Racial, Ethnic, and Geographic Diversity in Population Neuroscience

C. Elizabeth Shaaban[1,2], **Andrea L Rosso**[1]

[1]Department of Epidemiology, School of Public Health, 130 DeSoto St, University of Pittsburgh, Pittsburgh, PA, USA, 15261

[2]Alzheimer's Disease Research Center, University of Pittsburgh, UPMC Montefiore Suite 421, 200 Lothrop Street, Pittsburgh, PA, USA, 15213

Abstract

In this chapter, we consider lack of racial, ethnic, and geographic diversity in research studies from a public health perspective in which representation of a target population is critical. We review the state of the research field with respect to racial, ethnic, and geographic diversity in study participants. We next focus on key factors which can arise from lack of diversity and can negatively impact external validity. Finally, we argue that the public's health, and future research, will ultimately be served by approaches from both recruitment and representation science and population neuroscience, and we close with recommendations from these two fields to improve diversity in studies.

INTRODUCTION

Inclusion of diverse samples within research studies is a critical goal for ensuring justice, trust from communities, and validity of research findings. Justice is a key principle of human-subjects research as outlined in the Belmont report and requires that the risks and benefits of research are fairly distributed.[1,2] Further, trust in research findings and willingness to adhere to recommendations is eroded when members of the public do not see themselves reflected in research samples.[3,4] Importantly, validity of findings can also be compromised when samples lack diversity that reflects the target population for potential intervention.[2,3,5,6] Here, we focus on the impact of lack of diversity on external validity, or the ability to apply research findings to populations of interest. We consider diversity across racial/ethnic identity, geographic location (e.g., rural vs urban), and country of residence (e.g., high-income countries (HICs) vs low- and middle-income countries (LMICs)), though many more dimensions of diversity may be relevant to consider.

Research in population neuroscience fuses the knowledge and methods of brain, behavior, and population sciences to make inferences about exposure-outcome relations. Its goal is to understand better variation in development, aging, and disease related to the brain at the population level. These inferences are only useful for public health and clinical

Corresponding Author: C. Elizabeth Shaaban, PhD, MS, MPH, Department of Epidemiology, School of Public Health, 130 DeSoto Street, A547 Public Health Building, University of Pittsburgh; Pittsburgh, PA 15261; Beth.Shaaban@pitt.edu.

implementation if the findings are relevant to a target population of interest. Often, the target population for a research study is not explicitly defined and/or is presumed to be universal, regardless of how selected the study sample may be. If the study sample does not represent the intended target population, then results may have limited usefulness for informing implementation strategies. This is true even when the internal validity of the study is high. Therefore, ensuring diversity across a number of characteristics relevant to the target population is critical as is thorough characterization of included samples to understand to which population(s) inferences can be made. Without representative samples, we risk producing scientific findings with limited usefulness in the real world. We refer the reader to another chapter in this book (Population neuroscience: understanding concepts of generalizability and transportability and their application to improving the public's health) for a detailed introduction to issues of internal and external validity.

Box. Definitions of key terms

- Differential item functioning (DIF) – differences in how assessment items are interpreted across subgroups, which indicate that the construct is not being measured similarly across samples;

- Effect measure modification – difference in association between an exposure and outcome based on another variable; may be a difference (indicated by a significant interaction term) in direction of effects (positive in one group, negative in another), significance (stratified results that are significant in one group but not in the other), or in magnitude of effect (associations are significant and in the same direction in both groups but larger in one compared to the other); see also the other chapter in this book indicated above;

- External validity – extent to which study results represent the truth in the source population (generalizability) or the truth in the target population (transportability); also the other chapter in this book indicated above;

- High-income countries (HICs)—countries with a per capita gross national income $13,846[7]

- Internal validity – extent to which study results represent the truth within the study sample; also the other chapter in this book indicated above;

- Low- and middle-income countries (LMICs)—countries with a per capita gross national income $13,845[7];

- Population attributable fraction (PAF) – proportion of disease cases within a population that can be attributed to the exposure, or risk factor, of interest;

- Psychometric properties – the extent to which a measurement instrument is valid (accurately measures what is intended to be measured), reliable (performs consistently across time and individuals), and responsive (has the ability to detect change);

- Study sample – the participants included in a study; should be representative of the target population to which inferences are intended to be applied; see also the other chapter in this book indicated above;

- Target population – the population to which sample estimates are intended to be applied and for which inferences are made; see also the other chapter in this book indicated above.

