

VIEWPOINT

Ensuring the Best Use of Data

The Adolescent Brain Cognitive Development Study

Wilson M. Compton, MD, MPE

National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland.

Gayathri J. Dowling, PhD

National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland.

Hugh Garavan, PhD

Department of Psychiatry, University of Vermont, Burlington.

Data sharing is increasingly acknowledged to be a feature of a healthy scientific ecosystem, maximizing the benefits from the often costly business of collecting scientific data and enhancing discovery. Thus, timely data sharing from large research projects is an explicitly stated policy of the National Institutes of Health (NIH). Making data openly and freely available and encouraging researchers to use them for additional analyses ensures the maximum return on the scientific investments that the NIH, and ultimately the US taxpayer, have made.

The Adolescent Brain Cognitive Development (ABCD) study is a prime and successful example of the open data-sharing philosophy of the NIH. This ambitious 10-year study of brain development and child health in the United States is in its third year of collecting neuroimaging, genetic, and behavioral information and has completed baseline data collection on 11 878 US children who were recruited at age 9 to 10 years. The study is designed to measure brain development using structural and functional magnetic resonance imaging and to investigate the role of various biological, environmental, and behavioral factors on brain, cognitive, and social/emotional development.¹ Researchers have been encouraged to use this rich, open data set.

So far, the scientific community has responded. Two batches of ABCD data have been released, the first in February 2018, which included the children from the first year of recruitment, and the second in April 2019, which included the full baseline sample. Multiple research groups have already published analyses on neurobiological correlates of screen time,² neurocognitive associations with problem behaviors,³ construct validity and psychometric properties of a measure of prodromal psychotic-like symptoms,⁴ minority sexual orientation and gender identity,^{5,6} eating disorders,⁷ and neurobiological associations with anhedonia,⁸ and more articles are appearing regularly.

The uptake of the initial ABCD data set for such diverse analyses is encouraging. However, to use the data most effectively, it is important to understand the strengths as well as the limitations of the data set. For example, while the study sample and design are well suited for conducting cross-sectional and longitudinal analyses, it would not be appropriate to take the ABCD study cohort as fully representative of the US population for the purposes of calculating population prevalence estimates.⁹ Yet, some researchers conducting secondary analyses of ABCD data have done so, which could potentially produce misleading conclusions.

A Research Letter by Calzo and Blashill⁵ describes the sample as a "US representative cohort." These authors similarly describe the sample as "US representative" in a more recent article.⁶ Another Research Letter

by Rozzell and colleagues⁷ also describes the ABCD sample as "nationally representative." Because "representative" is neither easily defined nor proven, we propose that a more nuanced and accurate description of the sampling frame is needed by researchers reporting results of the ABCD study.

The ABCD study sought to recruit a sample that mirrors US population demographics by recruiting through geographically, demographically, and socioeconomically diverse school systems surrounding each of the 21 research sites. Informed by epidemiological methods, a stratified probability sample of schools was selected for each site based on sex, race/ethnicity, socioeconomic status, and urbanicity to minimize systematic sampling biases in recruitment at the school level. As a result, the ABCD study has approximated the diversity of the US population on sex, race/ethnicity, and socioeconomic status. Data analytic approaches can adjust for demographic differences between the ABCD sample and the population, but it does not necessarily follow that the data therefore provide unbiased population rates of disorders or particular behaviors. Other characteristics that are associated with these disorders or behaviors may not have been fully sampled. As stated in the ABCD design description article,⁹ "Designing the ABCD sample demographics to match those of the national target population does not guarantee that the sample will be representative across all of the many dimensions (demographics, family and individual factors, community and environment, behaviors, exposures) that may influence a child's development."

Although the ABCD sample was recruited at 21 sites spread throughout the United States, site selection was driven by the locations of research teams deemed meritorious by the NIH peer review system; thus, not all 9-year-old and 10-year-old children in the United States had an equal chance of being invited to participate in the study. For instance, we know that ABCD under-recruited rural families because of neuroimaging facilities tending to be in mostly urban research centers. Also, research participation is voluntary and incomplete within any particular school, and the response rates for individual students within the schools are not incorporated into the sample weighting schema. In addition, the final sample all self-selected into the study; self-selection of this sort can skew a sample away from being truly representative because of biases associated with participating in research. Indeed, although the ABCD sample includes children from various socioeconomic levels, it does overrepresent families with parents who earn higher incomes and have completed more years of education. Consequently, while the sample is designed to be epidemiologically informed and to minimize

Corresponding

Author: Wilson M. Compton, MD, MPE, National Institute on Drug Abuse, National Institutes of Health, 6001 Executive Blvd, MSC 9581, Bethesda, MD 20892-9581 (wcompton@nida.nih.gov).

selection bias,⁹ how representative it is of the US population will likely vary by the specific measure being investigated, the subset of the sample included in an analysis, and the extent to which the weighting methods or model covariates capture factors that affect the outcome of interest.

