MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220281

Race and Racialization in Mental Health Research and Implications for Developing and Evaluating Machine Learning Models: A Rapid Review

Marta M. Maslej[a], Nelson Shen[b,c], Iman Kassam[b], Terri Rodak[d], Laura Sikstrom[a]

[a]*Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health (CAMH), Toronto, Canada*
[b]*Campbell Family Mental Health Research Institute, CAMH, Toronto, Canada*
[c]*Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada*
[d]*CAMH Library, Department of Education, CAMH, Toronto, Canada*

Abstract

Machine learning models are often trained on sociodemographic features to predict mental health outcomes. Biases in the collection of race-related data can limit the development of useful and fair models. To assess the current state of this data in mental health research, we conducted a rapid review guided by Critical Race Theory. Findings reveal limitations in the measurement and reporting of race and ethnicity, potentially leading to models that amplify health inequities.

Keywords:

Continental Population Groups; Mental Health; Machine learning.

Introduction

Machine learning (ML) models are increasingly being developed to predict mental health outcomes, such as treatment response in Major Depressive Disorder (MDD) [1]. However, their performance relies heavily on the quality of data they are trained on. When biases and inaccuracies are introduced into data, ML models can perpetuate health disparities for disadvantaged groups [2].

Race or ethnicity have emerged as predictors of MDD outcomes in some studies, making them candidate features for ML modelling [1]. Cognitive responses to racism have been implicated in MDD, particularly as they relate to gender and socioeconomic status [3]. The observed impacts of demographic factors on health outcomes have prompted efforts to rethink the use of racial and ethnic identifiers in clinical research.

Yet, a 2008 review of 421 mental health studies concluded that the operationalization of race or ethnicity was superficial, vague, and simplistic. These variables were rarely defined and consistently used as proxies for other constructs [4]. Given ML models are often trained on demographic features, this inaccurate operationalization presents a barrier to developing useful and fair models. Furthermore, minority groups often have less training data available, which may account for more prediction errors in racialized or low-income samples [5]. This bias might be particularly apparent for groups defined by intersecting features, such as sex and race.

Rapid advancements in ML make it important to revisit how race or ethnicity are measured and operationalized in mental health research, since biases can be amplified when baked into ML data and models. According to Critical Race Theory (CRT) [6], racialization (or the social process of dividing people into different groups) is at the root of health disparities, not race as a biological factor. Thus, race should not be used as a proxy for racism. Instead, CRT recommends measuring racialization (e.g., discrimination) and identities based on intersecting features to identify at-risk populations. Additionally, relevant stakeholders should be involved to define racialization within their communities. Our study is the first to use CRT principles to guide a review of race and ethnicity in recent mental health research.

Methods

We conducted a rapid review to characterize the current state of collecting race, racialization, and ethnicity data in the context of MDD. This review updated findings from previous work [4], providing current evidence to support policy and decision making [7]. We searched Medline, PsycInfo and CINAHL databases, using subject headings, keywords, and Boolean logic to search for MDD and race/ethnicity. The search was executed in June 2020 and limited to primary studies of adults published in English after 2005. Articles were selected via a title and abstract screen, followed by full text review. Articles were included if they focused the relation between MDD and race/ethnicity. To facilitate timely review, we excluded studies of samples with co-morbid physical conditions. Drawing on CRT [6], we developed a template [8] to extract a range of study features. We report preliminary results relevant for ML modelling, i.e., operationalization of race/ethnicity, measurement of racialization, use of proxy indicators, and intersectional analyses.

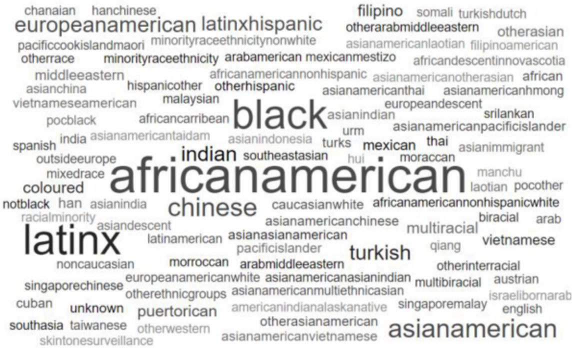
Results

The search yielded 10,467 citations (5,892 were unique), of which 975 were selected for full-text review and 207 were eligible for analysis. Most studies (71%) were conducted in the US, followed by the Netherlands (6%), Canada (3%), Malaysia (3%), and other countries. Sample sizes ranged from 12 – 807,048 (M = 9391, SD = 58,866, Median = 381).

