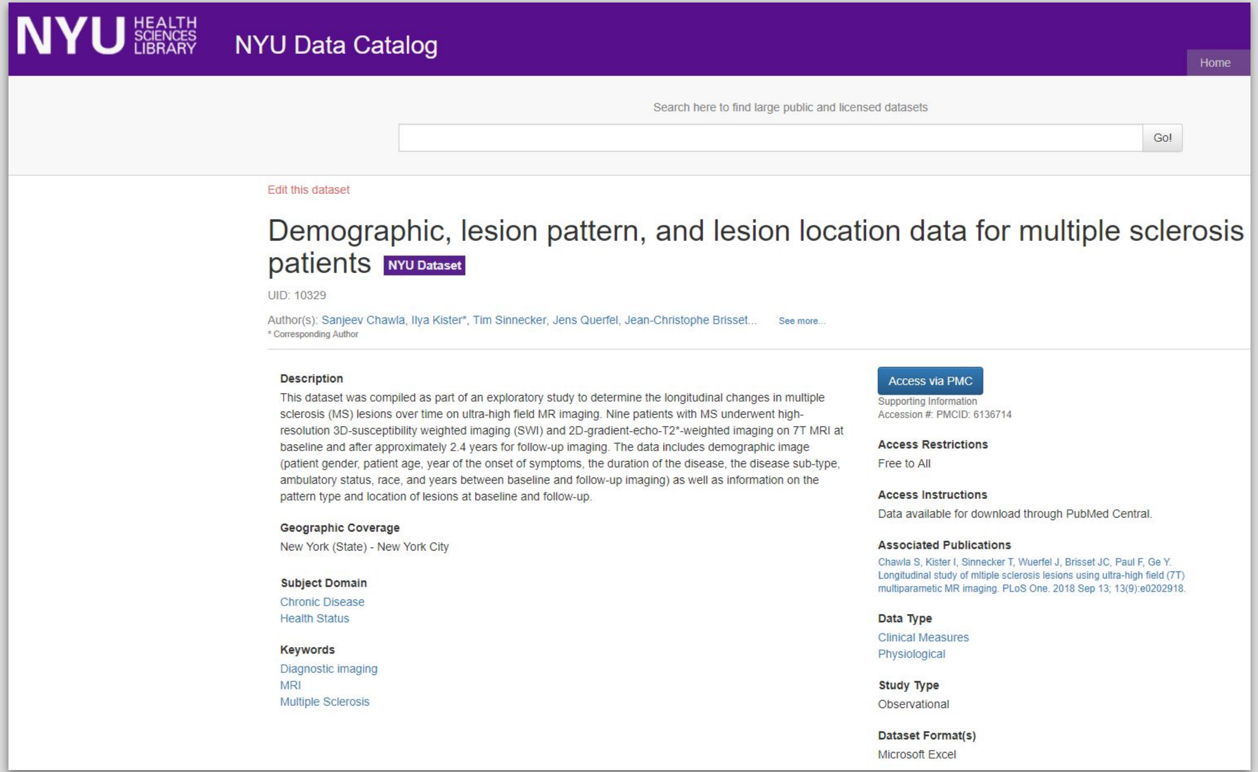
Observational

Dataset Format(s)

CSV

Electronic Health Record Data

Background: Example Data Catalog Record



Search here to find large public and licensed datasets

Edit this dataset

Demographic, lesion pattern, and lesion location data for multiple sclerosis patients NYU Dataset

UID: 10329

Author(s): Sanjeev Chawla, Ilya Kister*, Tim Sinnecker, Jens Querfel, Jean-Christophe Brisset... See more...

* Corresponding Author

Description

This dataset was compiled as part of an exploratory study to determine the longitudinal changes in multiple sclerosis (MS) lesions over time on ultra-high field MR imaging. Nine patients with MS underwent high-resolution 3D-susceptibility weighted imaging (SWI) and 2D-gradient-echo-T2*-weighted imaging on 7T MRI at baseline and after approximately 2.4 years for follow-up imaging. The data includes demographic image (patient gender, patient age, year of the onset of symptoms, the duration of the disease, the disease sub-type, ambulatory status, race, and years between baseline and follow-up imaging) as well as information on the pattern type and location of lesions at baseline and follow-up.