Understanding the methodological specifics of the study, such as the way participants were recruited, will ensure the scientific value of all ABCD analyses and forestall the drawing of potentially erroneous or misleading conclusions from these data. While we discourage inaccurate and potentially misleading descriptions of this research resource, we encourage full use of the data via the National Institute of Mental Health (NIMH) Data Archive (<https://data-archive.nimh.nih.gov/abcd>) for scientific discovery. For a study of trajectories and correlates of brain development, the size and diversity of the sample provide many opportunities for scientific discovery that will surely yield insights generalizable to most US youth.

The rapid release of data for use by the scientific community is a key strength of the ABCD study, and we encourage use of the data to enhance knowledge of youth development and outcomes. The diversity of the study sample and the range of neurobiological, genetic, behavioral, and other measures being collected make it ripe for analyses that the study planners and researchers could not envision. We expect that publications over the next decade based on analyses of the annual releases of data will demonstrate the study's power for this kind of research. However, to ensure that the scientific output from the study is valid and accurate, details of the sampling frame need to be considered to accurately describe the study as having a population-based, demographically diverse sample but one that is not necessarily representative of the US population. Attention to these and other methodological details will help to ensure the scientific rigor and validity of the findings.¹

ARTICLE INFORMATION

Published Online: July 15, 2019.

doi:10.1001/jamapediatrics.2019.2081

Conflict of Interest Disclosures: Dr Compton reports long-term stock holdings in General Electric Company, Pfizer Inc, and 3M Companies. No other disclosures were reported.

Disclaimer: The opinions expressed in this commentary do not necessarily represent the opinions of the National Institute on Drug Abuse, the National Institutes of Health, or the US Department of Health and Human Services.

Additional Contributions: We acknowledge the support of Eric Wargo, PhD, National Institute on Drug Abuse, in the development of this commentary and thank the 9988 families with their 11 878 children who have participated in the Adolescent Brain Cognitive Development study.

REFERENCES

1. Volkow ND, Koob GF, Croyle RT, et al. The conception of the ABCD study: from substance use

to a broad NIH collaboration. *Dev Cogn Neurosci*. 2018;32:4-7. doi:10.1016/j.dcn.2017.10.002

2. Paulus MP, Squeglia LM, Bagot K, et al. Screen media activity and brain structure in youth: evidence for diverse structural correlation networks from the ABCD study. *Neuroimage*. 2019;185:140-153. doi:10.1016/j.neuroimage.2018.10.040

3. Thompson WK, Barch DM, Bjork JM, et al. The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: findings from the ABCD study's baseline neurocognitive battery. *Dev Cogn Neurosci*. 2019;36:100606. doi:10.1016/j.dcn.2018.12.004

4. Karcher NR, Barch DM, Avenevoli S, et al. Assessment of the Prodromal Questionnaire-Brief Child Version for measurement of self-reported psychoticlike experiences in childhood. *JAMA Psychiatry*. 2018;75(8):853-861. doi:10.1001/jamapsychiatry.2018.1334

5. Calzo JP, Blashill AJ. Child sexual orientation and gender identity in the Adolescent Brain Cognitive Development cohort study. *JAMA Pediatr*. 2018;172

(11):1090-1092. doi:10.1001/jamapediatrics.2018.2496

6. Blashill AJ, Calzo JP. Sexual minority children: Mood disorders and suicidality disparities. *J Affect Disord*. 2019;246:96-98. doi:10.1016/j.jad.2018.12.040

7. Rozzell K, Moon DY, Klimek P, Brown T, Blashill AJ. Prevalence of eating disorders among US children aged 9 to 10 years: data from the Adolescent Brain Cognitive Development (ABCD) study. *JAMA Pediatr*. 2019;173(1):100-101. doi:10.1001/jamapediatrics.2018.3678

8. Pornpattananangkul N, Leibenluft E, Pine DS, Stringaris A. Association between childhood anhedonia and alterations in large-scale resting-state networks and task-evoked activation [published online March 13, 2019]. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2019.0020

9. Garavan H, Bartsch H, Conway K, et al. Recruiting the ABCD sample: design considerations and procedures. *Dev Cogn Neurosci*. 2018;32:16-22. doi:10.1016/j.dcn.2018.04.004