Letters

COMMENT & RESPONSE

In Reply Our recent article[1] elicited letters raising several issues. Wright and colleagues provide important criticisms of the framing of our article. We agree that *diagnosis* is a more accurate description of phecodes derived from *International Statistical Classification of Diseases, Ninth and Tenth Revisions, Clinical Modification* (*ICD-9/10-CM*), billing codes in electronic health records (EHRs), and *prevalence of diagnoses in EHRs* is more accurate than *prevalence*. Though some recent work has shown that using 2 or more EHR codes has good sensitivity for a confirmed diagnosis of certain severe disorders,[2] we cannot ensure this is the case across all diagnoses and populations. There is likely rampant underdiagnosis of heavily stigmatized disorders, as conditions like alcohol use disorders are generally missed in routine medical care,[3] and few patients with alcohol use disorders go on to receive treatment.[4] The goal of our analysis was to provide an overview of available information in All of Us, not to infer that these rates reflect patterns in the broader population.

In response to Kuplicki and Thompson, we have revised our query of the All of Us database and identified 214 206 participants with available EHR data using codes developed by the All of Us workbench team. Our previous report had erroneously included 331 380 participants. In addition, we expanded the EHR query to include all records under the parent Observational Medical Outcomes Partnership Common Data Model categories of drug-related disorders, mental disorders, substance abuse, sleep disorders, and mental state findings. We converted these codes to phecodes using *ICD-9/10-CM* codes. As Kuplicki and Thompson noted, All of Us used additional source vocabularies (eg, SNOMED). However, when we filtered to unique identifiers across each of the source vocabularies, *ICD-9/10-CM* codes were the majority (86.9%) (**Table**). Some participants had EHR data from multiple source vocabularies, and 10 579 (75.9%) of those with a "missing" source vocabulary identifier had corresponding *ICD-10-CM* codes and concept descriptions. We therefore used all participants with available *ICD-9/10-CM* codes and converted them to phecodes as this allowed us to group various codes into parent concepts and make our outcomes comparable with ongoing analyses in other biobanks.[2]

With this revised analysis, we identified increases in the prevalence of diagnoses compared with what we originally reported. For example, we now report a prevalence of diagnoses for any mood disorder of 22.14% (95% CI, 21.17%-22.52%) vs our initial prevalence of 11.0% (95% CI, 10.68%-11.32%), and we identified 35 776 participants (16.70%) with a major depressive disorder diagnosis, compared with 30 544 (9.3%) originally, which is comparable with rates in the general US population. Our original findings indicated that the prevalence rates among the All of Us cohort were generally lower than that of the general population. In our revised analysis, estimates for many of the diagnoses were lower than nationally representative population-based estimates, except those for mood disorders, sleep disorders, and schizophrenia. We attribute the differences between our corrected and original estimates and between those by Kuplicki and Thompson to our use of the PheWAS statistical package (as detailed in the article Supplement), which may need maps expanded to additional source vocabularies, similar to other programs.[5] We included more details of our analytic choices and their limitations in the corrected Supplement. We believe this can serve as a useful guide for readers who want to use this relatively new data resource. Importantly, while we were able to identify more participants with a diagnosis, we observed similar patterns of comorbidity and relationships with sociodemographic correlates.

Table. Unique Identifiers by Source Vocabulary

Source vocabulary	Unique identifiers, No. (%)
ICD-10-CM	101 016 (52.7)
ICD-9-CM	65 566 (34.2)
SNOMED	9781 (5.1)
Missing	13 938 (7.3)
None	1248 (0.7)
Other	50 (0.03)
Total	**191 599 (100)**

Abbreviations: *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification*; *ICD-10-CM*, *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification*; SNOMED, Systematized Nomenclature of Medicine.

We are grateful for the authors' comments on our article. All of Us provides a rich source of data with various nuances to be considered. We hope this lesson is taken to heart by all those interested in working with this new and complex resource. We have requested that our original article be corrected to include the term *diagnoses* vs *disorders*, our revised methods of analysis, and corrected estimates of prevalence of diagnoses in the title, text, Tables, Figure, and Supplement. We apologize to the readers of *JAMA Psychiatry* for any confusion this has caused. The article has now been corrected.[6]

Peter B. Barr, PhD
Tim B. Bigdeli, PhD
Jacquelyn L. Meyers, PhD

Author Affiliations: Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, New York (Barr, Bigdeli, Meyers); VA New York Harbor Healthcare System, Brooklyn (Barr, Bigdeli, Meyers).

Corresponding Author: Peter B. Barr, PhD, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, 450 Clarkson Ave, MSC 1203, Brooklyn, NY 11203 (peter.barr@downstate.edu).

Published Online: September 14, 2022. doi:10.1001/jamapsychiatry.2022.2739

Conflict of Interest Disclosures: None reported.

1. Barr PB, Bigdeli TB, Meyers JL. Prevalence, comorbidity, and sociodemographic correlates of psychiatric diagnoses reported in the All of Us research program. *JAMA Psychiatry*. 2022;79(6):622-628. doi:10.1001/jamapsychiatry.2022.0685

2. Bigdeli TB, Voloudakis G, Barr PB, et al. Penetrance and pleiotropy of polygenic risk scores for schizophrenia, bipolar disorder, and depression in the VA health care system. *medRxiv*. Published online January 1, 2022. doi:10.1101/2022.02.16.22271088

3. Smothers BA, Yahr HT, Ruhl CE. Detection of alcohol use disorders in general hospital admissions in the United States. *Arch Intern Med*. 2004;164(7):749-756. doi:10.1001/archinte.164.7.749

4. Mekonen T, Chan GCK, Connor J, Hall W, Hides L, Leung J. Treatment rates for alcohol use disorders: a systematic review and meta-analysis. *Addiction*. 2021;116(10):2617-2634. doi:10.1111/add.15357

5. Zheng NS, Feng Q, Kerchberger VE, et al. PheMap: a multi-resource knowledge base for high-throughput phenotyping within electronic health records. *J Am Med Inform Assoc*. 2020;27(11):1675-1687. doi:10.1093/jamia/ocaa104

6. Errors in terminology, data query, and prevalence rates. *JAMA Psychiatry*. Published online September 14, 2022. doi:10.1001/jamapsychiatry.2022.2817