WHERE IS THE FIELD CURRENTLY?

For much of its history, research in brain and cognitive science has focused on convenience samples, often without specifying a target population or determining representativeness. Both brain and behavioral research studies have focused on samples from countries that are Western, Educated, Industrialized, Rich, and Democratic (WEIRD).[8,9] For example, a review of cognitive psychology literature that was completed from 2003–2007 demonstrated that 96% of study participants came from Western industrialized countries and that the majority were from the United States.[9] Within those Western industrialized samples, the majority of samples were composed of undergraduate students. While such results would appropriately apply if the researchers wished to understand something specific to a country's undergraduates (e.g., how undergraduates' brains function under the influence of alcohol), these samples were not representative of the target populations of interest, which included other age groups.[9] In addition, as of 2009, 90% of neuroimaging studies were conducted in Western countries.[10]

With the growing awareness of the need for broader representation in research, several groups have recently set out to quantify the extent to which lack of diversity has been a problem to date. Below, we provide examples of this work to demonstrate the current state of research. Most of this work has focused on racial/ethnic representation, largely in the United States, with some studies assessing representation across global geographies. Reviews have not assessed rural representativeness but given that most research centers are located in more urban settings, it is likely that individuals living in urban and suburban settings are overrepresented in research studies. In addition, most studies of representativeness have been done within the field of Alzheimer's disease and related dementias (ADRD), as represented below. There is likely a similar state of lack of representation across other sources of diversity (for example, socioeconomic status, sexual orientation) and for other neurological conditions, but the evidence is not yet documented.

Several reviews have documented diversity of samples included in research on dementia[11–13], including ADRD neuroimaging studies[14] and drug trials.[15] All have demonstrated that samples are largely unrepresentative of target populations. Contemporary dementia research is almost exclusively conducted within North America and Europe (89% of reviewed studies) and the samples overwhelmingly included White (i.e. individuals of European ancestry) participants (median of 89% White participants).[11] Many studies of dementia incidence and prevalence have not conducted analyses by sex or gender (see chapter on "Sex and gender in population neuroscience") for definitions and more information), and most studies that have evaluated this have been conducted in HICs.[12,13] Neuroimaging

studies of Alzheimer's disease in the United States have also included predominantly non-Hispanic White participants. While there has been a notable increase in representation by race and ethnicity in recent years, particularly for Black/African American participants, the numbers still do not reflect population demographics.[14] Lack of racial/ethnic representation in Alzheimer's drug trials is even lower, with the median percentage of White participants being 95% and with no significant change over time.[15] The authors of this review note that eligibility criteria, including exclusions based on comorbid psychiatric or cardiometabolic conditions and requirements for a caregiver to attend study visits, were a key source of racial/ethnic underrepresentation. In addition, global diversity for drug trials was low with 79% of studies including sites in North America, 60% including sites in Europe, and few studies including sites in South America (15%) or Africa (7%) (does not total 100% because some studies included sites in more than one part of the world).

These reviews note the lack of reporting on race/ethnicity with only 22% of dementia studies[11] and 50% of Alzheimer's drug trials[15] reporting the race/ethnicity of participants. The review of Alzheimer's neuroimaging studies only included papers that reported race/ethnicity and thus excluded a large number of studies (n=1,160) for not reporting race/ethnicity.[14] Among the studies which could be included, a large number required the authors to indirectly derive race/ethnicity estimates (n=1,745/2,464 (71%)). The authors were able to do this because the papers came from large, well-published cohorts that have previously reported these data. This is in comparison to the n=719/2,464 (29%) studies that directly reported race/ethnicity data within the review. Another notable finding from these reviews is that several large cohorts served as the primary source of data for numerous published studies. For example, 3 cohorts, all based in the United States, represented 21% of study samples in overall dementia research.[11] Among the AD neuroimaging studies which indirectly reported race, 10 cohorts were the source of 94% of study samples.[14]

