RESEARCH ARTICLE

REVISED An assessment of the autism neuroimaging literature for the prospects of re-executability [version 2; peer review: 2 approved]

Steven M. Hodge[1], Christian Haselgrove[1], Leah Honor[2], David N. Kennedy[1], Jean A. Frazier[1]

[1]Eunice Kennedy Shriver Center, Department of Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts, 01655, USA
[2]Lamar Soutter Library, University of Massachusetts Medical School, Worcester, Massachusetts, 01655, USA

v2 **First published:** 24 Aug 2020, **9**:1031
https://doi.org/10.12688/f1000research.25306.1
Latest published: 04 Mar 2021, **9**:1031
https://doi.org/10.12688/f1000research.25306.2

Abstract

Background: The degree of reproducibility of the neuroimaging literature in psychiatric application areas has been called into question and the issues that relate to this reproducibility are extremely complex. Some of these complexities have to do with the underlying biology of the disorders that we study and others arise due to the technology we apply to the analysis of the data we collect. Ultimately, the observations we make get communicated to the rest of the community through publications in the scientific literature.
Methods: We sought to perform a 're-executability survey' to evaluate the recent neuroimaging literature with an eye toward seeing if the technical aspects of our publication practices are helping or hindering the overall quest for a more reproducible understanding of brain development and aging. The topic areas examined include availability of the data, the precision of the imaging method description and the reporting of the statistical analytic approach, and the availability of the complete results. We applied the survey to 50 publications in the autism neuroimaging literature that were published between September 16, 2017 to October 1, 2018.
Results: The results of the survey indicate that for the literature examined, data that is not already part of a public repository is rarely available, software tools are usually named but versions and operating system are not, it is expected that reasonably skilled analysts could approximately perform the analyses described, and the complete results of the studies are rarely available.
Conclusions: We have identified that there is ample room for improvement in research publication practices. We hope exposing these issues in the retrospective literature can provide guidance and motivation for improving this aspect of our reporting practices in the future.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 2 (revision) 04 Mar 2021	✓ report	
version 1 24 Aug 2020	? report	✓ report

1. **Karsten Specht**, University of Bergen, Bergen, Norway

2. **Travis Riddle**, National Institutes of Health, Bethesda, USA
Adam Thomas, National Institutes of Health, Bethesda, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Reproducible Science, Neuroimaging, Autism, Data Sharing, Re-executability

Corresponding author: David N. Kennedy (David.Kennedy@umassmed.edu)

Author roles: Hodge SM: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Haselgrove C**: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Honor L**: Conceptualization, Investigation, Methodology, Writing – Review & Editing; **Kennedy DN**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Frazier JA**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the National Institute of Mental Health [R01 MH083320; for CANDIShare] and the National Institute of Biomedical Imaging and Bioengineering [P41 EB019936; for ReproNim].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Hodge SM, Haselgrove C, Honor L *et al.* **An assessment of the autism neuroimaging literature for the prospects of re-executability [version 2; peer review: 2 approved]** F1000Research 2021, **9**:1031
https://doi.org/10.12688/f1000research.25306.2

First published: 24 Aug 2020, **9**:1031 https://doi.org/10.12688/f1000research.25306.1

REVISED **Amendments from Version 1**

In this version we made a number of important upgrades in response to the reviewers comments. First we clarify the scope of this specific survey. We try to do a better job of setting up the scope of the survey in the introduction. Specifically, we note that there are at least three domains in a publication where sufficient information for re-execution needs to be considered: the subject selection (can another researcher generate a comparable group); the data acquisition (can another researcher collect the same data); and the analysis (can another researcher perform the same analysis). All of these areas are important but we are only addressing the 'analysis' aspect in this manuscript. Second, we indicate in Table 1 the 'type' of MRI imaging (structural, functional, etc.) and which reviewers were involved. Third, we make the distinction that our assessment is aimed at evaluating the quality of the reporting (can I do what was reported), rather than the content (is what was reported the right or best thing to do?). Fourth, we have clarified the text regarding the validation cases (pilot assessment) and the dual raters for consensus evaluation of each publication. Fifth, we have tried to clarify the meaning and ways that the 'complete results availability' can be satisfied in the Discussion. Finally, the Results, Discussion and Conclusions sections have been updated to better reflect the appropriate content.

Any further responses from the reviewers can be found at the end of the article

Introduction

There is concern about the status of reproducibility in science in general and neuroimaging neuroscience in particular (Button *et al.*, 2013; Gorgolewski & Poldrack, 2016). A particularly germane concern was expressed by Kapur and colleagues in lamenting: "a profusion of statistically significant, but minimally differentiating, biological findings; 'approximate replications' of these findings in a way that neither confirms nor refutes them" (Kapur *et al.*, 2012). The replication of a specific finding (or reproducibility of a specific analysis), as reflected in a publication, has many details and nuances to it (Kennedy *et al.*, 2019). Often, we are searching for the 'generalizability' of a finding: does the finding hold true when using 'similar' data and a 'similar' analysis. The similarity of data (or analysis) is a fuzzy concept. One could have a population with the same number of subjects with the same diagnosis, having the same mean age and same gender distribution as a target population; however, if the diagnosis in question is a 'spectrum'-diagnosis (for example, autism, schizophrenia, depression, etc.), despite the 'sameness' of my sample in the aforementioned categories, the detailed nature of the characteristics of my sample in the features of the diagnosis itself can still be quite variable. At the level of a biological finding, we typically do not predicate the finding on an exact acquisition protocol, or a specific analysis protocol; rather, it is implicit in our finding that it should hold for other valid acquisitions and analyses of the reported types. There is increasing evidence that this implicit assumption of similarity, when it relates to the specific details of acquisition or analysis, does not necessarily hold (Glatard *et al.*, 2015).

Some have argued that the starting point for the structured exploration of the generalizability of a specific finding (and thus a cornerstone to the quest for reproducibility) lies in the original finding itself being re-executable (Ghosh *et al.*, 2017; Kennedy, 2019). Starting from the re-execution of a finding will allow for the systematic exploration of the generalizability of that finding, over changes in data and analysis. To date, when new studies find different findings from prior studies, it is too easy to simply argue that differences in the subject population or analysis workflow differences account for the discrepancy. In this paper we concentrate on assessing the technical prospects of re-executability of a publication. As introduced above, there are many other factors that will contribute to the actual generalization of the findings including subject population details, data acquisition details, the nature of the processing and statistics (even if they can be re-executed), the underlying biological effect size, if present, etc. (see Figure 1). Take for example, the subject population. Too often researchers communicate a finding based on a convenience sample without any statement indicating that the results might not generalize to a sample that more accurately reflects human diversity (e.g. DeJesus *et al.*, 2019; Henrich *et al.*, 2010; Hruschka *et al.*, 2018; Rad *et al.*, 2018). Comprehensive and standardized description of all these additional factors are critical as well, but are beyond the scope of this evaluation. Our groups and others are looking into reporting standards for these areas as well.

The potential impact of reproducibility issues become most obvious when trying to make sense of the accumulated literature on specific topic areas (Rane *et al.*, 2015). For this reason, we have chosen a particular area, 'autism' as a way to focus the literature for this survey, so that the conclusions we reach can have potential specific implications for that topic area. We feel that the autism focus, however, will generate findings that will have similar implications to other psychiatric and developmental application areas.

In this paper, we: 1) develop a specification for what constitutes an assessment of the technical re-executability for a given publication in each of the domains of: data, software, execution environment, statistics and results; 2) codify this assessment in survey form; and 3) apply the survey to a subset of the autism neuroimaging literature published recently (~2018). From the results of this survey, we can begin to generalize the state of the re-executability of the recent autism neuroimaging literature, in order to identify trends and opportunities for the enhancement of the re-executability status in support of greater overall generalizability (and hence reproducibility) of the literature. The survey template could also be applied as part of the publication review process, in order to prospectively attempt to enhance these aspects of reproducibility.

Methods

Survey development

Following the concept of a 're-executable publication' (Kennedy, 2019), in order to assess the prospects of re-execution of a given paper, we assess 1) the availability of the starting data, 2) the perceived completeness of the analysis description (both data processing and statistical assessment), and 3) the availability of the detailed complete results (in order to

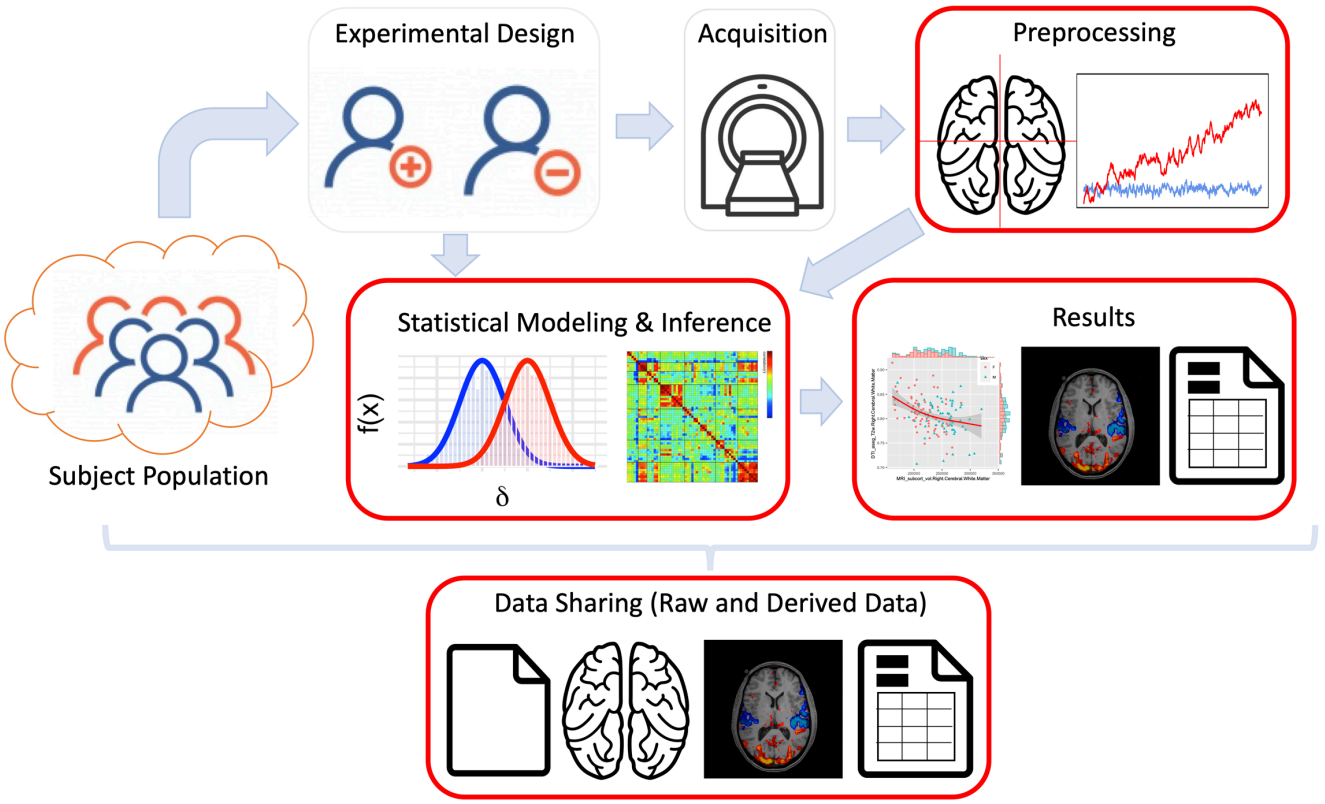

Figure 1. Essential elements of a publication. Elements of a publication that comprise a starting point for a structured exploration of the generalizability of a specific finding. The outlined areas define the technical prospects of re-executability of a finding that are evaluated in this survey.

verify accuracy of re-execution). Regarding the 'availability of the starting data', we assess if the publication indicates how someone[1] (other than the authors themselves) could appropriately access the data. The 'precision of the analysis description' ultimately asks if a reader who is reasonably skilled in the necessary domains, could precisely carry out the prescribed analysis steps. Specifically, are the software versions, operating system and complete parameters somehow made available to the reader? The 'detailed complete results' assesses if the publication indicates how to obtain the complete results, in order to both verify that the re-execution generates the same result and to overcome the limitations of only a selected summary being presented, which impedes a more complete meta-analysis of the literature.