Geographic Coverage

New York (State) - New York City

Subject Domain

Chronic Disease
Health Status

Keywords

Diagnostic imaging
MRI
Multiple Sclerosis

Supporting Information
Accession #: PMCID: 6136714

Access Restrictions

Free to All

Access Instructions

Data available for download through PubMed Central.

Associated Publications

Chawla S, Kister I, Sinnecker T, Wuerfel J, Brisset JC, Paul F, Ge Y. Longitudinal study of mltiple sclerosis lesions using ultra-high field (7T) multiparametic MR imaging. PLoS One. 2018 Sep 13; 13(9):e0202918.

Data Type

Clinical Measures
Physiological

Study Type

Observational

Dataset Format(s)

Microsoft Excel

Datasets in Supplemental Files of Published Articles

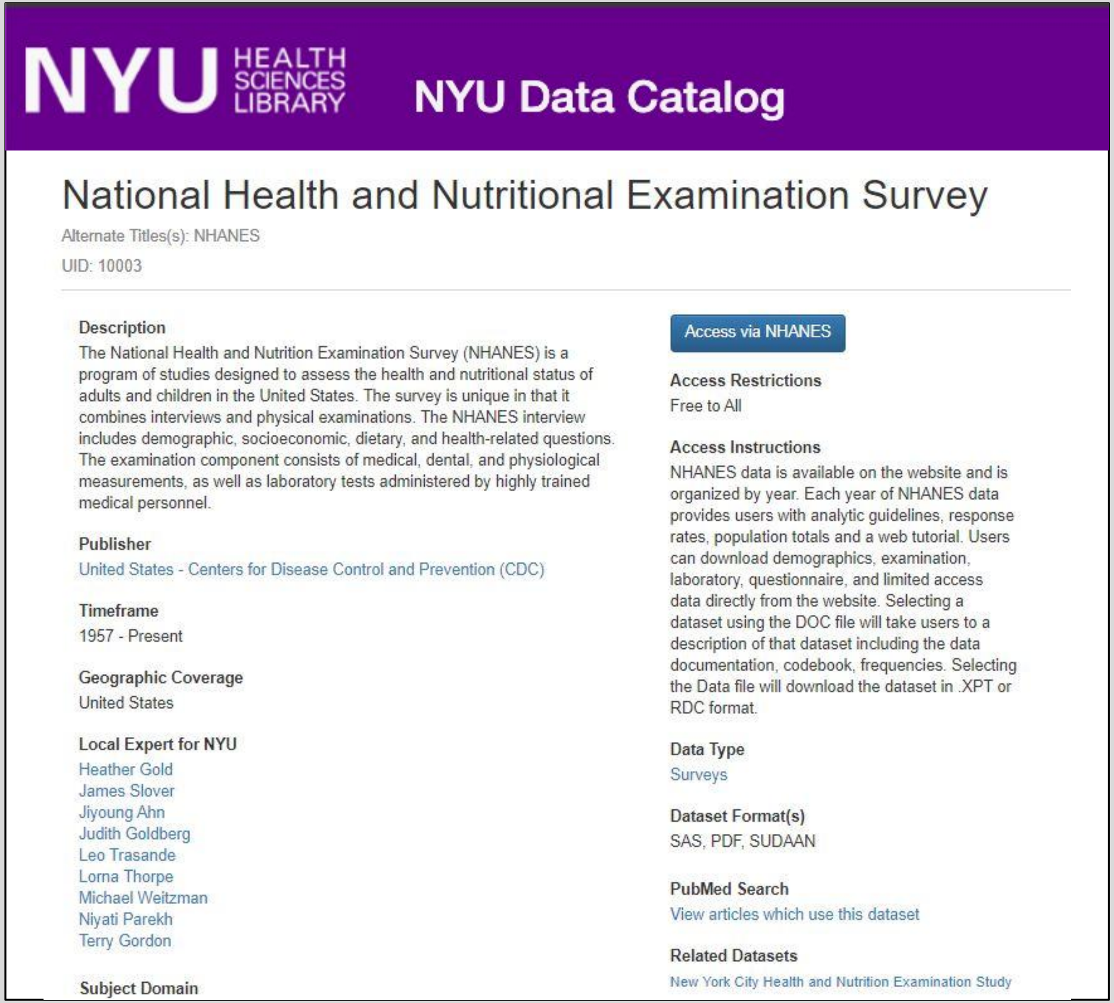
What this means for re-use

- Examples of facilitating re-use with discovery metadata
- Librarians on the ground gaining experience with facilitating re-use

