Biological Psychiatry: CNNI

Commentary

What Is Not Measured Cannot Be Counted: Sample Characteristics Reported in Studies of Hippocampal Volume and Depression in Neuroimaging Studies

Katherine M. Keyes, Noah T. Kreski, Victoria A. Joseph, Ava D. Hamilton, Mark L. Hatzenbuehler, Katie A. McLaughlin, and David G. Weissman

The field of population neuroscience is expanding (1), with increasingly large samples, geographic and demographic diversity of participants, and longitudinal studies accelerating the progress of the field. As progress accelerates, there is an increasing ability to consider how social context and structural factors, as well as sampling methods and sample composition, modify associations from neuroimaging studies and influence the magnitude of observed effect sizes across studies.

The early-life social environment fundamentally shapes brain development and mental health. Substantial evidence demonstrates that identity, wealth, resources, oppression, and exposure to stress matter for health and brain development. Factors such as socioeconomic status and experiences of adversity have long been studied as determinants of neural development (2), and an increasing body of work has identified broader structural determinants and societal arrangements (e.g., structural forms of stigma, state-level antipoverty policies) as critical determinants of risk for brain health and development (3,4). Indeed, in one recent study, we found that the association between family income and youth hippocampal volume varied significantly across U.S. states, and that among states with high costs of living, more generous benefits for families of lower socioeconomic status reduced the influence of family income on hippocampal volume (4).

Yet our ability to detect signals of risk across multilevel and multilayered determinants is bounded by the characteristics of the samples that we recruit into our studies, the number of sites included in the study and their geographic dispersion, how measures are constructed and reported, and the generalizability of collected samples to relevant target populations. Our work and others' have found that observed associations in neuroimaging samples change as a function of sociodemographic sample distributions (5–7), underscoring the importance of clearly identifying target populations of interest for neuroscience hypotheses and conducting representative sampling from those populations. Key to identifying the structural determinants of neural development is sufficient diversity within samples, as well as diversity in the number of sites that are included in neuroimaging studies.

One example of a commonly studied association in neuroimaging studies with potential variation across samples based on sociodemographic characteristics and environmental exposures is the link between hippocampal volume and depression, which is among the most replicable neurobiological signals within biological psychiatry identified through neuroimaging. Hippocampal volume in both the left and right hemispheres is reduced among people with active major depression (8,9). Although mechanisms underlying this association remain speculative, establishing validity, temporality, and long-term consequences for cognitive aging are important areas for biological psychiatry, given the significant role of the hippocampus in learning and memory.

Yet across existing reviews and meta-analyses, heterogeneity in the association between hippocampal volume and depression has been noted across patients at different ages, or with different histories of depressive episodes, treatment, and other factors. Variation in the magnitude of the association between hippocampal volume and depression suggests that there are important moderators of this relationship. Such moderators likely include aspects of identity and environmental experience that shape psychiatric risk as well as the conditions under which the brain develops. Sample composition may contribute to meaningful differences in the resulting effect size, but to our knowledge this has never been examined. Such variation is important to identify to reconcile heterogeneous findings across samples.

We undertook a systematic review and meta-regression approach with the aim of identifying whether and to what extent variation in sample composition and social context is associated with the magnitude of effect size in the relation between hippocampal volume and depression. We searched review articles through September 2021. Inclusion criteria were patients with a primary diagnosis of major depressive disorder assessed using international diagnostic criteria (DSM-5 or ICD-10), a comparison group of nonaffected individuals, magnetic resonance imaging as the primary measurement tool, and a continuous measure of hippocampal volume as the dependent variable. Exclusion criteria included patients presenting any other neuropsychiatric or metabolic condition. Individual articles on hippocampal volume and depression were extracted, and an electronic search of published literature was supplemented by hand searching literature review and reference lists. Article abstracts for individual studies were screened independently by 3 authors (NTK, ADH, VAJ), and those moving to full-text screen were also assessed by the 3 authors. We then extracted data on depression measures, hippocampal volume,

  https://doi.org/10.1016/j.bpsc.2023.01.007
Biological Psychiatry: Cognitive Neuroscience and Neuroimaging May 2023; 8:492–494 www.sobp.org/BPCNNI ISSN: 2451-9022

intracranial volume (ICV), the reported sample distributions of sex/gender, race and ethnicity, parent socioeconomic status (targeting measures of parental education, income, and employment), and location of data collection (city, state or other geographic identifier, and country).

Figure 1 describes our search strategy and the number of final studies identified. Of the 335 studies screened, 235 were selected for full-text review. We included studies for which there was sufficient information, accounting for ICV, to calculate a standardized effect size that could be compared across studies, whether through *F* tests or adjusted mean differences. Of the 235 studies screened for full-text review, 52 reported an association between depression and hippocampal volume, but only 20 controlled for total ICV and had sufficient information to report an effect size. Among the remaining 20 studies, all reported on sex or gender (though inconsistently), 15 had identifiable sample geography (though typically only at the level of country or continent), 12 reported parental education, and 3 reported race/ethnicity. No other demographic factors appeared with sufficient regularity for analysis.