In each of the three assessment areas, the survey distinguished between the theoretical potential for reproduction (such as complete descriptions of data used, software and commands executed, and statistical tests applied) and the practical potential for reproduction (whether the data is in fact accessible, whether the software is still available and will run). While the survey did not require the raters to actually reproduce the various steps, they were asked to use their professional judgement and past experience to determine the potential reproducibility. In these 'judgement' questions we allow responses of 'Yes', 'Approximately', 'I'm not sure', and 'No' to allow some degree of confidence in these judgements. For 'results availability', we coded 'Yes' if all of the results were indicated as being available, 'Partially' if some of the results were indicated as being available, and 'No' if none of the results were indicated as available or no indication of the results availability was provided. Note that our assessments are not if the analysis or data accessibility is 'optimum', or even 'correct', but rather if the assessor could redo the approach as described.

Figure 2 provides an overview of the survey design.

The survey was constructed in Google Forms. The details of the logic and wording of the survey forms was piloted (10 articles, three raters) within our own group, and then released for public comment to the BrainHack Slack[2] channel in August, 2018. The final complete (serialized) text of the survey is provided in S1 (see *Extended data*; Hodge *et al.*, 2020c).

[1]Although maybe not 'everyone', depending on the specific details of the data use agreement.

[2]Currently archived in the BrainHack Mattermost 'general' channel: https://mattermost.brainhack.org/brainhack/channels/general

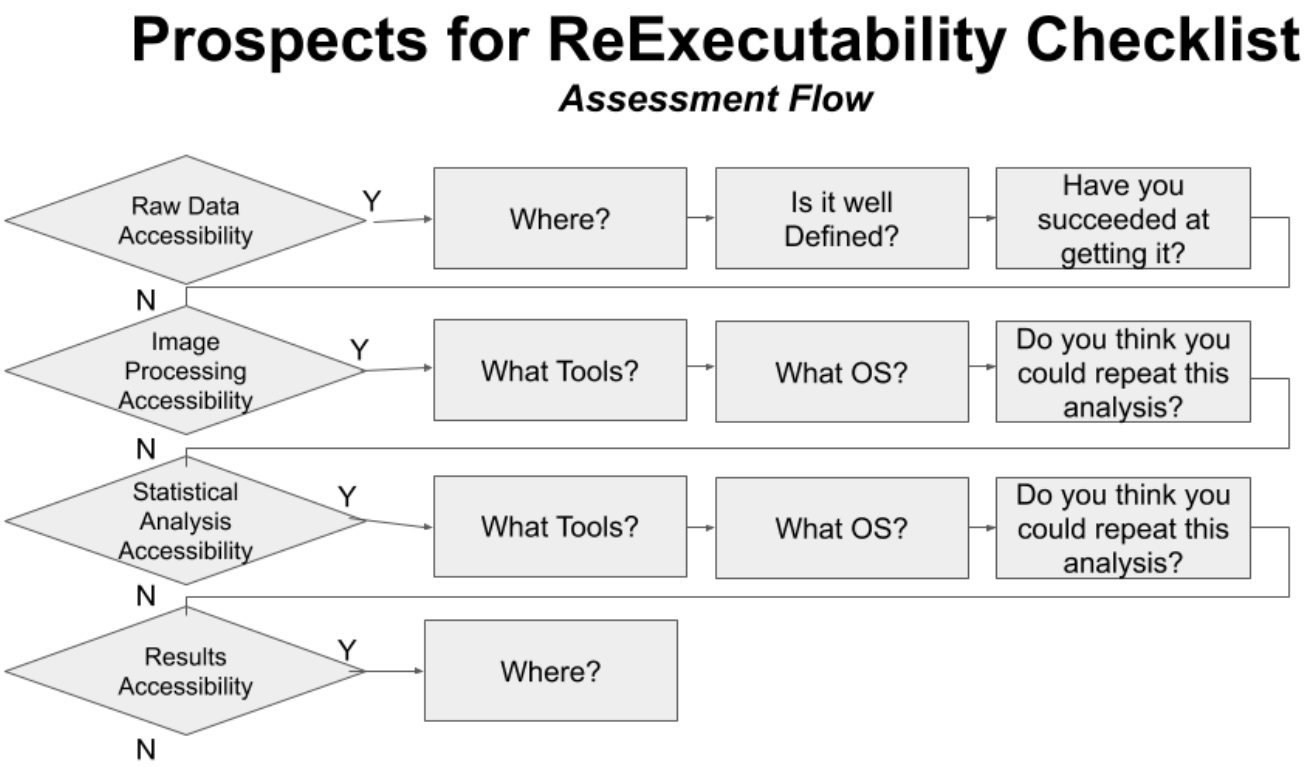

Figure 2. High-level survey design. OS, operating system.

Literature identification

On January 23, 2019, the following PubMed query was executed:

> ((("autistic disorder"[MeSH Terms] OR ("autistic"[All Fields] AND "disorder"[All Fields]) OR "autistic disorder"[All Fields] OR "autism"[All Fields]) AND ("magnetic resonance imaging"[MeSH Terms] OR ("magnetic"[All Fields] AND "resonance"[All Fields] AND "imaging"[All Fields]) OR "magnetic resonance imaging"[All Fields] OR "mri"[All Fields])) AND ("2014/01/25"[PDat] : "2019/01/23"[PDat] AND "humans"[MeSH Terms])

This is the expansion of the general query for 'autism AND MRI, qualified to select publications between 1/25/2014 - 1/23/2019 and where the MeSH term includes 'human'. This query generated 811 resultant publications at the time of the query (see S2, *Underlying data;* Hodge *et al.*, 2020a). We note that re-running the query today will generate additional results due to publications that have been added to PubMed after the search date but with publication dates within the defined range.

Survey application

Starting from the most recent publication and working backwards, we reviewed the title and abstract to verify publications that were indeed neuroimaging studies (not a case report or review), in English, related to autism and for which we could access the full text of the article. Working backwards from publication date, we selected the first 50 publications that met the above criteria. Of these 50 publications, 38 were available as free full text on PubMed, three were available as a PDF through a general Google Scholar search (publisher/author provided), two were available in PDF format from ResearchGate, and seven did not seem to be available without institutional access. The survey was applied to each paper by one of three raters (DNK, SMH, CH). Each of the final results were reviewed by a second rater (DNK or SMH) and consensus reached with the original rater if discrepancies were found.

Results

Literature selected

The final set of publications used in this report is tabulated in Table 1. The publication dates span from September 16, 2017 to October 1, 2018. Publications from 27 different journals are included. The publications selected covered a number of different MRI-based techniques (structural N=20, task-based fMRI N=14, resting-state fMRI N=13, diffusion MRI N=11 and magnetic resonance spectroscopy N=5)[3]. In this table we indicate what 'type'of imaging was performed: structural MRI (S), task-based functional MRI (F), resting state fMRI (RS), diffusion MRI (D), magnetic resonance spectroscopy (MRS), arterial spin labeling (ASL).

[3]A number of publications included data from multiple MRI types.

Table 1. Publications Included in the survey.

First Author	Title	Reference	PMID	Imaging Type	Assessors (Primary/ Check)
Marusak HA (Marusak *et al.*, 2018)	Mindfulness and dynamic functional neural connectivity in children and adolescents.	Behav Brain Res. 2018 Jan 15;336:211-218. doi: 10.1016/ j.bbr.2017.09.010. Epub 2017 Sep 5.	28887198	RS	Rev1/Rev3
Ramot M (Ramot *et al.*, 2017)	Direct modulation of aberrant brain network connectivity through real-time NeuroFeedback.	Elife. 2017 Sep 16;6. pii: e28974. doi: 10.7554/eLife.28974.	28917059	F	Rev2/Rev3
Bruno JL (Bruno *et al.*, 2017)	Longitudinal identification of clinically distinct neurophenotypes in young children with fragile X syndrome.	Proc Natl Acad Sci U S A. 2017 Oct 3;114(40):10767-10772. doi: 10.1073/ pnas.1620994114. Epub 2017 Sep 18.	28923933	S	Rev2/Rev3
Bottelier MA (Bottelier *et al.*, 2017)	Age-dependent effects of acute methylphenidate on amygdala reactivity in stimulant treatment-naive patients with Attention Deficit/Hyperactivity Disorder.	Psychiatry Res Neuroimaging. 2017 Nov 30;269:36-42. doi: 10.1016/j.pscychresns.20 17.09.009. Epub 2017 Sep 12.	28938219	F	Rev2/Rev3
Hotier S (Hotier *et al.*, 2017)	Social cognition in autism is associated with the neurodevelopment of the posterior superior temporal sulcus.	Acta Psychiatr Scand. 2017 Nov;136(5):517-525. doi: 10.1111/acps.12814. Epub 2017 Sep 22.	28940401	S	Rev2/Rev3
Chien YL (Chien *et al.*, 2017)	Altered white-matter integrity in unaffected siblings of probands with autism spectrum disorders.	Hum Brain Mapp. 2017 Dec;38(12):6053-6067. doi: 10.1002/ hbm.23810. Epub 2017 Sep 20.	28940697	D	Rev2/Rev3
Braden BB (Braden *et al.*, 2017)	Executive function and functional and structural brain differences in middle-age adults with autism spectrum disorder.	Autism Res. 2017 Dec;10(12):1945-1959. doi: 10.1002/ aur.1842. Epub 2017 Sep 21.	28940848	S, F	Rev3/Rev1
Hegarty JP 2nd (Hegarty *et al.*, 2018)	A proton MR spectroscopy study of the thalamus in twins with autism spectrum disorder.	Prog Neuropsychopharmacol Biol Psychiatry. 2018 Feb 2;81:153-160. doi: 10.1016/ j.pnpbp.2017.09.016. Epub 2017 Sep 21.	28941767	MRS	Rev3/Rev1
Joshi G (Joshi *et al.*, 2017)	Integration and Segregation of Default Mode Network Resting-State Functional Connectivity in Transition-Age Males with High-Functioning Autism Spectrum Disorder: A Proof-of- Concept Study.	Brain Connect. 2017 Nov;7(9):558-573. doi: 10.1089/ brain.2016.0483.	28942672	RS	Rev1/Rev3
Carlisi CO (Carlisi *et al.*, 2017)	Shared and Disorder-Specific Neurocomputational Mechanisms of Decision-Making in Autism Spectrum Disorder and Obsessive-Compulsive Disorder.	Cereb Cortex. 2017 Dec 1;27(12):5804-5816. doi: 10.1093/ cercor/bhx265.	29045575	F	Rev1/Rev3
White T (White *et al.*, 2018b)	Paediatric population neuroimaging and the Generation R Study: the second wave.	Eur J Epidemiol. 2018 Jan;33(1):99-125. doi: 10.1007/s10654- 017-0319-y. Epub 2017 Oct 24.	29064008	S, RS, D	Rev1/Rev3
Zhang F (Zhang *et al.*, 2018)	Whole brain white matter connectivity analysis using machine learning: An application to autism.	Neuroimage. 2018 May 15;172:826-837. doi: 10.1016/j.neur oimage.2017.10.029. Epub 2017 Oct 25.	29079524	D	Rev1/Rev3
Stanfield AC (Stanfield *et al.*, 2017)	Dissociation of Brain Activation in Autism and Schizotypal Personality Disorder During Social Judgments.	Schizophr Bull. 2017 Oct 21;43(6):1220-1228. doi: 10.1093/ schbul/sbx083.	29088456	F	Rev3/Rev1

