Transparency and reproducibility in the Adolescent Brain Cognitive Development (ABCD) study

Daniel A. Lopez [a,s], Carlos Cardenas-Iniguez [b], Punitha Subramaniam [c], Shana Adise [d], Katherine L. Bottenhorn [b], Paola Badilla [e], Ellen Mukwekwerere [f], Laila Tally [g], Omoengheme Ahanmisi [h], Isabelle L. Bedichek [i], Serena D. Matera [j], Gabriela Mercedes Perez-Tamayo [c], Nicholas Sissons [k], Owen Winters [l], Anya Harkness [m], Elizabeth Nakiyingi [f], Jennell Encizo [f], Zhuoran Xiang [n], Isabelle G. Wilson [o], Allison N. Smith [p], Anthony R. Hill [a], Amanda K. Adames [q], Elizabeth Robertson [l], Joseph R. Boughter [f], Arturo Lopez-Flores [a], Emma R. Skoler [e], Lyndsey Dorholt [r], Bonnie J. Nagel [a,s], Rebekah S. Huber [a,c,s,*]

[a] Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States
[b] Department of Population and Public Health Sciences, Keck School of Medicine of the University of Southern California, Los Angeles, CA, United States
[c] Department of Psychiatry, University of Utah, Salt Lake City, UT, United States
[d] Division of Endocrinology, Diabetes and Metabolism, Children's Hospital of Los Angeles, Los Angeles, CA, United States
[e] Institute for Behavioral Genetics, University of Colorado, Boulder, CO, United States
[f] Department of Psychology, Psychiatry, and Radiology, Washington University in St. Louis, St. Louis, MO, United States
[g] Center for Children and Families and Department of Psychology, Florida International University, Miami, FL, United States
[h] Department of Diagnostic Radiology & Nuclear Medicine, University of Maryland, Baltimore, MD, United States
[i] Institute for Drug and Alcohol Studies, Virginia Commonwealth University, Richmond, VA, United States
[j] The Frederick J. and Marion A. Schindler Cognitive Neurophysiology Laboratory Department of Neuroscience and The Ernest J. Del Monte Institute for Neuroscience University of Rochester School of Medicine and Dentistry, Rochester, NY, United States
[k] Departments of Psychiatry and Radiology, University of Vermont, Burlington, VT, United States
[l] Department of Psychiatry, Medical University of South Carolina, Charleston, SC, United States
[m] Center for Health Sciences, SRI International, Menlo Park, CS, United States
[n] Department of Psychology, Yale University, New Haven, CT, United States
[o] Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI, United States
[p] Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[q] Department of Psychiatry, University of California, San Diego, San Diego, CA, United States
[r] Department of Psychology, University of Minnesota, Minneapolis, MN, United States
[s] Center for Mental Health Innovation, Oregon Health & Science University, Portland, OR, United States

ARTICLE INFO

ABSTRACT

Background: Transparency can build trust in the scientific process, but scientific findings can be undermined by poor and obscure data use and reporting practices. The purpose of this work is to report how data from the Adolescent Brain Cognitive Development (ABCD) Study has been used to date, and to provide practical recommendations on how to improve the transparency and reproducibility of findings.

Methods: Articles published from 2017 to 2023 that used ABCD Study data were reviewed using more than 30 data extraction items to gather information on data use practices. Total frequencies were reported for each extraction item, along with computation of a Level of Completeness (LOC) score that represented overall endorsement of extraction items. Univariate linear regression models were used to examine the correlation between LOC scores and individual extraction items. Post hoc analysis included examination of whether LOC scores were correlated with the logged 2-year journal impact factor.

Results: There were 549 full-length articles included in the main analysis. Analytic scripts were shared in 30 % of full-length articles. The number of participants excluded due to missing data was reported in 60 % of articles, and

---

* Corresponding author at: Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States.
  E-mail address: huberr@ohsu.edu (R.S. Huber).

information on missing data for individual variables (e.g., household income) was provided in 38 % of articles. A table describing the analytic sample was included in 83 % of articles. A race and/or ethnicity variable was included in 78 % of reviewed articles, while its inclusion was justified in only 41 % of these articles. LOC scores were highly correlated with extraction items related to examination of missing data. A bottom 10 % of LOC score was significantly correlated with a lower logged journal impact factor when compared to the top 10 % of LOC scores (β=-0.77, 95 % −1.02, −0.51; *p*-value < 0.0001).

Conclusion: These findings highlight opportunities for improvement in future papers using ABCD Study data to readily adapt analytic practices for better transparency and reproducibility efforts. A list of recommendations is provided to facilitate adherence in future research.

1. Introduction

Transparency in science is necessary for reproducibility. Transparency pertains to the openness and accessibility of methods and analytic decisions involved in the research process (Korbmacher et al., 2023). Despite the benefits of transparency, there exists substantial variance in the reporting of methodology. Consequently, this impacts reproducibility in science. Reproducibility refers to the minimum requirements necessary to generate the same analytic results when using a common dataset (Parsons et al., 2022). Although separate concepts, many concerns related to reproducibility stem from past failed efforts related to replicability, the quality of obtaining consistent results across studies with the same research question using new data (Ioannidis, 2005; Sciences NAo, Policy et al., 2019). A landmark study examining 100 psychology papers reported that only 36 % of studies reporting statistically significant findings could be replicated, and that effect sizes were routinely overestimated (Collaboration, 2015). This problem is not specific to any discipline. A reproducibility project in the cancer biology field reported a 46 % replication success rate, and reproducibility hurdles included a lack of publicly available analytic code and failures to report statistical analyses in 81 % and 40 % of reviewed papers, respectively (Errington et al., 2021a, 2021b). There are multiple reasons for the transparency and reproducibility problems. A survey of 1576 researchers found that low reproducibility may be due to the unavailability of methods and code, while more statistical rigor and better mentorship emerged as practices that could improve reproducibility (Baker, 2016a). Many of these issues are rooted in an academic culture that relies on "statistically significant" results rather than focusing on the practical relevance of findings (Joober et al., 2012; Nuzzo, 2014). The reliance on statistical thresholds can lead to questionable research practices (e.g., *p*-hacking) (Andrade, 2021). These shortcomings in the research culture can hinder transparency efforts and compromise the public's trust in science.

Making data available to the broader research community through an open science model promotes collaboration, replication of findings, and acceleration of scientific discovery (Parsons et al., 2022). Large publicly available datasets, like the Adolescent Brain Cognitive Development (ABCD) Study®, have made it possible for researchers to access and analyze research data from various domains such as behavioral, genetics, and neuroimaging (Volkow et al., 2018). The ABCD Study is the largest longitudinal study of brain development in children in the United States with 11,875 youth enrolled at the ages of 9–10 years from 21 sites that will be followed for approximately 10 years (Garavan et al., 2018). This study is a critical resource for the field of developmental neuroscience that offers researchers a wealth of data to better understand factors that influence adolescent brain development. Data from the study has already been used to help chart neurodevelopmental trajectories (Bethlehem et al., 2022) and its findings will continue to impact our current understanding of adolescent development. Furthermore, findings from the study may inform education, policy, and transform clinical practice by identifying early markers of psychopathology and substance use that can lead to targeted interventions and prevention strategies. Recent research, largely in psychology and neuroscience, has stressed the importance of transparency in methodology to improve the

reproducibility of findings (Klapwijk et al., 2021; Nichols et al., 2017). Among these recommendations include openness in sharing analytic decisions (e.g., number of excluded participants) (Klapwijk et al., 2021; Nichols et al., 2017; Munafò et al., 2017). Whether guidance from recent initiatives has influenced the reporting of ABCD Study findings is not known.