Stories of Re-Use

Local Experts & Re-Use

- Metadata element designed to facilitate collaboration and re-use
- At NYU, generally used on large, third party datasets



NYU Data Catalog

National Health and Nutritional Examination Survey

Alternate Titles(s): NHANES

UID: 10003

Description
The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Publisher
United States - Centers for Disease Control and Prevention (CDC)

Timeframe
1957 - Present

Geographic Coverage
United States

Local Expert for NYU
Heather Gold
James Slover
Jiyoung Ahn
Judith Goldberg
Leo Trasande
Lorna Thorpe
Michael Weitzman
Niyati Parekh
Terry Gordon

Subject Domain

Access Restrictions
Free to All

Access Instructions
NHANES data is available on the website and is organized by year. Each year of NHANES data provides users with analytic guidelines, response rates, population totals and a web tutorial. Users can download demographics, examination, laboratory, questionnaire, and limited access data directly from the website. Selecting a dataset using the DOC file will take users to a description of that dataset including the data documentation, codebook, frequencies. Selecting the Data file will download the dataset in .XPT or RDC format.

Data Type
Surveys

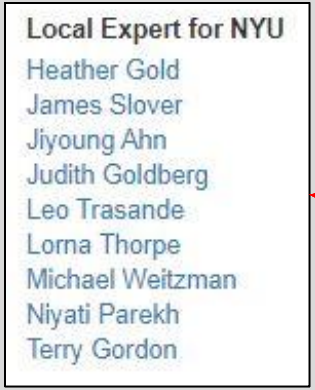
Dataset Format(s)
SAS, PDF, SUDAAN

PubMed Search
View articles which use this dataset

Related Datasets
New York City Health and Nutrition Examination Study

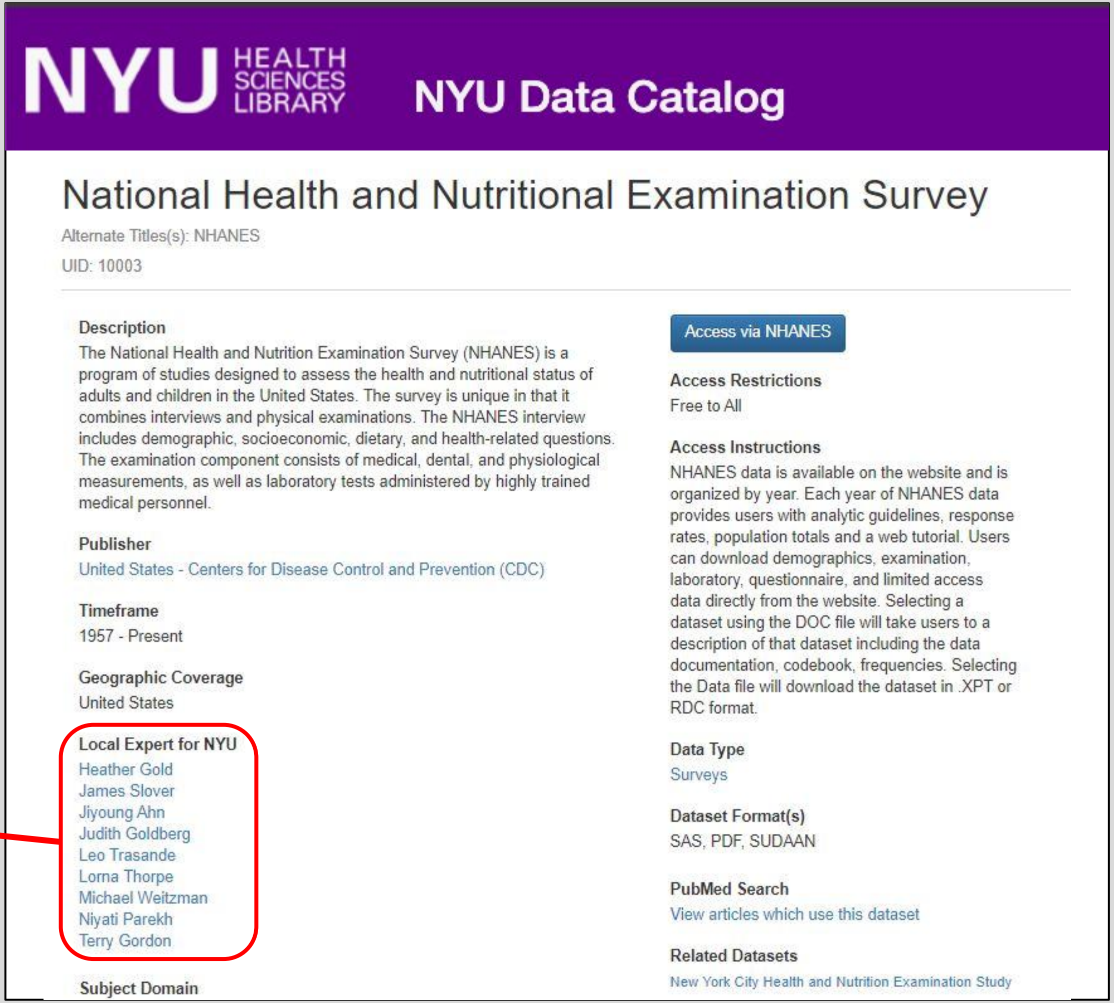
Local Experts & Re-Use

- Metadata element designed to facilitate collaboration and re-use
- At NYU, generally used on large, third party datasets



Local Expert for NYU

Heather Gold
James Slover
Jiyoung Ahn
Judith Goldberg
Leo Trasande
Lorna Thorpe
Michael Weitzman
Niyati Parekh
Terry Gordon



NYU Data Catalog

National Health and Nutritional Examination Survey

Alternate Titles(s): NHANES

UID: 10003

Description

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Publisher

United States - Centers for Disease Control and Prevention (CDC)

Timeframe

1957 - Present

Geographic Coverage

United States

Local Expert for NYU

Heather Gold
James Slover
Jiyoung Ahn
Judith Goldberg
Leo Trasande
Lorna Thorpe
Michael Weitzman
Niyati Parekh
Terry Gordon

Subject Domain

Access Restrictions

Free to All

Access Instructions