First Author	Title	Reference	PMID	Imaging Type	Assessors (Primary/ Check)
Ni HC (Ni *et al.*, 2018)	Neural correlates of impaired self-regulation in male youths with autism spectrum disorder: A voxel-based morphometry study.	Prog Neuropsychopharmacol Biol Psychiatry. 2018 Mar 2;82:233-241. doi: 10.1016/ j.pnpbp.2017.11.008. Epub 2017 Nov 9.	29129723	S	Rev3/Rev1
Murakami Y (Murakami *et al.*, 2018)	Autistic traits modulate the activity of the ventromedial prefrontal cortex in response to female faces.	Neurosci Res. 2018 Aug;133:28-37. doi: 10.1016/ j.neures.2017.11.003. Epub 2017 Nov 12.	29141188	F	Rev3/Rev1
Balci TB (Balci *et al.*, 2018)	Broad spectrum of neuropsychiatric phenotypes associated with white matter disease in PTEN hamartoma tumor syndrome.	Am J Med Genet B Neuropsychiatr Genet. 2018 Jan;177(1):101-109. doi: 10.1002/ajmg. b.32610. Epub 2017 Nov 20.	29152901	S	Rev3/Rev1
Naaijen J (Naaijen *et al.*, 2018)	Striatal structure and its association with N-Acetylaspartate and glutamate in autism spectrum disorder and obsessive compulsive disorder.	Eur Neuropsychopharmacol. 2018 Jan;28(1):118-129. doi: 10.1016/ j.euroneuro.2017.11.010. Epub 2017 Nov 21.	29169826	S, MRS	Rev1/Rev3
Abbott AE (Abbott *et al.*, 2018)	Repetitive behaviors in autism are linked to imbalance of corticostriatal connectivity: a functional connectivity MRI study.	Soc Cogn Affect Neurosci. 2018 Jan 1;13(1):32-42. doi: 10.1093/scan/nsx129.	29177509	RS	Rev3/Rev1
White T (White *et al.*, 2018a)	Automated quality assessment of structural magnetic resonance images in children: Comparison with visual inspection and surface-based reconstruction.	Hum Brain Mapp. 2018 Mar;39(3):1218-1231. doi: 10.1002/ hbm.23911. Epub 2017 Dec 5.	29206318	S	Rev3/Rev1
Forde NJ	Multi-modal imaging investigation of anterior cingulate cortex cytoarchitecture in neurodevelopment	Eur Neuropsychopharmacol. 2018 Jan; 28(1):13-23. doi: 10.1016/ j.euroneuro.2017.11.021. Epub 2017 Dec 7.	29223496	S, D, MRS	Rev1/Rev3
Wei L (Wei *et al.*, 2018)	Aberrant development of the asymmetry between hemispheric brain white matter networks in autism spectrum disorder.	Eur Neuropsychopharmacol. 2018 Jan; 28(1):48-62. doi: 10.1016/ j.euroneuro.2017.11.018. Epub 2017 Dec 7.	29224969	D, S	Rev1/Rev3
Wadsworth HM (Wadsworth *et al.*, 2018)	Action simulation and mirroring in children with autism spectrum disorders.	Behav Brain Res. 2018 Apr 2;341:1-8. doi: 10.1016/ j.bbr.2017.12.012. Epub 2017 Dec 13.	29247748	F	Rev1/Rev3
Bernas A (Bernas *et al.*, 2018)	Wavelet coherence-based classifier: A resting-state functional MRI study on neurodynamics in adolescents with high- functioning autism.	Comput Methods Programs Biomed. 2018 Feb;154:143-151. doi: 10.1016/ j.cmpb.2017.11.017. Epub 2017 Nov 16.	29249338	RS	Rev3/Rev1
Alexander LM (Alexander *et al.*, 2017)	An open resource for transdiagnostic research in pediatric mental health and learning disorders.	Sci Data. 2017 Dec 19;4:170181. doi: 10.1038/ sdata.2017.181.	29257126	RS, S, D	Rev3/Rev1
Gibbard CR (Gibbard *et al.*, 2018)	Structural connectivity of the amygdala in young adults with autism spectrum disorder.	Hum Brain Mapp. 2018 Mar;39(3):1270-1282. doi: 10.1002/ hbm.23915. Epub 2017 Dec 19.	29265723	S, D	Rev3/Rev1
Dona O (Dona *et al.*, 2017)	Temporal fractal analysis of the rs-BOLD signal identifies brain abnormalities in autism spectrum disorder.	PLoS One. 2017 Dec 22;12(12):e0190081. doi: 10.1371/ journal.pone.0190081. eCollection 2017.	29272297	RS	Rev1/Rev3

First Author	Title	Reference	PMID	Imaging Type	Assessors (Primary/ Check)
Feczko E (Feczko *et al.*, 2018)	Subtyping cognitive profiles in Autism Spectrum Disorder using a Functional Random Forest algorithm.	Neuroimage. 2018 May 15;172:674-688. doi: 10.1016/j.neur oimage.2017.12.044. Epub 2017 Dec 21.	29274502	F	Rev3/Rev1
Ciaramidaro A (Ciaramidaro *et al.*, 2018)	Transdiagnostic deviant facial recognition for implicit negative emotion in autism and schizophrenia.	Eur Neuropsychopharmacol. 2018 Feb;28(2):264-275. doi: 10.1016/ j.euroneuro.2017.12.005. Epub 2017 Dec 21.	29275843	F	Rev3/Rev1
Ktena SI (Ktena *et al.*, 2018)	Metric learning with spectral graph convolutions on brain connectivity networks.	Neuroimage. 2018 Apr 1;169:431-442. doi: 10.1016/j.neuroi mage.2017.12.052. Epub 2017 Dec 24.	29278772	RS	Rev3/Rev1
Hu Y (Hu *et al.*, 2018)	The neural substrates of procrastination: A voxel-based morphometry study.	Brain Cogn. 2018 Mar;121:11-16. doi: 10.1016/ j.bandc.2018.01.001. Epub 2018 Jan 6.	29309854	S	Rev2/Rev3
Kohls G (Kohls *et al.*, 2018)	Altered reward system reactivity for personalized circumscribed interests in autism.	Mol Autism. 2018 Jan 30;9:9. doi: 10.1186/ s13229-018-0195- 7. eCollection 2018.	29423135	F	Rev3/Rev1
Boets B (Boets *et al.*, 2018)	Alterations in the inferior longitudinal fasciculus in autism and associations with visual processing: a diffusion-weighted MRI study.	Mol Autism. 2018 Feb 8;9:10. doi: 10.1186/ s13229-018- 0188-6. eCollection 2018.	29449909	D	Rev3/Rev1
Stivaros S (Stivaros *et al.*, 2018)	Randomised controlled trial of simvastatin treatment for autism in young children with neurofibromatosis type 1 (SANTA).	Mol Autism. 2018 Feb 22;9:12. doi: 10.1186/ s13229-018- 0190-z. eCollection 2018.	29484149	MRS, ASL, S, RS	Rev3/Rev1
Floris DL (Floris *et al.*, 2018)	Network-specific sex differentiation of intrinsic brain function in males with autism.	Mol Autism. 2018 Mar 6;9:17. doi: 10.1186/ s13229-018- 0192-x. eCollection 2018.	29541439	RS	Rev3/Rev1
Adamson K (Adamson & Troiani 2018)	Distinct and overlapping fusiform activation to faces and food.	Neuroimage. 2018 Jul 1;174:393-406. doi: 10.1016/j.neuroim age.2018.02.064. Epub 2018 Mar 22.	29578027	F	Rev3/Rev1
Li SJ (Li *et al.*, 2018)	Alterations of White Matter Connectivity in Preschool Children with Autism Spectrum Disorder.	Radiology. 2018 Jul;288(1):209-217. doi: 10.1148/ radiol.2018170059. Epub 2018 Mar 27.	29584599	D	Rev3/Rev1
Sen B (Sen *et al.*, 2018)	A general prediction model for the detection of ADHD and Autism using structural and functional MRI.	PLoS One. 2018 Apr 17;13(4):e0194856. doi: 10.1371/ journal.pone.0194856. eCollection 2018.	29664902	S, RS	Rev1/Rev3
Tsoi L (Tsoi *et al.*, 2018)	Neural substrates for moral judgments of psychological versus physical harm.	Soc Cogn Affect Neurosci. 2018 May 1;13(5):460-470. doi: 10.1093/scan/nsy029.	29718384	F	Rev1/Rev3
Guzman GEC (Guzman *et al.*, 2018)	Identification of alterations associated with age in the clustering structure of functional brain networks.	PLoS One. 2018 May 24;13(5):e0195906. doi: 10.1371/ journal.pone.0195906. eCollection 2018.	29795565	RS	Rev1/Rev3
Karahanoƒülu FI (Karahanoğlu *et al.*, 2018)	Diffusion-weighted imaging evidence of altered white matter development from late childhood to early adulthood in Autism Spectrum Disorder.	Neuroimage Clin. 2018 Jun 7;19:840-847. doi: 10.1016/ j.nicl.2018.06.002. eCollection 2018.	29946509	D	Rev3/Rev1