Of these 20 studies, sample composition was rarely reported in a way that could be used for metaregression. The only reliable measure of sample composition that was routinely reported was the sex/gender distribution (percent male and female), but reporting of this variable was inconsistent. We examined whether sex/gender distribution modified the effect size of depression on hippocampal volume. In the 20-study sample, percent female was associated with a lower effect size (left hippocampal volume: B = −0.01 [95% CI −0.03 to 0.00], p = .07; right hippocampal volume: B = −0.01 [95% CI −0.02 to 0.00], p = .16); however, associations were not statistically significant and were skewed by one study with a 100% female sample (N = 41 respondents) that had one of the highest effect sizes across studies.

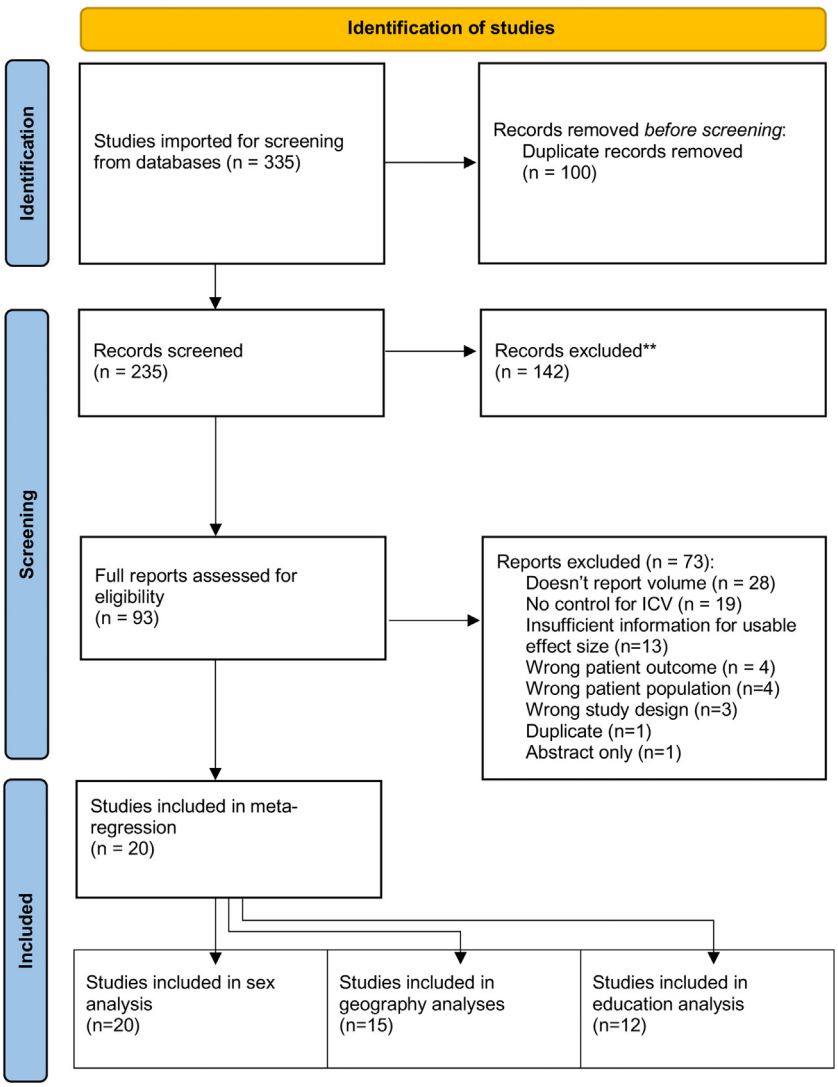

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram for systematic review of hippocampal volume and depression studies. PRISMA flow diagram for systematic review with search term strategy as follows: PubMed, Embase, PsycINFO, and Cochrane Library addressing hippocampal volume and depression using the following search terms: ("unipolar depression" OR "affective disorder" OR "depression" OR "MDD" OR "depressive" OR "depressed" OR "major depressive disorder" OR "mood disorders" OR "unipolar depression") AND ("hippocampus" OR "hippocampal" OR "hippocampi" OR "grey matter" OR "gray matter" OR "white matter") AND ("magnetic resonance imaging" OR "MRI" OR "imaging" OR "morphometry" OR "structural magnetic resonance" OR "functional magnetic resonance" OR "fMRI" OR "BOLD fMRI" OR "magnetic resonance neuroimaging" OR "tractography" OR "diffusion tensor imaging" OR "DTI" OR "fractional anisotropy" AND "volume" AND (English[language]) filters: humans). ICV, intracranial volume.

No other sample compositional characteristic was reported consistently enough for thorough analysis. Only 3 studies reported race or ethnicity, and 12 studies reported a single measure of socioeconomic status (i.e., parental education). No other measures of socioeconomic status were routinely reported. Samples were geographically skewed toward Europe (n = 6 of 15) and North America (n = 5 of 15), with fewer in Asia/Australia (n = 4 of 15), and no studies identified were from Africa or South America. Regional comparisons and assessment of parental education were underpowered, and we were therefore not able to assess moderation by social context or socioeconomic status, respectively.