The ABCD Study's influential role in developmental neuroscience and adolescent health necessitates a thorough accounting of how the data have been used. The open science model embraced by the ABCD Study also provides an opportunity to assess the reporting practices in the field, identify gaps, and provide recommendations. Assessing and improving reporting practices is essential for ensuring the reliability, reproducibility, and transparency of research findings. The ABCD Study, given its extensive scope and rigorous methodologies, can play a pivotal role in elevating reporting standards in developmental neuroscience.

The aims of the current study are two-fold. The first aim was to assess the reporting practices of publications, focusing on those using the ABCD Study publicly available data due to its use by many researchers. To do so, we reviewed and extracted information related to data use practices from papers that used the ABCD Study data. The second aim was to provide a set of recommendations on best practices when using the ABCD Study data that may also be applied more broadly to the field of developmental neuroscience. The motivation to examine the data use came from the goals of the ABCD Justice, Equity, Diversity, and Inclusion (JEDI) Responsible Use of Data Workgroup. The JEDI Workgroup strives to ensure that ABCD Study data are used in a way that prevents further stigmatization or marginalization of individuals. To this end, the JEDI Workgroup focuses on creating resources for ABCD data users to encourage best practices and facilitate responsible data use.

2. Methods

A list of articles mentioning the ABCD Study were provided to the authors by the ABCD Data Analysis, Informatics & Resource Center (DAIRC). The list of articles (n=676) included full research articles, abstracts, news articles, and research letters that were published from April 2017 (the date of the first ABCD Study publication) to May 2023. All articles were obtained either through open access or via institutional membership. Articles were considered for inclusion if they were in English and peer reviewed. Abstracts and research letters were analyzed separately. News articles, reviews, commentaries, and methods papers (i.e., manuscripts that predominantly described the protocol and instruments used in the ABCD Study) were excluded.

Articles were reviewed by a team of 28 researchers (22 Research Assistants at 15 separate ABCD Study sites and 6 PhD level researchers) across the ABCD Consortium. All reviewers received training on data extraction, and the full list of measures was accompanied by a written and video tutorial that included a more in-depth explanation with visual examples (i.e., an extract from an ABCD Study paper). Each reviewer was assigned up to 25 articles. Google Forms was used for the collation of items extracted from each article reviewed, and then the submission was assessed for eligibility for inclusion in the main analysis by the first author (DAL), who then checked each Google Forms submission to determine the year of publication and whether the article was an

abstract, research letter, methods, or a full-length paper. The original list of articles included some duplicates due to resubmissions (e.g., an article was resubmitted due to an erratum or retraction). This resulted in certain manuscripts being reviewed more than once. Discrepancies between duplicate submissions (e.g., due to a different response for the same question) were compared after a separate review of the manuscript by DAL. The duplicate submission with the worst overall accuracy as determined by the lead reviewer was removed from the main analysis. The overall accuracy rate between duplicate submissions was 83.9 %. In total, there were 83 method papers and 15 duplicate articles excluded from the main analysis. In addition, reviewers were asked to flag any problematic articles that required further discussion. A note section was included at the end of each form where reviewers could provide any additional information that they felt was important. Flagged articles (n=8 full-length articles) and note sections were evaluated by DAL and addressed when necessary. Reviewers communicated with DAL to resolve any questions and to fix data entry errors prior to the analysis. The main analysis included only full-length manuscripts that analyzed ABCD Study data.

2.1. Data extraction items

There were 32 items extracted from each research article, and 10 gated items that were only asked if the previous question was endorsed. The data extraction items were created after a review of the neuroimaging best practices literature (Klapwijk et al., 2021; Nichols et al., 2017; Gilmore et al., 2017). The items covered a wide range of issues related to transparency and reproducibility. Items related to replicability were not included due to the focus on the ABCD dataset. Data extraction items were initially created by the lead reviewer and were then reviewed by 5 PhD level researchers. Following feedback from the PhD level researchers, revisions were made and a final list of the data extraction items were created. The full set of data extraction items was then piloted by three RAs and two PhD level researchers. Individual items had a "Yes" or "No" option to note the presence or absence of the practice, and four measures included a third option. The full list of items can be found at: https://osf.io/qkefw/ and in the supplementary materials.

Overall, items 1 through 12 were concerned with analysis-level reproducibility, and items 13–26 broadly covered the transparent reporting of methods and results. Items 1–6 addressed the sharing of analysis scripts, software used, ABCD data release and Digital Object Identifier (DOI) information. Items 7–12 dealt with missing data issues, and whether articles considered limitations related to nonrandom missingness. Items 13–17 inquired about the inclusion of sociodemographic variables and had a gated question asking whether the article explained the inclusion of the variable as has been recommended for sound statistical modeling (Greenland and Pearce, 2015). Sociodemographic variables were emphasized due to their widespread habitual inclusion in models. Item 18 and 18b asked whether the article mentioned the manipulation of any variable (e.g., categorizing a continuous variable), and whether there was reasoning for the manipulation. Items 19–26 were concerned with the reporting of effect estimates and *p*-values, multiple comparisons, testing of statistical assumptions, and exploration of variables. Item 24 asked whether the researchers used the Data Exploration and Analysis Portal (DEAP) for their analysis. DEAP is a statistical analysis platform hosted by the DAIRC that can be used to analyze ABCD data (Heeringa and Berglund, 2020). Items 27–29 were used to further identify characteristics of the article (e.g., whether imaging or genetics data were used). Items 30–32 were related to preregistration, the discussion of limitations, and whether author contributions were included.

2.2. Level of completeness

A Level of Completeness score was calculated for each article to summarize overall adoption of practices. Completeness has previously been used to assess adherence to recommended practices in articles analyzing large cohort studies (Gibson et al., 2023). The Level of Completeness score was calculated by first converting binary variables into a numeric variable (Yes=1, No=0) and then summing up the following items: 1, 1b, 2, 2b, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13b, 14b, 15b, 16b, 18b, 19, 19b, 20, 21b, 22, 23, 25, 26, 30, 31. The included items can be found in the supplementary materials (Figure S1). Items 10, 11, and 31 were recoded from a categorical to a binary variable to improve overall interpretation. Items 10 and 11 had the "no participants were excluded" responses recoded as a 0. Item 31 had a "no missing data/attrition" response that was coded as a 0 (n = 71). Item 28b was related to rationalization of brain regions when using imaging data and was not included to avoid score inflation. Item 32 (whether author contributions were indicated in the manuscript) was not included in the Level of Completeness score due to the journal specific nature of the item. The possible range of Level of Completeness scores was 0–28, with a higher score indicating a greater level of adoption of data extraction items.

The decision to use all items for a summary score was made after a review of the underlying structure of the data extraction items. An exploratory factor analysis (EFA) with an oblimin rotation was conducted on the data extraction items using Mplus version 8.1 (Muthén et al., 2017). Binary items were treated as dichotomous variables by using the tetrachoric correlation structure. The results of the EFA (Tables S1-S4) did not find evidence of a meaningful clustering between items.

2.3. Statistical analysis

The percentage of endorsement of each item was reported for articles included in the main analysis. Percentages were calculated and reported separately for abstracts and research letters. Additional analyses used univariate linear regression to examine the correlation between individual extraction items and Level of Completeness scores. The purpose of the regression was to examine the overall influence of each extraction item on Completeness score. Results of the univariate linear regression represent the score difference associated with individual items. Statistical analysis of the data was performed using R version 4.2.2 and R Studio version 2023.06.01 build 524 (Team R, 2019; Team RC, 2013). Figures were created using the ggplot2 package in R (Wickham, 2011). Checks for normality were performed using the car package in R (Fox et al., 2012). Data preparation was completed using Microsoft Excel and the dplyr package in R (Wickham et al., 2019; Corporation M, 2018). R scripts are available at: github.com/Daniel-Adan-Lopez/ABCD_Transparency_Reproducibility.