First Author	Title	Reference	PMID	Imaging Type	Assessors (Primary/ Check)
Zhao G (Zhao *et al.*, 2018)	Reduced structural complexity of the right cerebellar cortex in male children with autism spectrum disorder.	PLoS One. 2018 Jul 11;13(7):e0196964. doi: 10.1371/journal. pone.0196964. eCollection 2018.	29995885	S	Rev3/Rev1
Yan W (Yan *et al.*, 2018)	Aberrant hemodynamic responses in autism: Implications for resting state fMRI functional connectivity studies.	Neuroimage Clin. 2018 Apr 13;19:320-330. doi: 10.1016/ j.nicl.2018.04.013. eCollection 2018.	30013915	RS	Rev3/Rev1
Kim N (Kim *et al.*, 2018)	Aberrant Neural Activation Underlying Idiom Comprehension in Korean Children with High Functioning Autism Spectrum Disorder.	Yonsei Med J. 2018 Sep;59(7):897-903. doi: 10.3349/ ymj.2018.59.7.897.	30091324	F	Rev3/Rev1
Duret P (Duret *et al.*, 2018)	Gyrification changes are related to cognitive strengths in autism.	Neuroimage Clin. 2018 Aug 4;20:415-423. doi: 10.1016/ j.nicl.2018.04.036. eCollection 2018.	30128280	S	Rev3/Rev1
Na S (Na *et al.*, 2018)	White matter network topology relates to cognitive flexibility and cumulative neurological risk in adult survivors of pediatric brain tumors.	Neuroimage Clin. 2018 Aug 10;20:485-497. doi: 10.1016/ j.nicl.2018.08.015. eCollection 2018.	30148064	D	Rev3/Rev1
Chin R (Chin *et al.*, 2018)	Recognition of Schizophrenia with Regularized Support Vector Machine and Sequential Region of Interest Selection using Structural Magnetic Resonance Imaging.	Sci Rep. 2018 Sep 14;8(1):13858. doi: 10.1038/s41598-018- 32290-9.	30218016	S	Rev2/Rev3
Gertsvolf N (Gertsvolf *et al.*, 2018)	Association between Subcortical Morphology and Cerebral White Matter Energy Metabolism in Neonates with Congenital Heart Disease.	Sci Rep. 2018 Sep 19;8(1):14057. doi: 10.1038/s41598-018- 32288-3.	30232359	S, MRS	Rev2/Rev3
Gray JC (Gray *et al.*, 2018)	No evidence for morphometric associations of the amygdala and hippocampus with the five-factor model personality traits in relatively healthy young adults.	PLoS One. 2018 Sep 20;13(9):e0204011. doi: 10.1371/ journal.pone.0204011. eCollection 2018.	30235257	S	Rev2/Rev3
Vavla M (Vavla *et al.*, 2018)	Functional and Structural Brain Damage in Friedreich's Ataxia.	Front Neurol. 2018 Sep 6;9:747. doi: 10.3389/ fneur.2018.00747. eCollection 2018.	30237783	S, F, D	Rev2/Rev3
Mann C (Mann *et al.*, 2018)	The effect of age on vertex-based measures of the grey-white matter tissue contrast in autism spectrum disorder.	Mol Autism. 2018 Oct 1;9:49. doi: 10.1186/ s13229-018-0232- 6. eCollection 2018.	30302187	S	Rev3/Rev1

Survey results

A high-level summary of the survey results are represented in Figure 3. The complete set of question-by-question results are provided in S3 (see *Underlying data*; (Hodge *et al.*, 2020c).

Publication availability: 38 of the 50 (76%) publications appear to have 'free full text' available, according to the PubMed search. Of these, 33 are indexed in PubMed Central. Overall, 43 were freely available through either PubMed Central, Google Scholar or publisher or other websites.

Data availability: 17 of the 50 (34%) publications make reference to the availability of the data used in the publication. However, the publications that indicate availability are mostly reusing data from the large repositories, whereas the publications that do not indicate data availability are principally locally conducted studies. Thus, this indicates that a large fraction of the data being used in publications are not available to the community. 3 of these 17 indicate 'available upon request'. For the data that is available, the following resources are indicated: ABIDE 1 (Di Martino *et al.*, 2014), ABIDE 2 (Di Martino *et al.*, 2017), FCP/INDI (Mennes *et al.*, 2013), COINS (Scott *et al.*, 2011), LORIS (Das *et al.*, 2012), NITRC (Kennedy *et al.*, 2016), Preprocessed Connectomes Project (Puccio *et al.*, 2016), UKBiobank (Miller *et al.*, 2016), Brain Genomics Superstruct Project (Holmes *et al.*, 2015), ADHD-200 (HD-200 Consortium, 2012), and Human Connectome Project (Glasser *et al.*, 2016).

Image analysis: Virtually all of the publications surveyed indicate the imaging analysis software used (45 of 50, 90%). Most publications indicate the use of multiple tools. However, specific tool versions are indicated only about half of the time. Thirty-five different publicly released tools (plus a number of in-house packages) are used in this collection of 50 papers. Not surprisingly, the following tools are used in over 10 publications each: SPM (Ashburner *et al.*, 1998), FSL (Jenkinson *et al.*, 2012), and FreeSurfer (Makris *et al.*, 2003). The specific operating system used is rarely reported (1 of 50, 2%). Overall, our raters felt that in 80% of the publications a skilled image analyst could (or might be able to) repeat the analysis.

Statistical analysis: In approximately two thirds of the publications (66%), the statistical software is indicated, again with variable indication of version and no reporting of the operating system upon which the software was running. In summary, our raters felt that in 29 of the 50 papers (58%), a skilled statistical analyst could (or might be able to) repeat the analysis.

Results availability: Availability of the detailed results is fairly rare. All or partial results are available in seven of the 50 publications (14%).

Other observations: Two publications which were clinical trials indicated preregistration (with the EU Clinical Trials Register and ClinnicalTrials.gov). None of the non-clinical trials publications reviewed indicated pre-registration (Nosek *et al.*, 2019).

Discussion

The recent past literature of autism neuroimaging presents a somewhat consistent picture with respect to the prospects of re-executability with regard to the characteristics we examined in this report. Concerns of this sort have been raised in numerous contexts. The Organization for Human Brain Mapping's Committee on Best Practices in Data Analysis and Sharing (COBIDAS)[4], for example, digs very deeply into the recommendations for reporting and sharing in the literature. The work here is complementary as it takes a high-level gestalt view of re-executability.

Data availability is low, as we would expect to see given the current state of affairs. Figure 3 indicates that there may be a trend towards better data availability (more "Yes" values in the data access column as PubMed ID increases, a good proxy for relative date of publication).

While 80% of the publications were deemed to have repeatable image analysis, the low rate of specifying software version and vanishing rate of specifying operating system is troublesome, since these details can make a difference in results (Ghosh *et al.*, 2017; Glatard *et al.*, 2015). Even if there are currently only limited software options in some analysis domains, which may implicitly implicate the operating system used, such limitations are not guaranteed to persist through time and should not be assumed for the reader.

A smaller fraction of papers indicates statistical software other than image analysis software, perhaps in the belief that the statistical techniques are more important than the software used to implement the technique.

In both cases there is a distinct difference between the theoretical and practical ability to reproduce both the image analysis and statistical analysis. Rater confidence in the ability to re-execute image analysis and statistical analysis are similar, regardless of the fraction of cases where the software is specified.

The complete results availability criterion was rarely met. Lack of results availability causes a number of problems. Primarily, it is harder to confirm replication (or the degree to which replication was or was not achieved) without the complete set of reported observations, not just the summary tables or figures. Resorting to visual interpretations of 'similarity' of published figures remains fraught with issues that can hamper true understanding of new results compared to prior results. Lack of detailed results sharing also compromises subsequent meta-analytic studies that would strive to integrate observations

[4]http://www.humanbrainmapping.org/cobidas/

Enter the publication PubMed ID (PMID)	Does the publication provide information about how to access the raw imaging data?	Are the software tools/packages specified?	Is the operating system used for tool execution specified?	Do you think a reasonably skilled image analyst could re-execute this analysis?	Are the statistical tools/packages specified?	Do you think a reasonably skilled statistical analyst could re-execute this statistical analysis?	Are the complete results (derived images, summary data, etc.) available?
28887198	No	No	No	No	Yes	No	No
28917059	Yes	Yes	No	Approximately	No	No	No
28923933	No	Yes	No	Yes	Yes	No	Partially
28938219	No	Yes	No	Yes	Yes	I'm not sure	No
28940401	No	Yes	No	I'm not sure	Yes	I'm not sure	No
28940697	No	Yes	No	No	Yes	I'm not sure	No
28940848	Yes	Yes	No	I'm not sure	Yes	Approximately	No
28941767	No	Yes	No	I'm not sure	Yes	Approximately	No
28942672	No	Yes	No	Approximately	Yes	Approximately	No
29045575	No	Yes	No	Approximately	Yes	Yes	No
29064008	No	No	No	No	No	No	No
29079524	No	Yes	No	Approximately	No	No	No
29088456	No	Yes	No	Approximately	Yes	Approximately	No
29129723	No	Yes	No	Approximately	Yes	Approximately	No
29141188	No	Yes	No	Approximately	Yes	No	No
29152901	No	No	No	No	No	No	No
29169826	No	Yes	No	Yes	Yes	Yes	No
29177509	No	Yes	No	Approximately	Yes	No	No
29206318	No	Yes	No	I'm not sure	Yes	Approximately	No
29223496	No	Yes	No	I'm not sure	Yes	Approximately	No
29224969	Yes	Yes	No	Approximately	No	No	No
29247748	No	Yes	No	Yes	Yes	No	No
29249338	Yes	Yes	No	I'm not sure	Yes	No	No
29257126	Yes	Yes	No	Yes	No	No	No
29265723	No	Yes	No	I'm not sure	No	No	No
29272297	Yes	No	No	No	No	No	No
29274502	No	Yes	No	I'm not sure	No	No	No
29275843	No	Yes	No	Approximately	Yes	Approximately	No
29278772	Yes	No	No	No	No	No	No
29309854	No	Yes	No	I'm not sure	Yes	No	No
29423135	Yes	Yes	No	Yes	Yes	Yes	Yes
29449909	Yes	Yes	No	Approximately	Yes	I'm not sure	No
29484149	No	Yes	No	I'm not sure	Yes	I'm not sure	No
29541439	Yes	Yes	No	Approximately	No	No	Partially
29578027	No	Yes	No	Approximately	Yes	Approximately	No
29584599	No	Yes	No	Approximately	No	No	No
29664902	Yes	Yes	No	Approximately	Yes	Approximately	No
29718384	No	Yes	No	Approximately	Yes	Approximately	No
29795565	Yes	Yes	No	Approximately	No	No	No
29946509	No	Yes	No	Approximately	Yes	Approximately	No
29995885	Yes	Yes	Yes	No	Yes	Approximately	No
30013915	Yes	Yes	No	Approximately	Yes	I'm not sure	No
30091324	No	Yes	No	Approximately	Yes	Yes	No
30128280	No	Yes	No	No	No	No	No
30148064	No	Yes	No	I'm not sure	Yes	I'm not sure	No
30218016	No	Yes	No	I'm not sure	No	No	Yes
30232359	No	Yes	No	No	No	No	Yes
30235257	Yes	Yes	No	Yes	No	No	Yes
30237783	Yes	Yes	No	No	Yes	Yes	Partially
30302187	Yes	Yes	No	Approximately	Yes	Approximately	No
Totals: Yes	17	45	1	7	33	5	4
No	33	5	49	10	17	24	43
Approximately/ Partially/I'm not sure				33		21	3

Figure 3. Survey results summary. The 50 publications are summarized on the main factors of data availability, software specification, statistical specification and results availability.

across multiple publications. Finally, lack of complete results exacerbates the publication bias (Jennings & Van Horn, 2012) through focus on the (relatively few) statistically significant observations while not reporting the large set of observations that are not significant. Examples of complete results availability include when the individual statistical maps for a fMRI analysis are available in a resource such as NeuroVault[5], the individual segmentation results of a processing workflow are available at NITRC[6] or Zenodo[7], etc.