Additional reviews have assessed diversity for other neurologic and psychiatric outcomes and have demonstrated a lack of diversity similar to that seen for ADRD research. Cognitive neuroscience studies underreport race/ethnicity and socioeconomic status, with only 14% and 18% of studies, respectively, providing descriptive data on these demographic characteristics.[16] Based on the few studies reporting on race/ethnicity, the authors demonstrate that cognitive neuroscience studies are much less diverse than, and not representative of, the target population. Two reviews assessed global diversity in studies of Parkinson's genetics in individuals of non-European descent[17] and in pediatric psychiatric disorders in LMICs.[18] Even among studies conducted in samples of individuals with non-European ancestry, representation may be limited. The review of Parkinson's genetic studies demonstrated that the majority of studies in those without European ancestry were conducted with Asian participants from Greater China (57%) with far fewer conducted in sub-Saharan Africa (4%), Southeast Asia (3%), or Central Asia (0.5%). Similarly, a review of studies of psychiatric disorders in children and adolescents in LMICs identified only 6 studies from 4 countries, including 3 from Brazil, 1 from China (Hong Kong), 1 from South Africa, and 1 from Mauritius.[18] Clearly, research across multiple disorders is not representative of the global population and even in countries such as the United States that are well-represented in research, there are large portions of the population that are underrepresented.

WHAT PROBLEMS ARISE FROM LACK OF DIVERSITY?

There are several common circumstances in which inferences from a non-diverse, non-representative sample could be invalid when applied to the target population, including differential prevalence of effect-measure modifiers across populations, differing ranges or prevalence of exposures, and differential item functioning (DIF) of key measures across groups (Fig. 1).

Differential prevalence of effect-measure modifiers

Presence of unaccounted for effect-measure modification of an exposure-outcome relation coupled with differences in prevalence of the effect-measure modifier by group can lead to non-comparability of findings across populations. See chapter on "Population neuroscience: understanding concepts of generalizability and transportability and their application to improving the public's health" for more detail about effect-measure modification. Briefly, effect-measure modification occurs when the relation between an exposure and outcome differs by level of a third variable—the effect-measure modifier. If effect-measure modification is not accounted for in analyses, through methods such as stratification or inclusion of interaction terms, the overall estimated effect size will be an average of the effect at all levels of the effect-measure modifier, weighted by the distribution of that effect-measure modifier within the sample. Therefore, differences in prevalence or distribution of the effect-measure modifier across populations would lead to different overall effect estimates by population.

An example of this can be seen in a comparison of strengths of exposure-outcome associations in a highly selected, clinic-based sample, the Alzheimer's Disease Neuroimaging Initiative (ADNI), and a community-based sample, Atherosclerosis Risk in Communities (ARIC).[19] The two samples differed in distribution of a number of important factors that could act as effect-measure modifiers in analyses of associations; for example, educational attainment, sex, and race. At least ¼ of tested associations varied significantly between the two studies and approximately ½ of the point estimates varied by >50%. Some of these differences were highly consequential. For example, the exposure-outcome association of *APOE4* genotype with amyloid beta (Aβ) positivity was an odds ratio (OR) of 2.8 in the community-based study but 8.6 in the clinic-based study. This suggests that a variable that modifies the effect of *APOE4* may have had a higher prevalence in the clinic-based study than in the community-based study. This type of difference matters for prioritization of intervention targets and choices about where to place public-health resources. For example, if the OR is truly 8.6 in the target population, interventionists might prioritize developing new AD interventions that target *APOE* (e.g. via modulation of APOE protein levels or gene therapy approaches), but they might make different decisions if the OR in the community is truly 2.6 while other risk-factors (e.g., physical health) have a stronger relationship with Aβ positivity. These concerns extend beyond clinic-based vs. community-based studies and are also relevant for convenience samples vs. population-representative community-based studies. A relevant example is the UK Biobank, a volunteer-based study with an impressive and well-characterized sample of 500,000 individuals from the United Kingdom. The study was highlighted in a correspondence about

non-representativeness.[20] The authors of the correspondence point out that associations can differ due to different prevalence of effect modifiers in the study sample vs. the target population (i.e. the entire UK), and that the large sample size does nothing to address this particular problem.[20]