2.4. Sensitivity analysis

A machine learning wrapper algorithm was used to estimate variable importance in relation to Level of Completeness score via feature selection. With the Boruta package (version 8) in R, we used a Random Forest based feature selection method to determine important and non-important attributes (Kursa and Rudnicki, 2010). A Boruta model was created using Level of Completeness score as the dependent variable and the data extraction items as the independent variable (i.e., features). Missing values were coded as "Not applicable" to prevent errors while running the Boruta wrapper algorithm.

2.5. Post hoc analysis

A post hoc analysis was conducted to examine the relationship between Level of Completeness score and journal impact factor. The most recently disseminated 2-year journal impact factor was collected for articles that scored in the top 10 % (n=56), middle 10 % (n=56), and bottom 10 % (n=56) of Completeness scores. Only full-length articles (i. e., not abstracts, methods papers, or research letters) were included in

the post hoc analysis. The impact factor was logged to reduce the influence of outliers (Figure S2). A univariate linear regression model was then used to examine the correlation between Completeness category (bottom 10 %, middle 10 %, top 10 %) and the log of the impact factor. Completeness score was also examined as a continuous measure.

3. Results

The initial review included 679 articles. Abstracts (n=18) and research letters (n=11) were separately analyzed. An additional two articles were removed due to not analyzing ABCD data. The final analysis included 549 full-length research articles (Fig. 1). Percentages of each response for the data extraction items can be found in Fig. 2.

3.1. Analysis-level reproducibility

A link to the analysis scripts were included in 30.2 % (n=166) of the articles, and 93.4 % (n=155) of the included links were active and working. Software information was mentioned in 91 % (n=499) of the articles. The software version (e.g., R version 4.2.2) was included in 75.3 % (n=376) of articles. The ABCD Study data release version (e.g.,

ABCD version 4.0) and digital-object identifier (DOI) were mentioned in 79 % (n=433) and 54 % (n=294) of articles, respectively. The ABCD Study visit number (e.g., Baseline, 1-year follow-up visit) was explicitly mentioned in 79.8 % (n=438) of articles.

3.2. Analytic samples

A table describing the study sample was included in 82.5 % of articles (n=453). The sample size of the final analytic model was mentioned in 84 % of articles (n=461), and 59.6 % of articles (n=327) mentioned the number of participants excluded from the final analyses due to missing data and/or exclusion criteria. The characteristics of the excluded sample were detailed in 42.8 % of papers that reported missingness in the data. Quantification of missing data for individual variables (e.g., the percent missing household income data) was included in 37.5 % (n=206) of articles. The missing data mechanism (e.g., missing completely at random) was discussed in 16.4 % (n=90) of articles. Imputation methods were used to account for missing data in 24.8 % (n=136) of articles, and in 36.9 % (n=76) of papers that mentioned missing data for individual variables. Imputation methods were also used in 61.1 % (n=55) of papers that examined the missingness
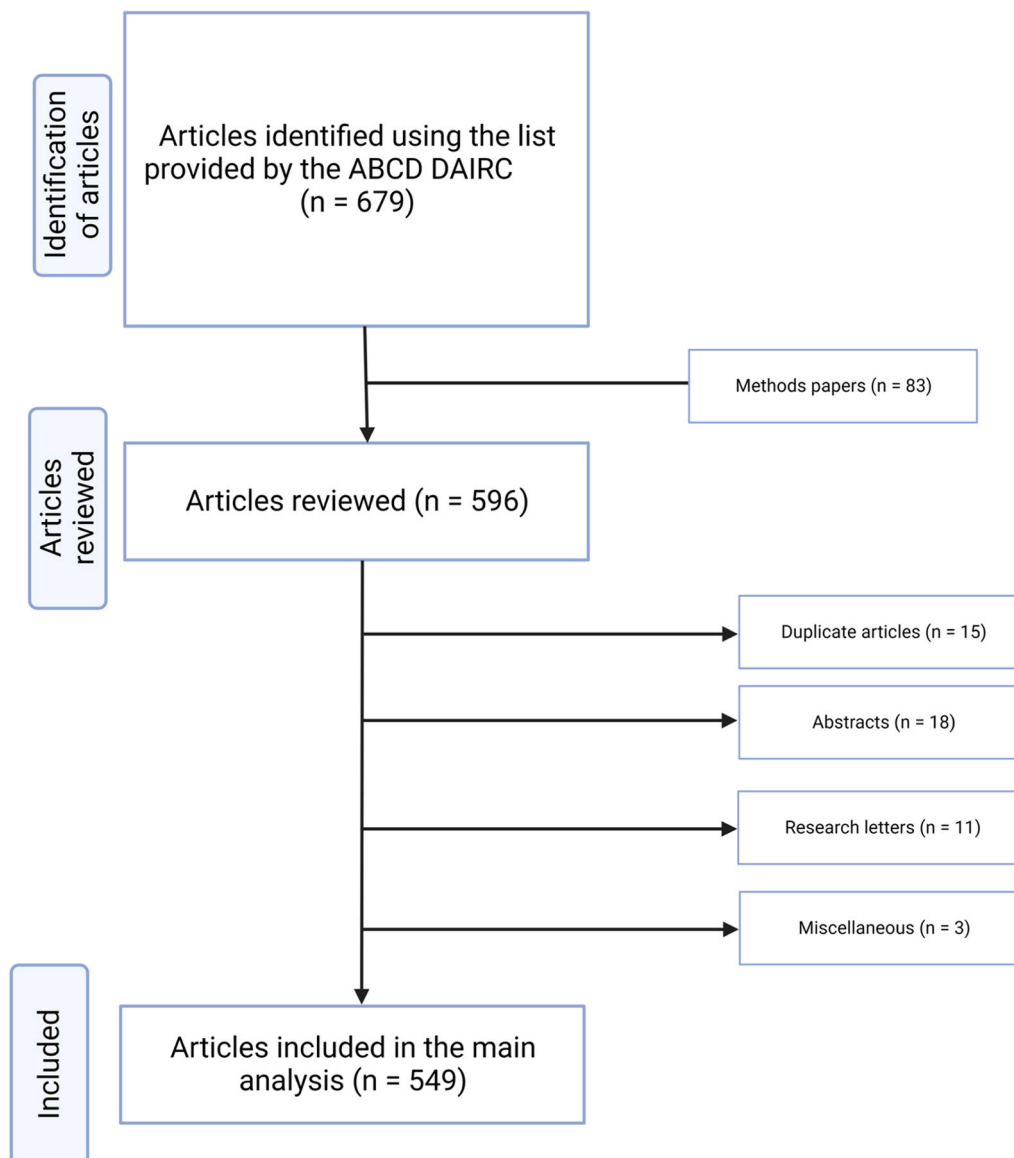