None of the reviewed publications indicated pre-registration (Nosek *et al.*, 2019). This is not surprising as pre-registration is a fairly new phenomenon, and its uptake in the literature can be expected to take a while. However, as a 'baseline' observation, it is still important to note, so that changes in the prevalence of the pre-registration practice can be monitored.

Limitations

The scope of our survey was rather limited; only 50 publications, and in a selected topic area, autism. However, as a retrospective starting point for evaluation, we believe that it fairly represents the qualitative impressions that investigators have about the nature of neuroimaging publications. We covered numerous neuroimaging subdomains: structural, diffusion, functional; and data and analytic practices in these subdomains can be rather variable. We acknowledge that the details of precise description and dissemination of data and methods may indeed vary by discipline. However, we argue that the 'best practice' principles that we are suggesting here are universal and domain-specific solutions are currently available. Also, even though fifty publications are included in the survey, a number of these publications share co-authors or originate from the same research groups. Specifically, 15 of these authors are listed on two or more publications, and 14 of the publications have authors that are also authors on other publications in this set.

The raters (DNK, CH, SMH) we used had over 15 years of neuroimaging research experience each; however, the specialties of each varied from more methodological/statistical to image analytic. This 'background' can influence the interpretation of how successfully other 'reasonably skilled' investigators could re-execute a given analysis. Familiarity with particular methods can both increase perceived confidence with its reuse ("Of course, everyone knows how to execute that common method") or decrease confidence ("There are so many details that I know could be varied, how do I know what was really done?"). In the absence of inclusion of explicitly re-executable data and methods in a publication (as in, for example, Ghosh *et al.*, 2017) the interpretation of the precision and completeness of the description with regard to re-executability will be somewhat imprecise and reader-dependent.

Finally, the assessment of each publication is performed on the accessible manuscript as published. It is possible that data and results sharing can have occurred after publication, but this fact may not be represented in the materials reviewed. Indeed, it would be a valuable service to facilitate a more prospective management of these critical re-execution factors that can support authors in making additional supporting data and methods available post publication.

[5]https://neurovault.org/
[6]https://nitrc.org
[7]https://zenodo.org/

Conclusions

In conclusion, we feel that the survey results presented here reflect a state of neuroimaging publication practices that leaves ample room for improvement. While reuse of existing data is good, the majority of new data being collected for use in publications is not made publicly available. While the listing of software used is good, important details for reproducibility, such as version, detailed parameters, operating system, etc. are not fully disclosed. Similarly, statistical assessment details are variably reported, making re-execution problematic and approximate. Finally, as very little of the complete results of a publication are disclosed, assessment of the similarity of future replication attempts is severely hampered. Given the overall state of uncertainty about how reproducible (and representative) specific neuroimaging findings are, it seems prudent to begin to tighten up the variables that we as authors do have in order to better support the effective accumulation of knowledge about conditions we study. Promoting best practices in ethical data sharing, complete analytic approach disclosure, and complete results reporting seem to be critical in integrating the complex set of observations we collectively have published about the brain and how it develops and ages. The implications of these observations are that authors should redouble their efforts to be comprehensive in their reporting, even after the publication, to make as accessible as possible the detailed methods and results that they are reporting on. Specifically, authors, reviewers and editors should insist on the complete declaration of: data source and availability status, all software and versions used for data analysis and statistical assessment, the operating system (and version) for data and statistical analysis, and the disposition of the analytic results. Such a 'checklist' would be a valuable asset for the community and will be the subject of future efforts. This future checklist should be developed in conjunction with journal specific guidelines, and other checklists (established in conjunction with the COBIDAS report (Nichols *et al.*, 2017), statistical reporting (Dexter & Shafer (2017), Nature Neuroscience Reporting Checklist, etc.). In such a way, publishers, editors and reviewers can impart more influence in the manuscripts that they encounter, in an effort to increase the transparency and completeness of the published record that they are playing their role in creating. Together, we hope that we can move the field forward and generate a literature that is more amenable to supporting the understanding of how our collective observations fit together in supporting the understanding of the brain.

Data availability

Underlying data

NITRC: CANDI Neuroimaging Access Point: S2_Raw_pubmed_Query_result.csv. http://doi.org/10.25790/bml0cm.68 (Hodge *et al.*, 2020a)

This project contains the following underlying data:

- S2_Raw_pubmed_Query_result.csv (complete PubMed query result from 1/23/2019)

NITRC: CANDI Neuroimaging Access Point: S3_CompleteSurveyData_v2.xlsx. http://doi.org/10.25790/bml0cm.81 (Hodge *et al.*, 2020b)

This project contains the following underlying data:

- S3_CompleteSurveyData.xlsx (complete survey results)

Extended data

NITRC: CANDI Neuroimaging Access Point: S1_Prospects for Reproducibility Check List_V2 - Google Forms.pdf http://doi.org/10.25790/bml0cm.66 (Hodge *et al.*, 2020c)

This project contains the following extended data:

- S1_Prospects for Reproducibility Check List_V2 - Google Forms.pdf (complete survey form)

References

Abbott AE, Linke AC, Nair A, *et al.*: **Repetitive Behaviors in Autism Are Linked to Imbalance of Corticostriatal Connectivity: A Functional Connectivity MRI Study.** *Soc Cogn Affect Neurosci.* 2018; **13**(1): 32–42.
PubMed Abstract | Publisher Full Text | Free Full Text

Adamson K, Troiani V: **Distinct and Overlapping Fusiform Activation to Faces and Food.** *NeuroImage.* 2018; **174**: 393–406.
PubMed Abstract | Publisher Full Text

Alexander LM, Escalera J, Ai L, *et al.*: **An Open Resource for Transdiagnostic Research in Pediatric Mental Health and Learning Disorders.** *Sci Data.* 2017; **4**: 170181.
PubMed Abstract | Publisher Full Text | Free Full Text

Ashburner J, Hutton C, Frackowiak R, *et al.*: **Identifying Global Anatomical Differences: Deformation-Based Morphometry.** *Hum Brain Mapp.* 1998; **6**(5–6): 348–57.
PubMed Abstract | Publisher Full Text | Free Full Text

Balci TB, Davila J, Lewis D, *et al.*: **Broad Spectrum of Neuropsychiatric Phenotypes Associated with White Matter Disease in PTEN Hamartoma Tumor Syndrome.** *Am J Med Genet B Neuropsychiatr Genet.* 2018; **177**(1): 101–9.
PubMed Abstract | Publisher Full Text

Bernas A, Aldenkamp AP, Zinger S: **Wavelet Coherence-Based Classifier: A Resting-State Functional MRI Study on Neurodynamics in Adolescents with High-Functioning Autism.** *Comput Methods Programs Biomed.* 2018; **154**: 143–51.
PubMed Abstract | Publisher Full Text

Boets B, Van Eylen L, Sitek K, *et al.*: **Alterations in the Inferior Longitudinal Fasciculus in Autism and Associations with Visual Processing: A Diffusion-Weighted MRI Study.** *Mol Autism.* 2018; **9**: 10.
PubMed Abstract | Publisher Full Text | Free Full Text

Bottelier MA, Schrantee A, Ferguson B, *et al.*: **Age-Dependent Effects of Acute Methylphenidate on Amygdala Reactivity in Stimulant Treatment-Naive Patients with Attention Deficit/Hyperactivity Disorder.** *Psychiatry Res Neuroimaging.* 2017; **269**: 36–42.
PubMed Abstract | Publisher Full Text

Braden BB, Smith CJ, Thompson A, *et al.*: **Executive Function and Functional and Structural Brain Differences in Middle-Age Adults with Autism Spectrum Disorder.** *Autism Res.* 2017; **10**(12): 1945–59.
PubMed Abstract | Publisher Full Text

Bruno JL, Romano D, Mazaika P, *et al.*: **Longitudinal Identification of Clinically Distinct Neurophenotypes in Young Children with Fragile X Syndrome.** *Proc Natl Acad Sci U S A.* 2017; **114**(40): 10767–72.
PubMed Abstract | Publisher Full Text | Free Full Text

Button KS, Ioannidis JPA, Mokrysz C, *et al.*: **Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience.** *Nat Rev Neurosci.* 2013; **14**(5): 365–76.
PubMed Abstract | Publisher Full Text

Carlisi CO, Norman L, Murphy CM, *et al.*: **Shared and Disorder-Specific Neurocomputational Mechanisms of Decision-Making in Autism Spectrum Disorder and Obsessive-Compulsive Disorder.** *Cereb Cortex.* 2017; **27**(12): 5804–16.
PubMed Abstract | Publisher Full Text | Free Full Text

Chien YL, Chen YJ, Hsu YC, *et al.*: **Altered White-Matter Integrity in Unaffected Siblings of Probands with Autism Spectrum Disorders.** *Hum Brain Mapp.* 2017; **38**(12): 6053–67.
PubMed Abstract | Publisher Full Text | Free Full Text

Chin R, You AX, Meng F, *et al.*: **Recognition of Schizophrenia with Regularized Support Vector Machine and Sequential Region of Interest Selection Using Structural Magnetic Resonance Imaging.** *Sci Rep.* 2018; **8**(1): 13858.
PubMed Abstract | Publisher Full Text | Free Full Text

Ciaramidaro A, Bölte S, Schlitt S, *et al.*: **Transdiagnostic Deviant Facial Recognition for Implicit Negative Emotion in Autism and Schizophrenia.** *Eur Neuropsychopharmacol.* 2018; **28**(2): 264–75.
PubMed Abstract | Publisher Full Text

Das S, Zijdenbos AP, Harlap J, *et al.*: **LORIS: A Web-Based Data Management System for Multi-Center Studies.** *Front Neuroinform.* 2012; **5**: 37.
PubMed Abstract | Publisher Full Text | Free Full Text

DeJesus JM, Callanan MA, Solis G, *et al.*: **Generic language in scientific communication.** *Proc Natl Acad Sci U S A.* 2019; **116**(37): 18370–18377.
PubMed Abstract | Publisher Full Text | Free Full Text

Dexter F, Shafer SL: **Narrative Review of Statistical Reporting Checklists, Mandatory Statistical Editing, and Rectifying Common Problems in the Reporting of Scientific Articles.** *Anesth Analg.* 2017; **124**(3): 943–47.
PubMed Abstract | Publisher Full Text

