Exploratory Tools for National Children's Study Data

Peter Gilbertson[1], Linda Andrews[1], Jodi Cisewski[1], Cora MacPherson[1], Christina Park[2], Anita Johnson[1], Jack Moye, Jr.[3]

[1]Social & Scientific Systems, Inc., Silver Spring, Maryland, USA; [2]National Institutes of Health, Office of the Director; [3]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development

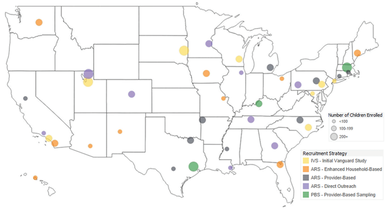
ABSTRACT

The National Children's Study Archive is an information, data, and sample repository for the National Children's Study, designed to make data and samples freely available for scientific research, with an approved research request. The study, which was active from 2009 to 2014, collected data and samples from over 12,000 mothers, fathers, and children across the US at 40 study locations. These participants are currently represented by nearly 14,000 variables, over 200,000 biological samples, and over 4,000 environmental samples in the Archive.

In order to facilitate meaningful evaluation and consumption of this wealth of data, the Archive has developed new exploratory tools. The Protocol Browser allows users to flow through the visit progression and visit instrumentation and identify available study datasets. The Participant Explorer allows users to investigate study participation by participant type (woman, child, father), demographics (e.g., education level, marital status), and data collection point. The Sample Explorer allows researchers to use demographics and study visit information to explore the available biological (blood, hair, nails, saliva, urine, vaginal swab, breast milk, cord blood, meconium, and placenta) and environmental (air, dust, water) primary and derivative samples that were collected from a subset of NCS families. Both the Participant and Sample Explorers aid researchers in determining potential sample sizes and the number of relevant biological and environmental samples that fit their criteria. The Variable Locator allows users to search the available NCS datasets for questions and variables of interest, returning specific data elements available. With these research tools, researchers targeting unique populations and topic areas can quickly establish if a particular topic area is represented in NCS data and samples. Once identified, researchers can use the Archive's proposal submission and review process to begin using NCS resources to pursue their scientific objectives.

BACKGROUND

The National Children's Study (NCS) Vanguard Study was a pilot for the planned Main Study cohort and was started in 2009. The NCS was conceived to be a nationally representative longitudinal cohort study of 100,000 children from before birth through age 21. The study aimed to evaluate environmental influences on child health and development. The Vanguard Study tested different recruitment strategies in 40 locations across 31 states throughout the U.S.



The study enrolled over 5,400 birth families and followed them through 2014. At protocol-specified study visits, questionnaires and interviews, neuro-psychosocial and cognitive assessments, and physical examination data were collected, along with nearly 19,000 biological and 4,000 environmental primary samples from which a sample repository of over 250,000 items was created.

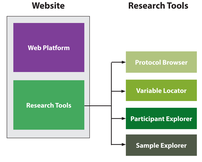
PROBLEM

Faced with the problem of sharing complex data and study documentation, the NICHD identified the need to create a single data archive and sample repository with exploratory tools to facilitate ongoing scholarly endeavors around the NCS. From this charge, the NCS archive (www.ncsarchive.s-3.net) was developed to present the available study data and samples, provide supporting documentation, monitor and track proposed research efforts, and provide a secure environment where researchers, new or seasoned, could collaborate with support staff.

SOLUTION: RESEARCH TOOLS

Recognizing the core NCS Archive user will shift over time from NCS experienced to NCS inexperienced users, it was identified that systems or tools need to be developed to quickly convey study information. To facilitate the meaningful evaluation and consumption of this wealth of data, the Archive developed four new exploratory tools: (1) Protocol Browser, (2) Variable Locator, (3) Participant Explorer, and (4) Sample Explorer (Figure 1).

Figure 1. The Archive support tools



The Protocol Browser

One of the challenges for the NCS is that the study was conducted with two main protocols—the Initial Vanguard Study (IVS) Protocol (2009–2010) and the Alternate Recruitment Substudy (ARS) Protocol (2011–2014). Each of these protocols included participant visits across multiple life stages (pre-conception, prenatal, perinatal, postnatal) each with different instrumentation. For a user to quickly assess how a participant progressed in the NCS, the Protocol Browser was developed.

The Protocol Browser (Figure 2) shows points of participant contact within each protocol and displays those points of contact by life stage, participant visit, and instruments administered. Additionally, for each administered instrument a research dataset is identified if available. In addition to the flow through an individual protocol, the browser allows users to compare protocols (Figure 3) to understand how and if participants in each protocol experience the study differently.