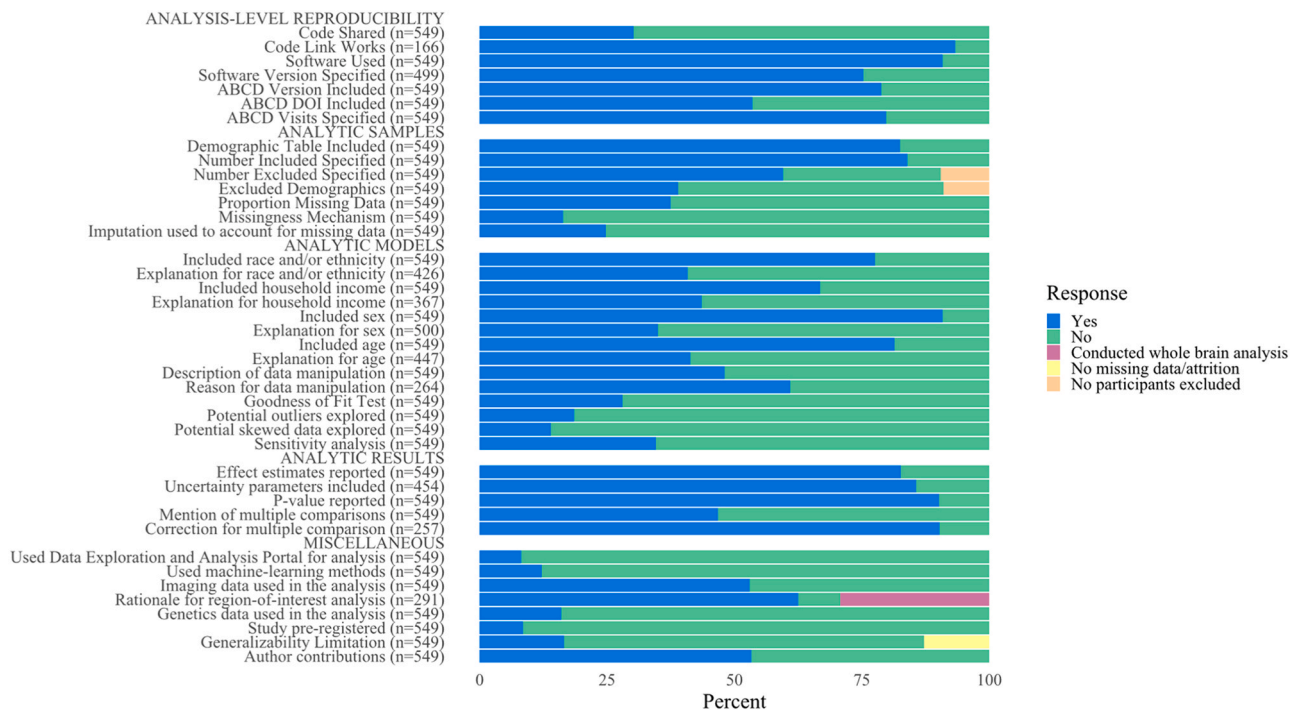**Fig. 1.** Flow chart summary of the article screening process.

Fig. 2. The percentage of response types for each data extraction item.

mechanisms (e.g., missing completely at random).

3.3. Analytic models

Race and/or ethnicity was included in 77.6 % of papers (n=426), and its inclusion was explained in 40.8 % of these articles (n=174). Household income was included in 66.8 % of papers (n=367), and its inclusion was explained in 43.6 % of these articles (n=160). Participant sex was included in 90.9 % (n=499) of reviewed articles, and its inclusion was explained in 35 % of these articles (n=175). Participant age was included and explained in 81.4 % (n=447) and 41.4 % (n=185) of articles, respectively.

Data manipulation (e.g., changing a variable from continuous to categorical) was mentioned in 48.1 % (n=264) of reviewed articles, and was explained in 61 % of these articles (n=161). Effect estimates (e.g., a beta coefficient or Cohen's d) were reported in 82.7 % (n=454) of articles included in the main analysis. A quantification of uncertainty (e.g., confidence intervals, standard errors) was included in 85.7 % (n=389) of articles that reported an effect estimate. A *p*-value was reported in 90.2 % (n=495) of articles included in the main analysis.

Multiple comparisons were mentioned in 46.8 % of articles (n=257), and correction for multiple comparisons was detailed in 90 % (n=232) of these articles. A sensitivity analysis was conducted in 34.6 % (n=190) of articles included in the main analysis. A goodness of fit test (e.g., Akaike Information Criterion) to select the final model was mentioned in 28.1 % (n=154) of articles. Potential outliers and skewed data were discussed in 18.6 % (n=102) and 14 % (n=77) of articles, respectively.

3.4. Miscellaneous

Imaging data were used in 53 % (n=291) of articles included in the main analysis, and 92 % (n=267) of these articles included a rationale for selecting certain regions of interest. Genetics data were analyzed in 16 % (n=89) of articles. Imaging and genetics data were included in 9.1 % of articles (n=50). Machine-learning methods were used in 12.2 % of articles included in the main analysis (n=67), and 76.1 % (n=51) of these also used imaging data. The Data Exploration and

Analysis Portal (DEAP) was used to analyze ABCD Study data in 8.2 % (n=45) of articles included in the main analysis. A statement about limitations to generalizability due to missing data was included in 28.6 % of articles that reported at least some missing data (n=69). Author contributions were included in 53.4 % of articles (n=293). Study preregistration was mentioned in 8.6 % (n=47) of articles included in the main analysis.

3.5. Level of completeness score

The Level of Completeness score ranged from 1 to 23 in full-length articles(Fig. 3). The mean score was 13.11 (median = 13.0, standard deviation [SD] = 3.92, variance = 15.4). The Level of Completeness score in abstracts (n=18) ranged from 0 to 13 (mean = 3.7, SD = 2.9). The range of scores in research letters (n=11) was 0–12 (mean = 8.64, SD=3.2).

Results of the univariate linear regression with individual data extraction items emphasize the importance of missing data (Table 1). Specifying the total number of participants excluded due to missing data was significantly correlated with Level of Completeness scores (β = 4.17, 95 % Confidence Interval [CI]: 3.6, 4.8; *p*-value <0.0001). A description of missingness for individual variables (e.g., household income) was positively correlated with Completeness scores (β = 4.1, 95 % CI: 3.5, 4.7; *p*-value = <0.0001). Mention of the ABCD data version and ABCD Study visits was significantly correlated with higher Completeness scores (β = 3.8 and β = 3.4, respectively). Using the DEAP was correlated with a decreased Completeness score (β = −1.9, 95 % CI: −3.1, −0.7; *p*-value = 0.017). There was no significant difference in Completeness scores in studies that used machine learning methods (*p*-value = 0.38). The year of publication was positively correlated with Level of Completeness score (β=0.83, 95 % CI: 0.55, 1.11; *p*-value <0.0001). In other words, more recent publications had higher Completeness scores. Results of all extraction items can be found in the supplementary materials (Table S5).
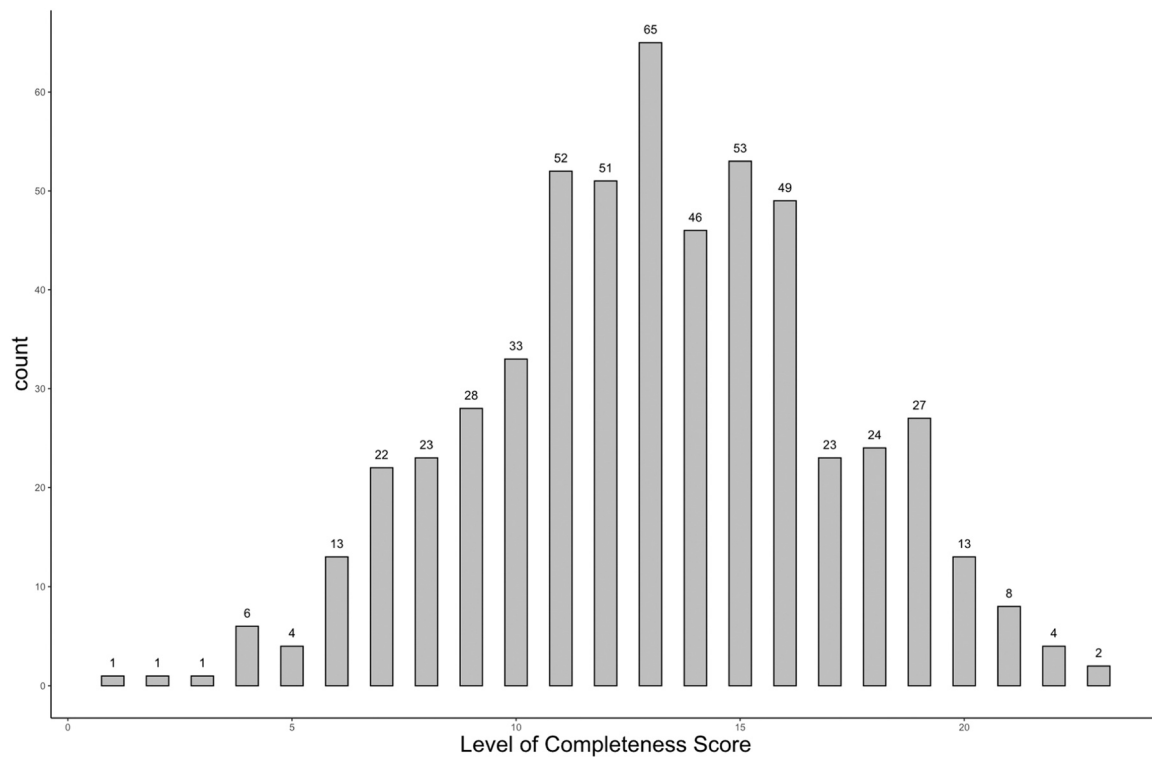
Fig. 3. The frequency and distribution of Level of Completeness score in the articles included in the main analysis.