Di Martino A, O'Connor D, Chen B, *et al.*: **Enhancing Studies of the Connectome in Autism Using the Autism Brain Imaging Data Exchange II.** *Sci Data.* 2017; **4**: 170010.
PubMed Abstract | Publisher Full Text | Free Full Text

Di Martino A, Yan CG, Li Q, *et al.*: **The Autism Brain Imaging Data Exchange: Towards a Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism.** *Mol Psychiatry.* 2014; **19**(6): 659–67.
PubMed Abstract | Publisher Full Text | Free Full Text

Dona O, Hall GB, Noseworthy MD: **Temporal Fractal Analysis of the Rs-BOLD Signal Identifies Brain Abnormalities in Autism Spectrum Disorder.** *PloS One.* 2017; **12**(12): e0190081.
PubMed Abstract | Publisher Full Text | Free Full Text

Duret P, Samson F, Pinsard B, *et al.*: **Gyrification Changes Are Related to Cognitive Strengths in Autism.** *Neuroimage Clin.* 2018; **20**: 415–23.
PubMed Abstract | Publisher Full Text | Free Full Text

Feczko E, Balba NM, Miranda-Dominguez O, *et al.*: **Subtyping Cognitive Profiles in Autism Spectrum Disorder Using a Functional Random Forest Algorithm.** *NeuroImage.* 2018; **172**: 674–88.
PubMed Abstract | Publisher Full Text | Free Full Text

Floris DL, Lai MC, Nath T, *et al.*: **Network-Specific Sex Differentiation of Intrinsic Brain Function in Males with Autism.** *Mol Autism.* 2018; **9**: 17.
PubMed Abstract | Publisher Full Text | Free Full Text

Gertsvolf N, Votava-Smith JK, Ceschin R, *et al.*: **Association between Subcortical Morphology and Cerebral White Matter Energy Metabolism in Neonates with Congenital Heart Disease.** *Sci Rep.* 2018; **8**(1): 14057.
PubMed Abstract | Publisher Full Text | Free Full Text

Ghosh SS, Poline JB, Keator DB, *et al.*: **A very simple, re-executable neuroimaging publication [version 2; peer review: 1 approved, 3 approved with reservations].** *F1000Res.* 2017; **6**: 124.
PubMed Abstract | Publisher Full Text | Free Full Text

Gibbard CR, Ren J, Skuse DH, *et al.*: **Structural Connectivity of the Amygdala in Young Adults with Autism Spectrum Disorder.** *Hum Brain Mapp.* 2018; **39**(3): 1270–82.
PubMed Abstract | Publisher Full Text | Free Full Text

Glasser MF, Smith SM, Marcus DS, *et al.*: **The Human Connectome Project's Neuroimaging Approach.** *Nat Neurosci.* 2016; **19**(9): 1175–87.
PubMed Abstract | Publisher Full Text | Free Full Text

Glatard T, Lewis LB, da Silva RF, *et al.*: **Reproducibility of Neuroimaging Analyses across Operating Systems.** *Front Neuroinform.* 2015; **9**: 12.
PubMed Abstract | Publisher Full Text | Free Full Text

Gorgolewski KJ, Poldrack RA: **A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research.** *PLoS Biol.* 2016; **14**(7): e1002506.
PubMed Abstract | Publisher Full Text | Free Full Text

Gray JC, Owens MM, Hyatt CS, *et al.*: **No Evidence for Morphometric Associations of the Amygdala and Hippocampus with the Five-Factor Model Personality Traits in Relatively Healthy Young Adults.** *PloS One.* 2018;

13(9): e0204011.
PubMed Abstract | Publisher Full Text | Free Full Text

Guzman GEC, Sato JR, Vidal MC, *et al.*: **Identification of Alterations Associated with Age in the Clustering Structure of Functional Brain Networks.** *PloS One.* 2018; **13**(5): e0195906.
PubMed Abstract | Publisher Full Text | Free Full Text

Hegarty JP, Gu M, Spielman DM, *et al.*: **A Proton MR Spectroscopy Study of the Thalamus in Twins with Autism Spectrum Disorder.** *Prog Neuropsychopharmacol Biol Psychiatry.* 2018; **81**: 153–60.
PubMed Abstract | Publisher Full Text | Free Full Text

Henrich J, Heine SJ, Norenzayan A: **The weirdest people in the world?** *Behav Brain Sci.* 2010; **33**(2–3): 61–83; discussion 83-135.
PubMed Abstract | Publisher Full Text

Hodge SM, Haselgrove C, Honor L, *et al.*: **S2_Raw_pubmed_Query_result.csv [Data].** Washington: NITRC. 2020a.
http://www.doi.org/10.25790/bml0cm.68

Hodge SM, Haselgrove C, Honor L, *et al.*: **S3_CompleteSurveyData.xlsx [Data].** Washington: NITRC. 2020b.
http://www.doi.org/10.25790/bml0cm.81

Hodge SM, Haselgrove C, Honor L, *et al.*: **S1_ ProspectsforReproducibilityCheckList_V2-GoogleForms.pdf [Document].** Washington: NITRC. 2020c.
http://www.doi.org/10.25790/bml0cm.66

Holmes AJ, Hollinshead MO, O'Keefe TM, *et al.*: **Brain Genomics Superstruct Project Initial Data Release with Structural, Functional, and Behavioral Measures.** *Sci Data.* 2015; **2**: 150031.
PubMed Abstract | Publisher Full Text | Free Full Text

Hotier S, Leroy F, Boisgontier J, *et al.*: **Social Cognition in Autism Is Associated with the Neurodevelopment of the Posterior Superior Temporal Sulcus.** *Acta Psychiatr Scand.* 2017; **136**(5): 517–25.
PubMed Abstract | Publisher Full Text

Hruschka DJ, Medin DL, Rogoff B, *et al.*: **Pressing questions in the study of psychological and behavioral diversity.** *Proc Natl Acad Sci U S A.* 2018; **115**(45): 11366–11368.
PubMed Abstract | Publisher Full Text | Free Full Text

Hu Y, Liu P, Guo Y, *et al.*: **The Neural Substrates of Procrastination: A Voxel-Based Morphometry Study.** *Brain Cogn.* 2018; **121**: 11–16.
PubMed Abstract | Publisher Full Text

Jenkinson M, Beckmann CF, Behrens TEJ, *et al.*: **FSL.** *Neuroimage.* 2012; **62**(2): 782–90.
PubMed Abstract | Publisher Full Text

Jennings RG, VanHorn JD: **Publication Bias in Neuroimaging Research: Implications for Meta-Analyses.** *Neuroinformatics.* 2012; **10**(1): 67–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Joshi G, Anteraper SA, Patil KR, *et al.*: **Integration and Segregation of Default Mode Network Resting-State Functional Connectivity in Transition-Age Males with High-Functioning Autism Spectrum Disorder: A Proof-of-Concept Study.** *Brain Connect.* 2017; **7**(9): 558–73.
PubMed Abstract | Publisher Full Text | Free Full Text

Kapur S, Phillips AG, Insel TR: **Why Has It Taken so Long for Biological Psychiatry to Develop Clinical Tests and What to Do about It?** *Mol Psychiatry.* 2012; **17**(12): 1174–79.
PubMed Abstract | Publisher Full Text

Karahanoğlu FI, Baran BI, Nguyen OTH, *et al.*: **Diffusion-Weighted Imaging Evidence of Altered White Matter Development from Late Childhood to Early Adulthood in Autism Spectrum Disorder.** *Neuroimage Clin.* 2018; **19**: 840–47.
PubMed Abstract | Publisher Full Text | Free Full Text

Kennedy DN: **The ReproPub: A Hybrid Research Object for Supporting Publication-Level Re-Execution and Generalization of Neuroimaging Research Findings.** *Zenodo.* 2019.
Publisher Full Text

Kennedy DN, Abraham SA, Bates JF, *et al.*: **Everything Matters: The ReproNim Perspective on Reproducible Neuroimaging.** *Front Neuroinform.* 2019; **13**: 1.
PubMed Abstract | Publisher Full Text | Free Full Text

Kennedy DN, Haselgrove C, Riehl J, *et al.*: **The NITRC Image Repository.** *NeuroImage.* 2016; **124**(Pt B): 1069–73.
PubMed Abstract | Publisher Full Text | Free Full Text

Kim N, Choi US, Ha S, *et al.*: **Aberrant Neural Activation Underlying Idiom Comprehension in Korean Children with High Functioning Autism Spectrum Disorder.** *Yonsei Med J.* 2018; **59**(7): 897–903.
PubMed Abstract | Publisher Full Text | Free Full Text

Kohls G, Antezana L, Mosner MG, *et al.*: **Altered Reward System Reactivity for Personalized Circumscribed Interests in Autism.** *Mol Autism.* 2018; **9**: 9.
PubMed Abstract | Publisher Full Text | Free Full Text

Ktena SI, Parisot S, Ferrante E, *et al.*: **Metric Learning with Spectral Graph Convolutions on Brain Connectivity Networks.** *Neuroimage.* 2018; **169**: 431–42.
PubMed Abstract | Publisher Full Text

Li SJ, Wang Y, Qian L, *et al.*: **Alterations of White Matter Connectivity in Preschool Children with Autism Spectrum Disorder.** *Radiology.* 2018; **288**(1): 209–17.
PubMed Abstract | Publisher Full Text

Makris N, Hodge SM, Haselgrove C, *et al.*: **Human Cerebellum: Surface-Assisted Cortical Parcellation and Volumetry with Magnetic Resonance Imaging.** *J Cogn Neurosci.* 2003; **15**(4): 584–99.
PubMed Abstract | Publisher Full Text

Mann C, Bletsch A, Andrews D, *et al.*: **The Effect of Age on Vertex-Based Measures of the Grey-White Matter Tissue Contrast in Autism Spectrum Disorder.** *Mol Autism.* 2018; **9**: 49.
PubMed Abstract | Publisher Full Text | Free Full Text

Marusak HA, Elrahal F, Andrews CA, *et al.*: **Mindfulness and Dynamic Functional Neural Connectivity in Children and Adolescents.** *Behav Brain Res.* 2018; **336**: 211–18.
PubMed Abstract | Publisher Full Text | Free Full Text

Mennes M, Biswal FB, Castellanos X, *et al.*: **Making Data Sharing Work: The FCP/INDI Experience.** *Neuroimage.* 2013; **82**: 683–91.
PubMed Abstract | Publisher Full Text | Free Full Text

Miller KL, Alfaro-Almagro F, Bangerter NK, *et al.*: **Multimodal Population Brain Imaging in the UK Biobank Prospective Epidemiological Study.** *Nat Neurosci.* 2016; **19**(11): 1523–36.
PubMed Abstract | Publisher Full Text | Free Full Text

Murakami Y, Sakai S, Takeda K, *et al.*: **Autistic Traits Modulate the Activity of the Ventromedial Prefrontal Cortex in Response to Female Faces.** *Neurosci Res.* 2018; **133**: 28–37.
PubMed Abstract | Publisher Full Text

Naaijen J, Zwiers MP, Forde NJ, *et al.*: **Striatal Structure and Its Association with N-Acetylaspartate and Glutamate in Autism Spectrum Disorder and Obsessive Compulsive Disorder.** *Eur Neuropsychopharmacol.* 2018; **28**(1): 118–29.
PubMed Abstract | Publisher Full Text