NHANES data is available on the website and is organized by year. Each year of NHANES data provides users with analytic guidelines, response rates, population totals and a web tutorial. Users can download demographics, examination, laboratory, questionnaire, and limited access data directly from the website. Selecting a dataset using the DOC file will take users to a description of that dataset including the data documentation, codebook, frequencies. Selecting the Data file will download the dataset in .XPT or RDC format.

Data Type

Surveys

Dataset Format(s)

SAS, PDF, SUDAAN

PubMed Search

View articles which use this dataset

Related Datasets

New York City Health and Nutrition Examination Study

Successes

- Local experts report being contacted about those datasets
- Local experts report becoming co-authors on papers generated from that data

Concerns

- Local experts complain that they are contacted by people outside of the institution that are not viable collaborators
- Local experts express concern about the amount of time their volunteer work takes

Takeaways

- Researchers have questions, even on incredibly well-documented data
- Researchers have limited time so answering questions can be difficult
- Researchers enter into collaborations based on conversations on datasets and re-use
- Addressing this gap in responsibility (e.g., acting as a local expert) is a key part of ensuring future success of initiatives like the NYU Data Catalog

Residents Research Practicum

- Third year residents worked with researchers and a librarian to develop original research through re-used datasets
- Residents worked as a group and located a dataset for re-use through through the NYU Data Catalog

Fred LaPolla, Research and Data Librarian, Lead of Data Education and Course Director

Takeaways

- Residents and practicum leads needed to discuss the datasets with the original creators, even as the dataset was well-documented
- A poster was accepted at a conference based on the resident's work
- Residents were not able to publish on their research due to limitations from one of the funders from the original research
- We only know of this interaction due to librarian participation - further investigation into user tracking is necessary

Future Efforts

- Further investigation into re-use use cases with the Data Discovery Collaboration Project
- Creation of infrastructure to help “Local Experts” manage their requests

Thank you

Nicole Contaxis, MLIS
Lead, NYU Data Catalog