Table 1

Univariate linear regression results with extraction items and Level of Completeness score.

	Level of Completeness Score	
	β (95 % CI)	*p*-value
ITEM		
Imaging Data Used (Ref = No)		
Yes	0.6 (-0.06, 1.25)	0.075
Genetics Data Used (Ref = No)		
Yes	0.52 (-0.38, 1.41)	0.26
Machine learning methods (Ref = No)		
Yes	-0.45 (-1.45, 0.55)	0.38
N Excluded (Ref = No)		
No participants excluded	-0.61 (-1.6, 0.42)	0.24
Yes	4.2 (3.6, 4.8)	2×10^{-16}
Proportion Missing Data (Ref = No)		
Yes	4.1 (3.5, 4.7)	2×10^{-16}
ABCD Version (Ref = No)		
Yes	3.8 (3.1, 4.6)	2×10^{-16}
ABCD Visits (Ref = No)		
Yes	3.4 (2.6, 4.2)	2×10^{-16}
DEAP Used (Ref = No)		
Yes	-1.9 (-3.1, −0.72)	0.002
Behavioral only (Ref=No)		
Yes	-0.14 (-1.3, 1.0)	0.81
Year of Publication	0.83 (0.55, 1.11)	9.5×10^{-9}

Note: Model results are from a univariate linear regression analysis

3.6. Results of the sensitivity analysis

Findings from the sensitivity analysis were like those found using univariate linear regression. Mentioning the number of participants excluded from the final analytic model (i.e., item #8) was the most important predictor of Level of Completeness score (Fig. 4). The following items were considered nonimportant contributors of Completeness score: Machine Learning and Author Contributions. The full Boruta attribute statistics can be found in Table S6.

3.7. Post hoc analysis

The post hoc analysis examined the relationship between Level of Completeness score and the logged journal impact factor (Figure S3). The Pearson correlation coefficient between the Level of Completeness score and logged journal impact factor was 0.43. There was 1 article in the top 10 %, 5 articles in the middle 10 %, and 4 articles in the bottom 10 % of Completeness scores that were in journals without an impact factor. The mean 2-year journal impact factor of the included articles (n=158) was 6.9 (range = 0.5–82.9). The mean impact factor for the bottom 10 %, middle 10 %, and top 10 % of Completeness scores was 3.9, 7.2, and 9.5, respectively. A bottom 10 % Completeness score was significantly correlated with a 0.77-point lower logged impact factor (*p*-value < 0.0001) when compared to articles in the top 10 % of Completeness scores (Table S6). There was no significant difference in the logged impact factor of manuscripts in the top 10 % and middle 10 % of Completeness scores (*p*-value=0.54). Completeness score was significantly correlated with the logged impact factor in models using Completeness as a continuous measure (β=0.056, 95 % CI: 0.04, 0.07; *p*-value<0.0001). In other words, a one-point increase in the Completeness score was correlated with a 0.06-point increase in the logged impact factor.

4. Discussion

The findings highlight shortcomings and optimistic reporting trends in the level of analytic information included in ABCD Study publications, with implications for transparency and reproducibility. Analytic scripts were shared in fewer than a third of publications included in the main analysis, and this proportion increased to 44 % of papers published in 2023. Issues related to missing data were intermittently addressed. The sociodemographic differences of participants included and excluded from the analysis were provided in fewer than half of publications. Descriptions of missingness for individual variables were included in around one-third of articles overall. The reasons for inclusion of socio-demographic variables (e.g., race and/or ethnicity, household income)
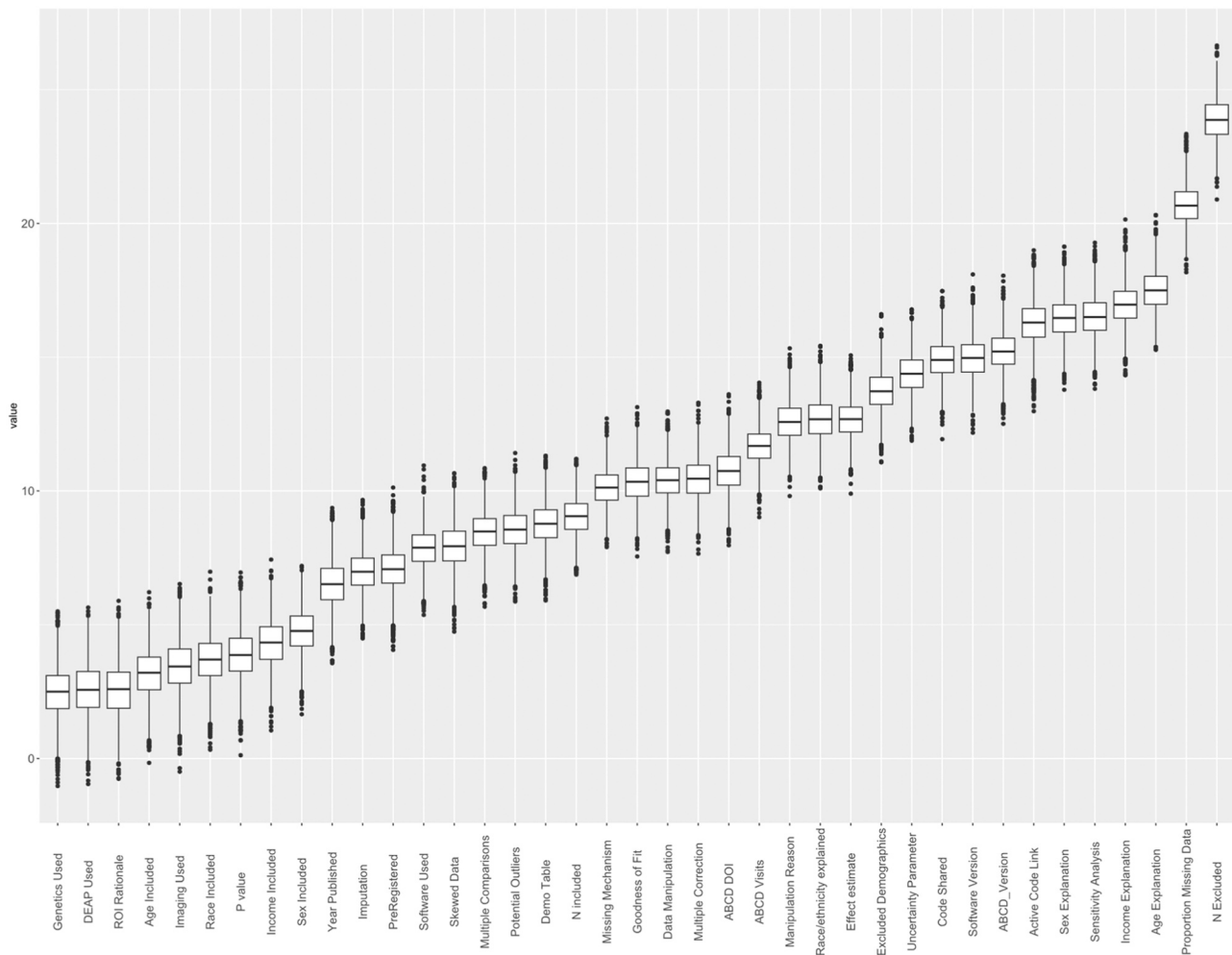
1