Na S, Li L, Crosson B, *et al.*: **White Matter Network Topology Relates to Cognitive Flexibility and Cumulative Neurological Risk in Adult Survivors of Pediatric Brain Tumors.** *Neuroimage Clin.* 2018; **20**: 485–97.
PubMed Abstract | Publisher Full Text | Free Full Text

Nichols TE, Das S, Eickhoff SB, *et al.*: **Best Practices in Data Analysis and Sharing in Neuroimaging Using MRI.** *Nat Neurosci.* 2017; **20**(3): 299–303.
PubMed Abstract | Publisher Full Text | Free Full Text

Ni HS, Lin HY, Tseng WYI, *et al.*: **Neural Correlates of Impaired Self-Regulation in Male Youths with Autism Spectrum Disorder: A Voxel-Based Morphometry Study.** *Prog Neuropsychopharmacol Biol Psychiatry.* 2018; **82**: 233–41.
PubMed Abstract | Publisher Full Text

Nosek BA, Beck ED, Campbell L, *et al.*: **Preregistration Is Hard, And Worthwhile.** *Trends Cogn Sci.* 2019; **23**(10): 815–18.
PubMed Abstract | Publisher Full Text

Puccio B, Pooley JP, Pellman JS, *et al.*: **The Preprocessed Connectomes Project Repository of Manually Corrected Skull-Stripped T1-Weighted Anatomical MRI Data.** *GigaScience.* 2016; **5**(1): 45.
PubMed Abstract | Publisher Full Text | Free Full Text

Rad MS, Martingano AJ, Ginges J: **Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population.** *Proc Natl Acad Sci U S A.* 2018; **115**(45): 11401–11405.
PubMed Abstract | Publisher Full Text | Free Full Text

Ramot M, Kimmich S, Gonzalez-Castillo J, *et al.*: **Direct Modulation of Aberrant Brain Network Connectivity through Real-Time NeuroFeedback.** *Elife.* 2017; **6**: e28974.
PubMed Abstract | Publisher Full Text | Free Full Text

Rane P, Cochran D, Hodge SM, *et al.*: **Connectivity in Autism: A Review of MRI Connectivity Studies.** *Harv Rev Psychiatry.* 2015; **23**(4): 223–44.
PubMed Abstract | Publisher Full Text | Free Full Text

Scott A, Courtney W, Wood D, *et al.*: **COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets.** *Front Neuroinform.* 2011; **5**: 33.
PubMed Abstract | Publisher Full Text | Free Full Text

Sen B, Borle NC, Greiner R, *et al.*: **A General Prediction Model for the Detection of ADHD and Autism Using Structural and Functional MRI.** *PLoS One.* 2018; **13**(4): e0194856.
PubMed Abstract | Publisher Full Text | Free Full Text

Stanfield AC, Philip RCM, Whalley H, *et al.*: **Dissociation of Brain Activation in Autism and Schizotypal Personality Disorder During Social Judgments.** *Schizophr Bull.* 2017; **43**(6): 1220–28.
PubMed Abstract | Publisher Full Text | Free Full Text

Stivaros S, Garg S, Tziraki M, *et al.*: **Randomised Controlled Trial of Simvastatin Treatment for Autism in Young Children with Neurofibromatosis Type 1 (SANTA).** *Mol Autism.* 2018; **9**: 12.
PubMed Abstract | Publisher Full Text | Free Full Text

The ADHD-200 Consortium: **The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience.** *Front Syst Neurosci.* 2012; **6**: 62.
PubMed Abstract | Publisher Full Text | Free Full Text

Tsoi L, Dungan JA, Chakroff A, *et al.*: **Neural Substrates for Moral Judgments of Psychological versus Physical Harm.** *Soc Cogn Affect Neurosci.* 2018; **13**(5): 460–70.
PubMed Abstract | Publisher Full Text | Free Full Text

Vavla M, Arrigoni F, Nordio A, *et al.*: **Functional and Structural Brain Damage**

in Friedreich's Ataxia. *Front Neurol.* 2018; **9**: 747.
PubMed Abstract | Publisher Full Text | Free Full Text

Wadsworth HM, Maximo JO, Donnelly RJ, *et al.*: **Action Simulation and Mirroring in Children with Autism Spectrum Disorders.** *Behav Brain Res.* 2018; **341**: 1–8.
PubMed Abstract | Publisher Full Text | Free Full Text

Wei L, Zhong S, Nie S, *et al.*: **Aberrant Development of the Asymmetry between Hemispheric Brain White Matter Networks in Autism Spectrum Disorder.** *Eur Neuropsychopharmacol.* 2018; **28**(1): 48–62.
PubMed Abstract | Publisher Full Text

White T, Jansen PR, Muetzel RL, *et al.*: **Automated Quality Assessment of Structural Magnetic Resonance Images in Children: Comparison with Visual Inspection and Surface-Based Reconstruction.** *Hum Brain Mapp.* 2018a; **39**(3): 1218–31.
PubMed Abstract | Publisher Full Text | Free Full Text

White T, Muetzel RL, Marroun HE, *et al.*: **Paediatric Population Neuroimaging and the Generation R Study: The Second Wave.** *Eur J Epidemiol.* 2018b; **33**(1): 99–125.
PubMed Abstract | Publisher Full Text | Free Full Text

Yan W, Rangaprakash D, Deshpande G: **Aberrant Hemodynamic Responses in Autism: Implications for Resting State fMRI Functional Connectivity Studies.** *Neuroimage Clin.* 2018; **19**: 320–30.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhang F, Savadjiev P, Cai W, *et al.*: **Whole Brain White Matter Connectivity Analysis Using Machine Learning: An Application to Autism.** *Neuroimage.* 2018; **172**: 826–37.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhao G, Walsh K, Long J, *et al.*: **Reduced Structural Complexity of the Right Cerebellar Cortex in Male Children with Autism Spectrum Disorder.** *PLoS One.* 2018; **13**(7): e0196964.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 15 March 2021

https://doi.org/10.5256/f1000research.54296.r80936

✔ **Karsten Specht**
Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

It was a pleasure to have discussed this interesting and important paper with the authors. I hope we will soon see more of those critical surveys, which may improve the reliability of neuroimaging studies and how they are conducted and reported.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroimaging, fMRI, MRS, reliability

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 22 September 2020

https://doi.org/10.5256/f1000research.27929.r70151

✔ **Travis Riddle**
Data Science and Sharing Team, Intramural Research Program, National Institute of Mental

Health, National Institutes of Health, Bethesda, MD, USA

Adam Thomas
Data Science and Sharing Team, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

This paper highlights an important concern regarding the quality of the science seen in published literature. We applaud the authors for undertaking this work, and agree with their general conclusions that there are many opportunities for researchers to improve their reporting. However, we feel that it is worth mentioning a few details in the paper that caused us some concern or confusion.

First, the paper leads with a summary of some of the issues surrounding reproducibility of science. We urge the authors to make note of another concern that is widely overlooked. Too often researchers communicate a finding based on a convenience sample without any statement indicating that the results might not generalize to a sample that more accurately reflects human diversity (e.g. DeJesus, Callanan, Solis & Gelman, 2019[1]; Hruschka, Medin, Rogoff & Henrich, 2018[2]; Rad, Martingano & Ginges, 2018[3]; Henrich, Heine, & Norenzayan, 2010[4]). Of course, this paper is about other reasons for reproducibility, but it seems appropriate to mention this, especially in light of the increased attention given to exclusionary social systems in other domains.

We also had some concern with the concepts of the 'precision of analysis' (methods paragraph 1). This issue in particular seems difficult to assess reliably, and so there might be a higher degree of measurement error for this concept in comparison to the other concepts. We appreciate that the authors allude to this difficulty later in the paper, when they state that more expertise could also lead to higher levels of measurement error, but here we feel that a more explicit note of caution that these variables in particular should be viewed with additional skepticism.

The description of how the assessment was applied to each paper was difficult to follow ('survey application' pg 4: "one of three raters applied the survey to each of these articles. Each of the final results..."). Does this mean that each paper was evaluated by 1 reviewer? It seems like it would be useful to have more than one person complete the review. This would allow the reader to have a sense of the degree of inter-rater reliability. If the reliability was low, that would lead us to be a little more credulous with respect to many of the subsequent findings. If there was more than one reviewer per paper, the authors should report some standard inter-rater agreement metrics. If not, an independent assessment by other raters (along with ratings) would be a wonderful addition to the work, if a bit effort-intensive. Adding a column to figure 2 listing which rater assessed which publication would be helpful. This column could be coded for anonymity (Rater 1, Rater 2) if the authors so choose.

We additionally found the category of 'results availability' to be a little vague. Especially so since it seems as though papers never reached this cutoff. What does it take for a paper to have complete results availability? Some models might have thousands of parameters or more, and some papers might include dozens or hundreds of fitted models. Does this mean that all parameter values would be reported, confidence intervals, model fit statistics, and so forth would be reported?

References

1. DeJesus J, Callanan M, Solis G, Gelman S: Generic language in scientific communication.

Proceedings of the National Academy of Sciences. 2019; **116** (37): 18370-18377 Publisher Full Text
2. Hruschka D, Medin D, Rogoff B, Henrich J: Pressing questions in the study of psychological and behavioral diversity. *Proceedings of the National Academy of Sciences*. 2018; **115** (45): 11366-11368 Publisher Full Text
3. Rad M, Martingano A, Ginges J: Toward a psychology ofHomo sapiens : Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*. 2018; **115** (45): 11401-11405 Publisher Full Text
4. Henrich J, Heine SJ, Norenzayan A: The weirdest people in the world?. *Behav Brain Sci*. 2010; **33** (2-3): 61-83; discussion 83 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroscience, Psychology, Computer Science, Neuroimaging

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 27 Jan 2021

David Kennedy, University of Massachusetts Medical School, Worcester, USA

We would like to thank the reviewer for their thoughtful comments and attention to this manuscript. Here we outline our responses to the comments, and indicate where in the manuscript we have made changes.

- We urge the authors to make note of another concern that is widely overlooked. Too often researchers communicate a finding based on a convenience sample without any statement indicating that the results might not generalize to a sample that more accurately reflects human diversity (e.g. DeJesus, Callanan, Solis & Gelman, 20191; Hruschka, Medin, Rogoff & Henrich, 20182; Rad, Martingano & Ginges, 20183;

Henrich, Heine, & Norenzayan, 20104). Of course, this paper is about other reasons for reproducibility, but it seems appropriate to mention this, especially in light of the increased attention given to exclusionary social systems in other domains.

Response: As we discuss in response to Reviewer 1 above, the details of the subject pool ascertainment and its' generalizability is beyond the scope of this manuscript, but as this is an important point, we have included it in our updated introduction.