Figure 2. Protocol Browser protocol flow through dataset

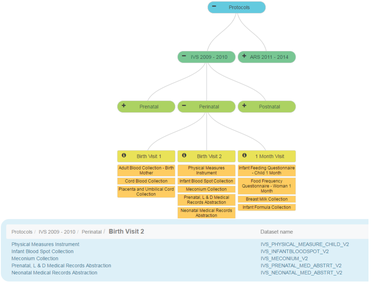
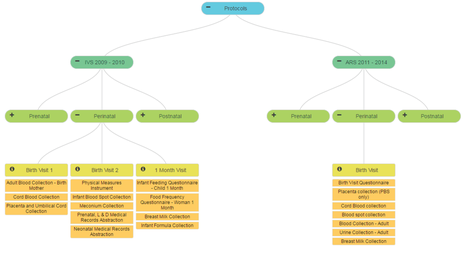


Figure 3. Protocol Browser protocol comparisons



The Variable Locator

The NCS Archive currently has information available from nearly 14,000 variables collected over the course of the NCS. Because many researchers tend to focus on one area of expertise, there needs to be a way to quickly evaluate what information was collected from NCS participants and how many data points, or responses, are available in that area of interest. To support this discovery the Variable Locator was created. The Variable Locator is a free text and variable name search for all data currently available in the NCS Archive. For example, a user can type in "sleep" and will be provided with a list of all variables and questions on sleeping, along with the number of participants who responded to the question and the number of records in the dataset (Figure 4). The Variable Locator includes the Study Phase, name of the corresponding dataset, variable name, variable label (question text), number of valid responses, number of legitimate skip responses, number of other responses, and total number of variable records. The locator allows users to use Boolean search terms AND, NOT, and OR to further filter their results. With this tool, a researcher can quickly identify the potential of NCS Archive data for their future research.

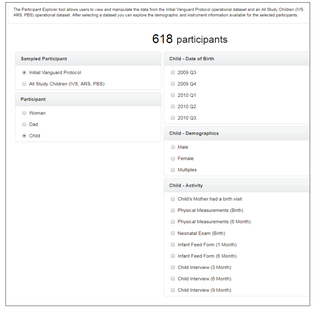
Figure 4. Variable Locator

The Participant Explorer

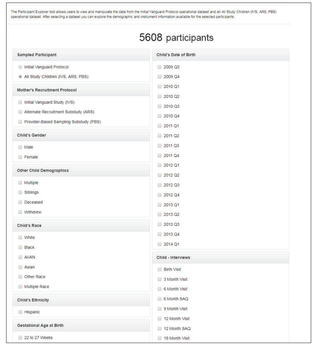
The Participant Explorer is a tool designed to allow users to investigate NCS participants and their study participation on a metadata level. Users can see participant counts by participant type (woman, child, father), demographics (race, ethnicity, education level, marital status), and data collection point. Additionally, child participant can be categorized by gestational age, birth weight, and quarter of birth between 2009 and 2014. Users interested in family populations can review the demographic and activity characteristics for the 3000+ women, children, and fathers involved in the IVS phase of the NCS. Researchers focused on child health can review information on the 5500+ child in all NCS recruitment phases of the NCS (Figure 5).

Figure 5. Participant Explorer

Participant Explorer: Initial Vanguard Protocol



Participant Explorer: All Study Children

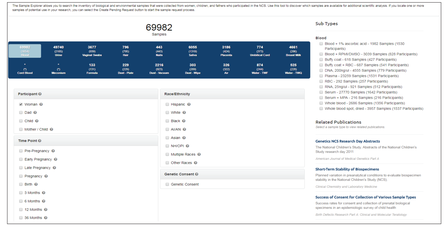


The Sample Explorer

There are currently 219,000 biological specimens and 4,600 environmental samples available connected with NCS participants. This large collection positions the Archive as a strong resource for future analytical studies. The Sample Explorer tool provides users with a method of exploring the available samples in the context of the participants that provided them.

The Sample Explorer allows researchers to use demographics, participant type, race, and ethnicity, and study visit information to explore the available biological (blood, hair, nails, saliva, urine, vaginal swab, breast milk, cord blood, meconium, and placenta) and environmental (air, dust, water) primary and derivative samples that were collected from a subset of NCS families. With the tool, researchers can identify the number of biological and environmental samples available that fit the types of participants of interest to their research. With sample number, sample demographic, and sample size information identified, a researcher can then put forth targeted research proposals requesting the samples of most interest to their research.

Figure 6. Sample Explorer



CONCLUSIONS

With these research tools, researchers targeting unique populations, focused topic areas, and specific biological or environmental samples can quickly establish whether the data or samples of interest are represented in the NCS Archive. These research tools empower researchers to explore and refine study information in their own areas of study. Once identified, researchers can use the Archive's proposal submission and review process to begin using NCS resources to pursue their scientific objectives.

For more information, please contact:

NICHD website https://www.nichd.nih.gov/research/NCS/
NCS Archive website https://ncsarchive.s-3.net
Archive Email NCSArchi

THE NATIONAL CHILDREN'S STUDY

NIH *Eunice Kennedy Shriver* National Institute of Child Health and Human Development