In summary, we were not able to conduct a rigorous meta-regression of whether sample composition or social context moderates the association between depression and hippocampal volume because the available studies simply do not report the details about their sample in ways that would allow researchers to investigate this question. This was the case for one of the most well-studied neuroimaging associations in the literature, and certainly suggests that investigations of other associations may encounter the same issues. While part of the lack of studies was due to other methodological issues (e.g., not accounting for ICV), consistent, thorough characterization of study samples is an urgently needed step toward improving population neuroscience.

Our failure to be able to conduct an analysis of how these sampling characteristics, including socioeconomic status, modify associations between depression and hippocampal volume implies that researchers recruiting these samples may not consider how sociodemographic factors and social context may influence the associations reported. This has important implications for how bodies of evidence are used for translational science that bridges neuroimaging discovery with public health and intervention. Neuroimaging studies, especially when recruiting from the general population, are generally a high bar for participation for most families (10), and participants with high levels of family education and resources are overrepresented in neuroimaging samples (10). The lack of diversity in sample composition across studies limits the generalizability of the findings that emerge in numerous ways. First, within individual studies, reporting sociodemographic characteristics and maximizing sociodemographic diversity will provide transparency and allow more power for testing potential social and structural mediators. Without attention to reporting of demographics and the implications of those demographics for associations with social context, the robustness of social and structural hypotheses cannot be confirmed. Second, to identify the impact of structural indicators of social inequality, samples need to be recruited and/or harmonized across geographic space; samples recruited from one area, or only urban areas, for example, are largely invariant to the structural aspects of the social environment (4). Maximizing the number of sites used in neuroimaging studies will allow more power for testing potential social and structural mediators.

The only characteristic we were able to routinely capture—sex/gender composition—did suggest potential moderation in the effect size of the association between hippocampal volume and depression, suggesting that sex-specific mechanisms should be investigated more thoroughly. Were we able to report moderation (or lack thereof) for other characteristics, new hypotheses and research directions may also have emerged. The lack of routine reporting of basic sociodemographics in neuroimaging samples limits scientific progress. This basic reporting of sample composition should be required for publication in biological psychiatry and neuroimaging journals.

Acknowledgments and Disclosures

This work was supported by National Institute on Drug Abuse Grant No. R01DA048853 (to KMK).

The authors report no biomedical financial interests or potential conflicts of interest.

Article Information

From the Department of Epidemiology, Columbia University Mailman School of Public Health, New York, New York (KMK, NTK, VAJ, ADH), and the Department of Psychology, Harvard University, Cambridge, Massachusetts (MLH, KAM, DGW).

Address correspondence to Noah T. Kreski, M.P.H., at ntk2109@cumc.columbia.edu.

Received Jan 19, 2023; accepted Jan 19, 2023.

References

1. Falk EB, Hyde LW, Mitchell C, Faul J, Gonzalez R, Heitzeg MM, *et al.* (2013): What is a representative brain? Neuroscience meets population science. Proc Natl Acad Sci 110:17615–17622.
2. Tooley UA, Bassett DS, Mackey AP (2021): Environmental influences on the pace of brain development. Nat Rev Neurosci 22:372–384.
3. Hatzenbuehler ML, Weissman DG, McKetta S, Lattanner MR, Ford JV, Barch DM, McLaughlin KA (2022): Smaller hippocampal volume among black and latinx youth living in high-stigma contexts. J Am Acad Child Adolesc Psychiatry 61:809–819.
4. Weissman DG, Hatzenbuehler ML, Cikara M, Barch D, McLaughlin KA (2021): Antipoverty programs mitigate socioeconomic disparities in brain structure and psychopathology among US youths. PsyArXiv. https://doi.org/10.31234/osf.io/8nhej.
5. LeWinn KZ, Sheridan MA, Keyes KM, Hamilton A, McLaughlin KA (2017): Sample composition alters associations between age and brain structure. Nat Commun 8:874.
6. Blanco C, Wall MM, Lindquist MA, Rodríguez-Fernández JM, Franco S, Wang S, Olfson M (2016): Generalizability of neuroimaging studies in 5 common psychiatric disorders based on the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). J Clin Psychiatry 77:e1618–e1625.
7. Charpentier CJ, Faulkner P, Pool ER, Ly V, Tollenaar MS, Kluen LM, *et al.* (2021): How representative are neuroimaging samples? Large-scale evidence for trait anxiety differences between fMRI and behaviour-only research participants. Soc Cogn Affect Neurosci 16:1057–1070.
8. Malykhin NV, Coupland NJ (2015): Hippocampal neuroplasticity in major depressive disorder. Neuroscience 309:200–213.
9. Amidfar M, Quevedo J, Réus GZ, Kim Y-K (2021): Grey matter volume abnormalities in the first depressive episode of medication-naïve adult individuals: A systematic review of voxel based morphometric studies. Int J Psychiatry Clin Pract 25:407–420.
10. Garcini LM, Arredondo MM, Berry O, Church JA, Fryberg S, Thomason ME, McLaughlin KA (2022): Increasing diversity in developmental cognitive neuroscience: A roadmap for increasing representation in pediatric neuroimaging research. Dev Cogn Neurosci 58:101167.