Fig. 4. Results of the machine learning feature selection method.

was explained in fewer than half of the papers. Around 20 % of articles mentioned problems with generalizability of results due to nonrandom missingness (e.g., disproportionate missing data/attrition in minority groups). Univariate regression models and machine learning methods highlighted missingness items as an important contributor of Level of Completeness scores. Finally, the post hoc analysis found a significant positive correlation between Level of Completeness scores and journal impact factor. Together, these findings highlight existing gaps in the statistical rigor and reporting of studies using ABCD Study data but indicate that some of these gaps are already being closed by improved reporting practices in recent years. Based on the study findings, we have provided recommendations (summarized in Table 2) for best reporting practices to improve transparency and reproducibility for future full-length ABCD Study articles. The recommendations are tailored to the ABCD Study but can be applied to other observational data in the field of developmental neuroscience.

4.1. Analysis-level reproducibility

Analytic scripts that detail how the data were processed and manipulated can create support for scientific claims (Baker, 2016b). Code sharing reluctance may be due in part to a lack of awareness. A survey by Stuart et al. of over 7700 researchers found that 33 % did not know where to deposit data and code, and 46 % were unsure how to usefully organize the data (Stuart et al., 2018). Reproducibility issues are heightened by the lack of code sharing. A case study comparing reproducibility before and after a journal policy change reported a 40 %

increase in the probability of reproducing results when analytic code was made available (Laurinavichyute et al., 2022). In that study, only 37 % of papers that provided data, but not code, were reproducible compared to 81 % of papers that included analytic scripts (Laurinavichyute et al., 2022). Altogether, this highlights the need for greater adoption of code sharing practices in the field of developmental neuroscience to improve transparency and reproducibility efforts.

However, the current study found that publications from 2023 (43.3 % shared code compared to 29 % in 2022 and 30 % in 2021) were more likely to share analytic scripts and highlighted the recent emphasis on improving transparency and reproducibility efforts. The ABCD Study affords an incredible opportunity to improve current code sharing practices as it provides resources for researchers interested in analyzing the ABCD dataset. For example, the ABCD-ReproNim (http://www.abcd-repronim.org) course was designed to improve reproducibility efforts and is still available and free to use at a self-guided pace. Additionally, publicly available analytic scripts on platforms like GitHub can be used by any researcher to determine which variables, methods, and assumptions were made while working with the open source ABCD data. Code sharing can also be used to reanalyze the data with alternative methods that can lead to new insights. The burden of providing well-documented code may deter some researchers from sharing analytic scripts, although even poorly documented code can be beneficial to improving reproducibility efforts in science (Gorgolewski and Poldrack, 2016; Barnes, 2010). Future ABCD data use agreements should consider requiring code sharing on public platforms (e.g., GitHub, Open Science Framework) to normalize transparency efforts. Based on our findings,

Table 2
List of practices and recommendations to improve transparency and reproducibility in ABCD Study publications.

Practice	Recommendations
1. Code sharing	Make analytic scripts publicly available on an open platform (e.g., GitHub, OSF). The scripts should include enough information to reproduce the findings of the study. Add detailed comments to each step of the code to improve interpretability.
2. Software Used	Clearly state the software and software version that was used for all steps of the analysis.
3. ABCD Version	State which ABCD data version is being used in the analysis (e.g., ABCD version 4.0).
4. ABCD Visits	State the ABCD Study visits that were included in the analysis.
5. Demographics Table	Include a table in the main paper or supplementary materials detailing the characteristics of the analytic sample.
6. Included Sample Size	State the final sample size of your analytic model after any exclusions.
7. Excluded Sample Size	State the number of participants that were excluded.
8. Excluded Demographics	Include a table in the main paper or supplementary materials detailing the characteristics of the excluded sample.
9. Missing data - covariates	Describe the amount of missing data in the variables included in your analysis.
10. Missing data - attrition	Describe the amount of missing data due to a participant not attending a study visit.
11. Covariate explanation	Provide a reason for including a covariate in the model for adjustment. The reason should include support for why it was necessary (e.g., a citation of previous research or meets the definition of a confounder).
12. Data Manipulation	Describe and include a reason for variable manipulation. For example, justify why a continuous measure was categorized, and why certain cut-offs were used.
13. Effect estimates	Include a measure of association along with quantification of uncertainty (e.g., confidence intervals, standard errors).
14. Sensitivity Analysis	Test whether the results are robust to any statistical assumptions. Some common assumptions include how a variable is defined or that the results are robust to unmeasured confounding.
15. Skewed Data	Examine skewness in the variables of interest to determine whether it meets model assumptions. Explore alternatives if model assumptions are not met (e.g., generalized linear models, outcome transformation).
16. Potential Outliers	Examine any potential outlying data points in your covariates. Decide whether these were due to measurement error or natural variability in the measurement.
17. Rationale for ROIs	Provide the rationale and method for selecting the particular regions of interest when analyzing imaging data. Explicitly detail any additional steps taken when analyzing the imaging data (e.g., regressing out motion parameters, adjustment for scanner).
18. Study preregistration	Preregister the study hypothesis and analytic plans prior to analyzing the ABCD data. Preregistration can help guard against publication bias and P-hacking. Preregistration can be done on open platforms like the Open Science Framework (OSF). Include a link of the preregistration in your manuscript.
19. Limitations	Describe any caveats to the findings. For example, that the generalizability of the findings may be impacted by nonrandom attrition.

we strongly recommend widespread sharing of analytic scripts that can be used to reproduce ABCD Study findings. The analytic scripts should, at minimum, include publicly accessible code that can recreate research results (e.g., statistical models). We highly encourage the inclusion of comments throughout the analytic script to improve interpretability. In addition to sharing code, we encourage researchers to share version of ABCD data and the visits included in the analyses as well as the software and version used for all steps of the analysis.

4.2. Analytic samples

Most papers using ABCD data included a description of the study sample. The high percentage is in line with previous findings of partic-ipant characteristics reporting in neuroimaging studies (Sterling et al., 2022). In contrast, issues related to missing data were not regularly addressed in ABCD papers. Missing data can introduce bias and weaken the generalizability of findings when the missingness is nonrandom (Dong and Peng, 2013; Lash et al., 2020). These data and previous

studies exploring missingness in research suggests that the rate of missingness reporting in papers using ABCD Study data is in line with that of the broader literature. A study of psychology papers (n=113) published in 2012 found that 56 % of papers mentioned missing data, and 6.7 % explored how missingness was related to other variables (Little et al., 2014). A review of cohort studies (n=82) found that 43 % reported the amount of missing data at follow-up visits, and that 32 % compared the characteristics of participants and nonparticipants (Karahalios et al., 2012). A separate review of developmental psychology papers (n=100) found that only 43 % of studies that mentioned missing data compared the two groups (Jeličić et al., 2009).