"In this paper we concentrate on assessing the technical prospects of re-executability of a publication. As introduced above, there are many other factors that will contribute to the actual generalization of the findings including subject population details, data acquisition details, the nature of the processing and statistics (even if they can be re-executed), the underlying biological effect size, if present, etc. (see Figure 1). Take for example, the subject population. Too often researchers communicate a finding based on a convenience sample without any statement indicating that the results might not generalize to a sample that more accurately reflects human diversity (e.g. DeJesus, Callanan, Solis & Gelman, 20191; Hruschka, Medin, Rogoff & Henrich, 20182; Rad, Martingano & Ginges, 20183; Henrich, Heine, & Norenzayan, 20104). Comprehensive and standardized description of all these additional factors are critical as well, but are beyond the scope of this evaluation. Our groups and others are looking into reporting standards for these areas as well."

- We also had some concern with the concepts of the 'precision of analysis' (methods paragraph 1). This issue in particular seems difficult to assess reliably, and so there might be a higher degree of measurement error for this concept in comparison to the other concepts. We appreciate that the authors allude to this difficulty later in the paper, when they state that more expertise could also lead to higher levels of measurement error, but here we feel that a more explicit note of caution that these variables in particular should be viewed with additional skepticism.

Response: In order to help the reader appreciate the cautionary note regarding these assessments, we have updated the notion of 'precision' to "perceived completeness" to help remind that the precision assessment is in the mind of the assessor. This is reflected in Methods paragraph one and elaborated upon a little more in Limitations paragraph two:

"In the absence of inclusion of explicitly re-executable data and methods in a publication (as in, for example, Ghosh, et al.) the interpretation of the precision and completeness of the description with regard to re-executability will be somewhat imprecise and reader-dependent."

- The description of how the assessment was applied to each paper was difficult to follow ('survey application' pg 4: "one of three raters applied the survey to each of these articles. Each of the final results..."). Does this mean that each paper was evaluated by 1 reviewer? It seems like it would be useful to have more than one person complete the review. This would allow the reader to have a sense of the degree of inter-rater reliability.

Response: We have attempted to clarify the text regarding the validation cases (pilot

assessment) and the dual raters for each publication.

"The survey was applied to each paper by one of three raters (DNK, SMH, CH). Each of the final results were reviewed by a second rater (DNK or SMH) and consensus reached with the original rater if discrepancies were found."

- [related] an independent assessment by other raters (along with ratings) would be a wonderful addition to the work, if a bit effort-intensive. Adding a column to figure 2 listing which rater assessed which publication would be helpful. This column could be coded for anonymity (Rater 1, Rater 2) if the authors so choose.

Response: Table 1 now has a column indicating which raters (Rev1, Rev2 or Rev3) reviewed each publication as the 'primary' or 'checking' reviewer.

- We additionally found the category of 'results availability' to be a little vague. Especially so since it seems as though papers never reached this cutoff. What does it take for a paper to have complete results availability?

Response: We agree that the 'complete results availability' was a lofty and somewhat variable goal statement. We have tried to clarify the meaning and ways that this can be satisfied in the updated text of paragraph five in the Discussion:

"The complete results availability criterion was rarely met. Lack of results availability causes a number of problems. Primarily, it is harder to confirm replication (or the degree to which replication was or was not achieved) without the complete set of reported observations, not just the summary tables or figures. Resorting to visual interpretations of 'similarity' of published figures remains fraught with issues that can hamper true understanding of new results compared to prior results. Lack of detailed results sharing also compromises subsequent meta-analytic studies that would strive to integrate observations across multiple publications. Finally, lack of complete results exacerbates the publication bias (Jennings and Van Horn 2012) through focus on the (relatively few) statistically significant observations while not reporting the large set of observations that are not significant. Examples of complete results availability include when the individual statistical maps for a fMRI annalysis are available in a resource such as NeuroVault, the individual segmentation results of a processing workflow are available at NITRC or Zenodo, etc."

Competing Interests: No competing interests were disclosed.

Reviewer Report 22 September 2020

https://doi.org/10.5256/f1000research.27929.r70168

Karsten Specht
Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

Summary
The article by Hodge and co-workers summarises an attempt in assessing the possibility to replicate 50 published neuroimaging studies on autism. The results indicate that the majority of the studies provide only partial information that would be required for replication of the study. In particular, are information about the operating system missing, only a few studies share their data or other files, and the description of the different analysis steps are sparsely described.

Assessment:
The article is well written with a clearly described method and results.
The study provides a suitable method that could easily be applied to other research topics, as well. However, the conclusions that can be drawn from this study are still limited in my view, since it would have been good to include further information in the survey, which I will list below:

1. In my opinion, the authors focus too much on the technical aspects of a study. Although the authors introduce that a “spectrum-diagnosis” might generate further problems, they do not follow this up in the survey. I would like to see at least one additional column that codes whether the diagnostic criteria and sample are replicable, i.e. are the patients well characterised (age, gender, education), are the diagnostic instrument mentioned, cut-off criteria, etc.

2. I suggest including another column (at least) in the supplementary material S3 that also lists the imaging modality, i.e. structural MRI, fMRI, MRS, DTI, since they also partly represent different disciplines and traditions in publishing. Further, some methods have only a very limited number of software tools, like MRS, which are often restricted to only one (type of) OS. So, reporting the software may make it almost obsolete to report the OS.
Therefore, doing a survey across different neuroimaging modalities may show some general deficiencies, but the other disciplines may need to improve on different aspects.

3. Similarly, concerning fMRI, it also makes a difference whether studies were analysed as whole-brain studies or as a focused region of interest analyses, and, in the latter case, whether the regions were derived from anatomical images or, for example, simply spheres. It would also be informative to know whether studies applied corrected p-values, and which one, and whether the effect sizes were reported.

4. Did the authors control how many studies came from the same lab? Some labs might have a kind of “tradition” in reporting results, which could bias the survey.

5. I think, the headlines of the article are a bit off since the “Discussion” mostly reports the results, and the “Conclusion” primarily discusses the results.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroimaging, fMRI, MRS, reliability

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 27 Jan 2021

David Kennedy, University of Massachusetts Medical School, Worcester, USA

We would like to thank the reviewer for their thoughtful comments and attention to this manuscript. Here we outline our responses to the comments, and indicate where in the manuscript we have made changes.

- [...] the authors focus too much on the technical aspects of a study. Although the authors introduce that a “spectrum-diagnosis” might generate further problems, they do not follow this up in the survey. I would like to see at least one additional column that codes whether the diagnostic criteria and sample are replicable, i.e. are the patients well characterised (age, gender, education), are the diagnostic instrument mentioned, cut-off criteria, etc.

Response: The survey is indeed focused on the technical ability to reproduce the analytic approach of a study. We try to do a better job of setting up the scope of the survey in the introduction. Specifically, we clarify that there are at least three domains in a publication where sufficient information for re-execution needs to be considered: the subject selection (can another researcher generate a comparable group); the data acquisition (can another researcher collect the same data); and the analysis (can another researcher perform the same analysis). All of these areas are important but we are only addressing the ‘analysis’ aspect in this manuscript. Development and application of a similar re-ascertainment survey for the subject selection is an excellent idea, and we hope to pursue such an endeavour in a future survey that would look at subject characteristics such as age, gender, education, diagnostic instruments, etc.
We have augmented the second paragraph of the Introduction to address this (and a

similar concern raised by Reviewer #2):

"In this paper we concentrate on assessing the technical prospects of re-executability of a publication. As introduced above, there are many other factors that will contribute to the actual generalization of the findings including subject population details, data acquisition details, the nature of the processing and statistics (even if they can be re-executed), the underlying biological effect size, if present, etc. (see Figure 1). Take for example, the subject population. Too often researchers communicate a finding based on a convenience sample without any statement indicating that the results might not generalize to a sample that more accurately reflects human diversity (e.g. DeJesus, Callanan, Solis & Gelman, 20191; Hruschka, Medin, Rogoff & Henrich, 20182; Rad, Martingano & Ginges, 20183; Henrich, Heine, & Norenzayan, 20104). Comprehensive and standardized description of all these additional factors are critical as well, but are beyond the scope of this evaluation. Our groups and others are looking into reporting standards for these areas as well."

- I suggest including another column (at least) in the supplementary material S3 that also lists the imaging modality, i.e. structural MRI, fMRI, MRS, DTI, since they also partly represent different disciplines and traditions in publishing.

Response: We have added a new column to Table 1 that indicates modality. While the details of the data and analysis procedures will vary by these modalities, the need to fully express the complete analysis should be independent of the specific modality.

- some methods have only a very limited number of software tools, like MRS, which are often restricted to only one (type of) OS. So, reporting the software may make it almost obsolete to report the OS.

Response: While this is certainly true in some situations, we suggest that a good best practice for reporting should be universal (and OS versions change and thus should be disclosed). We have added in the Discussion, third paragraph:

"Even if there are currently only limited software options in some analysis domains, which may implicitly implicate the operating system used, such limitations are not guaranteed to persist through time and should not be assumed for the reader."

- doing a survey across different neuroimaging modalities may show some general deficiencies, but the other disciplines may need to improve on different aspects.

Response: Again, while this is true, the general best practices and principles we're trying to elucidate here should be universal. What specific disciplines need to do to support these necessary practices may indeed vary by discipline. We try to elaborate on this in the Limitations section, first paragraph:

"We acknowledge that the details of precise description and dissemination of data and methods may indeed vary by discipline. However, we argue that the 'best practice' principles that we are suggesting here are universal and domain-specific solutions are currently available."

- concerning fMRI, it also makes a difference whether studies were analysed as whole-brain studies or as a focused region of interest analyses, and, in the latter case, whether the regions were derived from anatomical images or, for example, simply spheres.

Response: This specific factor is accounted for in the assessors interpretation of how confident they are about re-executing the procedure. If a focused region of interest study is reported, the assessor will determine how confident they are that they could arrive at the ROIs used.

- It would also be informative to know whether studies applied corrected p-values, and which one, and whether the effect sizes were reported.

Response: This is an important distinction that we did not clarify in the original manuscript. Our assessment is aimed at evaluating the quality of the reporting (can I do what was reported), rather than the content (is what was reported the right or best thing to do?). The latter assessment is really the purview of the original reviewers of the article itself, whereas the former (in other words, an attempt to generalize a reported finding) is a function that the community of readers would be engaged in and hence our assessment of the feasibility of this from the article. We've added a statement at end of the Survey Design section:

"Note that our assessments are not if the analysis or data accessibility is 'optimum', or even 'correct', but rather if the assessor could redo the approach as described."

- Did the authors control how many studies came from the same lab? Some labs might have a kind of "tradition" in reporting results, which could bias the survey.

Response: This is a fair point. We have reviewed the author lists of the articles included in the survey and indeed discovered that a number of these articles come from the same groups. This is now explicitly documented in the first paragraph of the Limitations section.

"Also, even though fifty publications are included in the survey, a number of these publications share co-authors or originate from the same research groups. Specifically, 15 of these authors are listed on two or more publications, and 14 of the publications have authors that are also authors on other publications in this set."

- the headlines of the article are a bit off since the "Discussion" mostly reports the results, and the "Conclusion" primarily discusses the results.

Response: The Results, Discussion and Conclusions sections have been updated to better reflect the appropriate content.

Competing Interests: No competing interests were disclosed.