Variation in levels of exposure

Non-linearities and range effects.—It is not unusual for levels of an exposure to vary across populations based on demographics, geography, or other characteristics. A difference in prevalence of exposure across populations does not necessarily imply a difference in the strength of association with outcomes of interest. But inferences about exposure-outcome relationships in one population may not apply to another population when there are non-linear associations across levels of exposure and levels of exposure differ substantially between those populations (Fig. 2). For example, if a particular threshold must be reached before an exposure causes disease, associations will not be observed in populations with overall low exposure levels. Further, this can cause inference problems if the range of exposure level is too narrow within a single population, even if the relation is linear. In this case, the narrow range of exposure will limit the ability to detect associations due to the lack of variance. This type of relationship is highly relevant in AD research. For example, there is a tipping point at which tau accumulation rapidly accelerates—when Aβ levels are 40 centiloids on Positron Emission Tomography (PET) imaging.[21] Because Aβ accumulation plateaus in the clinical stages of the disease, it is often found to be only weakly or not associated with cognition in that stage, while tau is strongly associated with cognition.[22,23] Researchers should report the mean and range or variance of their exposure variables within their samples to allow for comparison with other samples or with known distributions in the target population, when available. In addition, non-linearities of exposure-outcome relations should be assessed.

Population attributable fraction (PAF).—Another way in which differences in exposure prevalence can affect interpretation of results across populations is in calculation of the PAF for a given exposure or set of exposures. PAF is the proportion of disease cases within a population that can be attributed to the exposure or risk factor of interest and is calculated as

$$PAF = \frac{P_e x (RR - 1)}{P_e x (RR - 1) + 1}$$

where P_e is the prevalence of the exposure or risk factor and RR is the relative risk for the risk factor on the outcome of interest.

We discuss two examples of this—one using dementia-related PAF estimates in LMICs vs. HICs, and one using dementia-related PAF estimates across ethnoracial groups in the United States. The 2020 Lancet report on dementia prevention, intervention, and care calculated PAFs for established risk factors of dementia and suggested that at least 40% of risk for dementia is due to 12 modifiable dementia risk factors (this is the PAF summed across risk factors) while the remaining 60% is either non-modifiable or not yet identified.[24] This built

upon the group's 2017 work, which included 9 modifiable dementia risk factors and found that they accounted for an estimated 35% of dementia risk.[25] These estimates were based on meta-analyses to define both P_e and RR.

Although international studies were used to represent global estimates, it is important to point out that the studies used to obtain the P_e and RR estimates have generally been conducted in relatively homogeneous samples from HICs. Because the PAF is a function of both the P_e and RR, variations in their estimates will cause the PAF to vary. A summary of estimates of P_e, RR, and PAFs from various studies and populations may be found in Table 1. PAFs were weighted to account for correlations between risk factors and try to give an estimate of the unique contribution of each exposure to overall dementia risk. All included studies used the same RR estimates, while estimates of P_e were study specific. RRs represent a weighted average RR over relevant population subgroups. Because risk factor-dementia associations vary based on these subgroups defined by different prevalence of effect-measure modifiers, RRs will differ in populations with different distributions of subgroups.[20,26] The RRs drawn from the meta-analyses included in the 2017 and 2020 Lancet dementia reports will be relevant for global estimates to the extent that the prevalence of unmeasured effect measure modifiers is similar across those populations not represented. This assumption may be met for many variables but would probably not be met for several important variables such as age, education, or gender. Even if the strength of the association between the exposure and outcome does not differ between populations, the relative importance of the exposure as a target for public health interventions could differ substantially based on the P_e and the PAF, as demonstrated in Table 1.

LMIC vs HIC PAFs.: Based on data from the ELSI-Brazil,[27] the Brazilian member of the global Health and Retirement Studies, which was designed to be population-representative of individuals 50+ years of age in Brazil, the PAF for dementia was as 50.5%, higher than that from the 2020 Lancet report (39.7%) despite being calculated from fewer risk factors (10 vs 12).[28] The PAF estimates for the 9 dementia risk factors included in the 2017 Lancet report (35.0%) in several LMIC locations were also higher as follows: across 6 Latin American countries (Cuba, Dominican Republic, Mexico, Peru, Puerto Rico, and Venezuela), 55.8%; in India, 41.2%, and in China, 39.5%.[29] Not only are the RRs likely to have been different in these populations as discussed above, but the P_e estimates also varied substantially. For example, the prevalence of low education was 10.7% in the US, 40.0% globally, relying on mostly HIC data, and 92.2% in India.