Approximately one-third (35%) of studies did not pro-vide clear definitions of race or ethnicity, which were sometimes used as proxy indicators for racism or dis-crimination (in 16% of studies), nationality or immigration (16%), or culture (11%). Only half of the studies (52%) measured exposure to racialization or discrimination, typically at an interpersonal level (in 81% of these studies). Only 8% of studies involved racialized samples in the research process, and at least 60% of studies did not focus on intersectional identities.

Across the 207 studies, 267 unique terms described racial or ethnic groups (Figure 1). African American and White samples were most common (18% each), followed by Latinx (15%), Black (14%), and Other (14%). Most studies (73%) relied on self-reported race or ethnicity, but many (21%) did not describe how this construct was measured. Only 22% of studies provided criteria for excluding certain groups, which was often a research focus on specific racial or ethnic samples. Of studies that provided this criterion, 24% excluded groups (typically indigenous individuals) due to small samples.

Figure 1



Note. 100 of 267 unique descriptors are shown, with larger terms indicating more frequent use across studies.

Conclusions

Preliminary findings from our rapid review highlight persisting limitations in mental health research on how race-related data is collected, operationalized, and reported, which can present barriers to ML modelling. Descriptions of racial or ethnic groups were diverse, making it difficult to compile data and integrate findings across studies. Marginalized groups were often excluded due to small sample sizes, suggesting that these groups may be underrepresented in ML models. Often, there was a lack of clarity around the source of demographic data, making it difficult to gauge its quality. These findings highlight a need for consensus on the operationalization of race/ethnicity, and better reporting practices.

Contrary to CRT [6], most studies did not measure racialization, sometimes using self-reported race as a proxy for interpersonal discrimination. Most studies also did not focus on the intersection of race with other identities, like gender, which limits the identification of vulnerable subgroups. If race is related to MDD and other mental health conditions via chronic exposures to interpersonal or systemic racism, ML-based predictions of outcomes could improve if racialization were used as a feature, rather than membership to a broad group (e.g., African American, White, Latinx). The most commonly-used measure of racialization was the Everyday Discrimination Scale, but many others exist. Although all studies focused on race or ethnicity as related to MDD, few consulted racialized populations in the research process.

In conclusion, there is an urgent need to improve the collection of race-related data in mental health research. The absence of accurate data limits the development of useful and fair ML models. Without this improvement, we are more likely to amplify rather than resolve health inequities with technologies intended to support the mental health of at-risk groups.

Acknowledgements

This study was funded by the Canadian Institutes of Health Research Health Systems Impact Fellowships.

References

[1] K. Perlman, Benrimoh, D., ... and M. T. Berlim, A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder, J Affect Disord 243 (2019), 503-515.

[2] L. C. Brewer, K. L. Fortuna, … and L. A. Cooper, Back to the future: Achieving health equity through health informatics and digital health, JMIR mHealth uHealth 8 (2020), e14512.

[3] L. K. Hill, and L. S. Hoggard, Active coping moderates associations among race-related stress, rumination, and depressive symptoms in African American women, Dev. Psychopath 30 (2018), 1817.

[4] S. Møllersen and A. Holte, Ethnicity in mental health research: A systematic review of articles published 1990–2004, Nord J Psychiat 62 (2008), 322-328.

[5] I. Y. Chen, P. Szolovits, and M. Ghassemi, Can AI help reduce disparities in general medical and mental health care?, AMA J Ethics 21 (2019), 167-179.

[6] C. L. Ford and C. O. Airhihenbuwa, Critical race theory, race equity, and public health: Toward antiracism praxis, Am J Public Health 100 (2010), S30-35.

[7] M. J. Grant and A. Booth, A typology of reviews: An analysis of 14 review types and associated methodologies, Health Info Libr J 26 (2009), 91-108.

[8] Ethnicity, race, and racialization: A rapid review protocol. OSF (2020), https://osf.io/e2a6p/

Address for correspondence

Laura Sikstrom
1025 Queen St – 2341, Toronto ON M6J 1H1
Telephone: (416) 535-8501 x30082
Laura.Sikstrom@camh.ca