Missing data in the ABCD Study can arise from selected exclusion criteria (e.g., failed imaging quality control, refusal to answer) or from attrition (e.g., loss to follow-up) (Ewing et al., 2022). In the case of ABCD imaging data, exclusion criteria can result in nonrandom missingness that is highly correlated with sociodemographic variables. For example, more than 50 % of Black children that completed the EN-back and Stop Signal fMRI task during the baseline ABCD visit were excluded after data processing (Chaarani et al., 2021). The current study also found that techniques for dealing with missing data (e.g., multiple imputation) were used infrequently. The decision to use imputation methods depends on many different factors related to the study (e.g., the amount of missing data, the missingness mechanism). Imputation may be necessary when there is nonrandom or systematic missingness of data (Kleinke et al., 2020). Caution is warranted when considering imputation since it does not always reduce bias (Twisk et al., 2013). Altogether, there is considerable evidence that missing data is not being sufficiently addressed in papers using ABCD Study data. Future studies using the ABCD data should examine patterns of missingness and make a greater effort to describe the population excluded from the analytic sample. We recommend that data users state the final sample size after exclusions, include the number of participants excluded, and include a table that details the characteristics of the excluded sample. Researchers should also examine how missing data is correlated with their outcome variable and/or other covariates, since this may be an indicator of whether the missingness is ignorable (Twisk, 2013). For example, a researcher conducting a longitudinal analysis using the Child Behavior Checklist – Externalizing subscale as their outcome of interest can describe participants with and without a missing ABCD time point. A significant difference in the characteristics of the two groups can be an indicator of selection bias and should be mentioned as a potential limit to generalizability.

4.3. Analytic models

The current study found that sociodemographic variables were commonly included in statistical models without explanation. This finding is in line with previous reports from other studies. A study using randomly selected articles (n=60) from psychology journals found that 18.3 % of articles failed to provide an explanation for any control variables, and 53.3 % for at least one variable (Becker, 2005). A study of Management Research journals found that 19 % of publications (n=162) did not provide full justification for variable control (Carlson and Wu, 2011). Another study of Management Research publications (n=812) from 2005 to 2009 found that 18.2 % of articles provided no rationale for variable inclusion (Atinc et al., 2012).

The current study looked for any explanation for variable inclusion (e.g., a confounder, a citation supporting the inclusion). There are statistical and social reasons for justifying the inclusion of sociodemographic variables. Controlling for certain variables (e.g., a mediator) can inadvertently result in biased effect estimates and may even induce spurious associations (e.g., adjustment for a collider) (Schisterman et al., 2009; Wysocki et al., 2022). More thoughtful consideration of variable relationships (e.g., using directed acyclic graphs) can help determine whether statistical adjustment is necessary (VanderWeele and Robinson, 2014). The habitual inclusion of sociodemographic variables (e.g., race

and/or ethnicity) in statistical models has also been criticized for perpetuating stigmatization and inequity (Cardenas-Iniguez and Gonzalez, 2023). We strongly encourage that researchers provide support for the inclusion of all covariates in their statistical models. This practice can alleviate concerns over biased estimates due to improper statistical adjustment (e.g., conditioning on a collider). In particular, the use of any race and/or ethnicity variable should be thoroughly explained and justified. Researchers should determine whether inclusion, omission, or some other method (e.g., effect measure modification) is necessary when considering the use of race or ethno-racial variables. Researchers should consult recommended best practices (VanderWeele and Robinson, 2014; Cardenas-Iniguez and Gonzalez, 2023; Martinez et al., 2023) prior to using race and/or ethnicity variables in their analyses.

Finally, the ABCD dataset includes hundreds of socioeconomic and environmental variables that may offer superior explanatory power than sociodemographic variables. We encourage exploration of the ABCD data dictionary (https://data-dict.abcdstudy.org/) to determine whether there are variables that are more direct measures of differences in the study sample.

4.4. Strengths

The study had several strengths. First, we provided a comprehensive summary of reporting practices across papers using ABCD Study data. Since the ABCD Study is ongoing, the findings will hopefully influence the inclusion of important methodological and analytic information in future publications using the data, increasing transparency and reproducibility. Second, we have provided the full list of items as a resource for researchers to incorporate into their own manuscripts. A list of practices and recommendations is available (Table 2 or at OSF: https://osf.io/8wqft) and will ideally improve adherence to recommended best practices when using ABCD data and can also be applied more broadly to the field of developmental neuroscience. Third, we examined variable importance and quantified the importance of certain practices (e.g., exploration of missing data) to overall methodological and reporting rigor. We are hopeful that future work will investigate patterns of missingness in the ABCD Study dataset.

4.5. Limitations

There were several limitations in the study. First, the extraction items were created for use with the ABCD Study data and may not generalize outside the study, although we expect that certain items and the questions motivating their inclusion will have broad applicability. Second, the extraction items were heavily influenced by neuroimaging best practices recommendations and may not have captured important measures for other fields (e.g., genetics). Third, the large number of reviewers and papers makes it difficult to feasibly address any potential measurement error or subjectivity in reviews, and the potential impact thereof on findings reported here. We cannot rule out that errors were made in the extraction of data from each article. We expect that the large sample of papers that were included in the analysis minimized the influence that measurement error had on our findings. Fourth, the absence of underlying groupings in the extraction items meant that items were summed and treated equally for the Completeness score. We expect that certain items (e.g., code sharing) are more valuable contributors to transparency and reproducibility efforts.

4.6. Conclusion

The findings highlight glaring transparency issues in studies using ABCD data. The ABCD Study will influence many aspects of research involving adolescent development and ensuring that findings are reproducible is highly feasible and crucial to moving the field forward. We hope that our review of ABCD use of data sheds light on the many opportunities for improvement in research practices and strengthens

reproducibility efforts going forward.

CRediT authorship contribution statement

Paola Badilla: Investigation. **Katherine L. Bottenhorn:** Writing – review & editing, Investigation. **Laila Tally:** Investigation. **Ellen Mukwekwerere:** Investigation. **Carlos Cardenas-Iniguez:** Writing – review & editing, Investigation. **Shana Adise:** Writing – review & editing, Investigation. **Punitha Subramaniam:** Writing – review & editing, Investigation. **Arturo Lopez-Flores:** Investigation. **Joseph R. Boughter:** Investigation. **Lyndsey Dorholt:** Investigation. **Emma R. Skoler:** Investigation. **Isabelle L. Bedichek:** Investigation. **Omoengheme Ahanmisi:** Investigation. **Elizabeth Robertson:** Investigation. **Amanda K. Adames:** Investigation. **Serena D. Matera:** Investigation. **Rebekah S. Huber:** Writing – review & editing, Investigation, Conceptualization. **Bonnie J. Nagel:** Writing – review & editing. **Owen Winters:** Investigation. **Nicholas Sissons:** Investigation. **Elizabeth Nakiyingi:** Investigation. **Anya Harkness:** Writing – review & editing, Investigation. **Gabriela Mercedes Perez-Tamayo:** Investigation. **Daniel A. Lopez:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Anthony R. Hill:** Investigation. **Zhuoran Xiang:** Investigation. **Jennell Encizo:** Investigation. **Allison N. Smith:** Investigation. **Isabelle G. Wilson:** Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset is publicly available and can be downloaded at https://osf.io/qkefw/.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2024.101408.

References

Andrade, C., 2021. HARKing, cherry-picking, p-hacking, fishing expeditions, and data dredging and mining as questionable research practices. J. Clin. Psychiatry 82 (1), 25941.

Atinc, G., Simmering, M.J., Kroll, M.J., 2012. Control variable use and reporting in macro and micro management research. Organ. Res. Methods 15 (1), 57–74.

Baker, M., 2016a. 1,500 scientists lift the lid on reproducibility. Nature 533 (7604), 452–454.

Baker, M., 2016b. Why scientists must share their research code. Nature.

Barnes, N., 2010. Publish your computer code: it is good enough. Nature 467 (7317), 753.

Becker, T.E., 2005. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. Organ. Res. Methods 8 (3), 274–289.

Bethlehem, R.A.I., Seidlitz, J., White, S.R., et al., 2022. Brain charts for the human lifespan. Nature 604 (7906), 525–533.

Cardenas-Iniguez, C., Gonzalez, M.R., 2023. "We controlled for race and ethnicity…". Consid. Use Commun. race Ethn. neuroimaging Res.

Carlson, K.D., Wu, J., 2011. The Illusion of Statistical Control: Control Variable Practice in Management Research. Organ. Res. Methods 15 (3), 413–435.

Chaarani, B., Hahn, S., Allgaier, N., et al., 2021. Baseline brain function in the preadolescents of the ABCD Study. Nat. Neurosci. 24 (8), 1176–1186.

Collaboration, O.S., 2015. Estimating the reproducibility of psychological science. Science 349 (6251).

Corporation M. Microsoft Excel. 2018.

Dong, Y., Peng, C.Y., 2013. Principled missing data methods for researchers. Springerplus 2 (1), 222.

Errington, T.M., Mathur, M., Soderberg, C.K., et al., 2021a. Investigating the replicability of preclinical cancer biology. Elife 10, e71601.

Errington, T.M., Denis, A., Perfito, N., et al., 2021b. Challenges for assessing replicability in preclinical cancer biology. eLife 10, e67995.

Ewing, S.W.F., Dash, G.F., Thompson, W.K., et al., 2022. Measuring retention within the adolescent brain cognitive development (ABCD) SM study. Dev. Cogn. Neurosci. 54, 101081.

Fox, J., Weisberg, S., Adler, D., et al., 2012. Package 'car'. Vienna.: R. Found. Stat. Comput. 16 (332), 333.

Garavan, H., Bartsch, H., Conway, K., et al., 2018. Recruiting the ABCD sample: Design considerations and procedures. Dev. Cogn. Neurosci. 32, 16–22.

Gibson, M.J., Spiga, F., Campbell, A., et al., 2023. Reporting and methodological quality of studies that use Mendelian randomisation in UK Biobank: a meta-epidemiological study. BMJ Evid. -Based Med. 28 (2), 103–110.

Gilmore, R.O., Diaz, M.T., Wyble, B.A., et al., 2017. Progress toward openness, transparency, and reproducibility in cognitive neuroscience. Ann. N. Y. Acad. Sci. 1396 (1), 5–18.

Gorgolewski, K.J., Poldrack, R.A., 2016. A practical guide for improving transparency and reproducibility in neuroimaging research. PLoS Biol. 14 (7), e1002506.

Greenland, S., Pearce, N., 2015. Statistical foundations for model-based adjustments. Annu. Rev. Public Health 36, 89–108.

Heeringa, S.G., Berglund, P.A., 2020. A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. BioRxiv.

Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124.

Jeličić, H., Phelps, E., Lerner, R.M., 2009. Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. Dev. Psychol. 45 (4), 1195–1199.

Joober, R., Schmitz, N., Annable, L., et al., 2012. Publication bias: what are the challenges and can they be overcome? J. Psychiatry Neurosci. 37 (3), 149–152.

Karahalios, A., Baglietto, L., Carlin, J.B., et al., 2012. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Med. Res. Methodol. 12 (1), 96.

Klapwijk, E.T., van den Bos, W., Tamnes, C.K., et al., 2021. Opportunities for increased reproducibility and replicability of developmental neuroimaging. Dev. Cogn. Neurosci. 47, 100902.

Kleinke, K., Reinecke, J., Salfrán, D., et al., 2020. Applied Multiple Imputation: Advantages, Pitfalls, New Developments and Applications in R. Springer Nature.

Korbmacher, M., Azevedo, F., Pennington, C.R., et al., 2023. The replication crisis has led to positive structural, procedural, and community changes. Commun. Psychol. 1 (1), 3.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Lash, T.L., VanderWeele, T.J., Haneause, S., et al., 2020. Modern Epidemiology. Lippincott Williams & Wilkins.

Laurinavichyute, A., Yadav, H., Vasishth, S., 2022. Share the code, not just the data: A case study of the reproducibility of articles published in the Journal of Memory and Language under the open data policy. J. Mem. Lang. 125, 104332.

Little, T.D., Jorgensen, T.D., Lang, K.M., et al., 2014. On the joys of missing data. J. Pediatr. Psychol. 39 (2), 151–162.

Martinez, R.A.M., Andrabi, N., Goodwin, A.N., et al., 2023. Conceptualization, operationalization, and utilization of race and ethnicity in major epidemiology journals, 1995–2018: A systematic review. Am. J. Epidemiol. 192 (3), 483–496.

Munafò, M.R., Nosek, B.A., Bishop, D.V., et al., 2017. A manifesto for reproducible science. Nat. Hum. Behav. 1 (1), 1–9.

Muthén, B., Muthén, L., Mplus, 2017. Handbook of Item Response Theory. Chapman and Hall/CRC, pp. 507–518.

Nichols, T.E., Das, S., Eickhoff, S.B., et al., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20 (3), 299–303.

Nuzzo, R., 2014. Statistical errors. Nature 506 (7487), 150.

Parsons, S., Azevedo, F., Elsherif, M.M., et al., 2022. A community-sourced glossary of open scholarship terms. Nat. Hum. Behav. 6 (3), 312–318.

Schisterman, E.F., Cole, S.R., Platt, R.W., 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiol. (Camb., Mass) 20 (4), 488.

Sciences NAo, Policy, Affairs, G., et al., 2019. Reproducibility and Replicability in Science. National Academies Press.

Sterling, E., Pearl, H., Liu, Z., et al., 2022. Demographic reporting across a decade of neuroimaging: a systematic review. Brain Imaging Behav. 16 (6), 2785–2796.

Stuart, D., Baynes, G., Hrynaszkiewicz, I., et al., 2018. Pract. Chall. Res. data Shar.

Team R. RStudio: integrated development environment for R. 2015. 2019.

Team RC. R: A language and environment for statistical computing. 2013.

Twisk, J., de Boer, M., de Vente, W., et al., 2013. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J. Clin. Epidemiol. 66 (9), 1022–1028.

Twisk, J.W., 2013. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. cambridge university press.

VanderWeele, T.J., Robinson, W.R., 2014. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology 25 (4), 473–484.

Volkow, N.D., Koob, G.F., Croyle, R.T., et al., 2018. The conception of the ABCD study: From substance use to a broad NIH collaboration. Dev. Cogn. Neurosci. 32, 4–7.

Wickham, H., 2011. ggplot2. Wiley Interdiscip. Rev.: Comput. Stat. 3 (2), 180–185.

Wickham, H., François, R., Henry, L., et al., 2019. Package 'dplyr'. A Gramm. Data Manip. R. Package Version 8.

Wysocki, A.C., Lawson, K.M., Rhemtulla, M., 2022. Statistical Control Requires Causal Justification. Adv. Methods Pract. Psychol. Sci. 5 (2), 25152459221095823.