PAFs in various US ethnoracial groups.: In the US, there are notable variations in dementia risk and risk factor prevalence between ethnoracial groups. Thus, national PAF estimates (41.0% based on 12 risk factors) do not necessarily apply to the entire US population, nor do PAF estimates for one ethnoracial group apply to others within the US.[30] For example, based on a study using data from ARIC, low education was most common for Hispanic individuals in the US and least common for non-Hispanic Asian and White individuals leading to a greater PAF due to low education and in sum for those identifying as Hispanic vs. those identifying as non-Hispanic Asian or White (Table 1).[30] Hypertension was present in 61.0% of non-Hispanic Black individuals but <40% in the other ethnoracial

groups. The consequences of this difference are a greater PAF due hypertension and in sum for those identifying as non-Hispanic Black vs. those identifying as non-Hispanic Asian or White (Table 1).[30] In addition, non-Hispanic White individuals had lower exposure to air pollution with a P_e of 17.2%, while the remaining US ethnoracial groups had P_es that were more than double that. This leads to a lower PAF due to air pollution and in sum for those identifying as non-Hispanic White vs. those identifying as Hispanic or non-Hispanic Black.

Differential item functioning (DIF) and different psychometric properties of instruments across groups

DIF indicates that an item does not measure the same construct across different groups. DIF can also impact sum-scored instruments when items within the instrument function differently; this is important for population-based health research, which often uses sum scores from established scales.[31] DIF and differences in psychometric properties of instruments across groups can influence validity of findings between convenience samples and the broader population. This can arise due to differences in interpretation of wording in questionnaires or in reaction to the content of an instrument. For example, the Center for Epidemiologic Studies Scale-Depression (CES-D), a commonly used questionnaire that generates a sum score for depressive symptomology, has been shown to have differential functioning by race,[32,33] ethnicity,[34] immigration status,[35] country of study,[36] and sexual orientation,[37] though it should be noted that many other studies have demonstrated measurement invariance of the CES-D across different groups, including those listed above. Additional examples of potentially different measurement properties by populations include autism spectrum characteristics by gender or country of origin[38,39] and suicidality by ethnicity.[40] DIF is beginning to be assessed for the tasks used in task-related functional MRI.[41] Differences in psychometric properties (i.e. validity, reliability, responsiveness) of an assessment across groups can also lead to differences in sensitivity/specificity to an underlying construct and as a result, differences in ability to detect significant associations within certain groups.

RECOMMENDATIONS TO IMPROVE DIVERSITY IN RESEARCH STUDIES

Clearly there is a lot of room to enhance diversity and external validity in research, and we now move to a discussion of recommended strategies for improvement. Two main categories of strategies can be used to enhance diversity and external validity in studies: 1) study-design approaches implemented before data collection, including specific sampling and recruitment strategies, and 2) analytic approaches that can enhance external validity in some instances when data have already been collected. Ultimately, we see enhancing diversity and external validity as requiring a merging of recruitment and representation science with population neuroscience. A number of study-design problems have limited diversification of samples, and recruitment and representation science offers strategies to address these barriers. Recruitment and representation science is a growing field focused on improving recruitment strategies that enhance inclusion and retention of individuals from underrepresented backgrounds.[42,43] This field is identifying best practices that should be consulted to improve representation of samples.

A major barrier to diversity in research has been a lack of understanding by researchers of the needs of diverse communities that would allow them to participate in research. Often, researchers have attempted to diversify their samples by continuing to utilize recruitment strategies and data collection approaches that they had been using and which have consistently resulted in the non-generalizable/non-transportable samples in prior research. This includes largely passive recruitment methods that rely on volunteers to see and respond to advertisements to participate. Location and method of recruitment should be carefully considered and sufficient resources devoted to recruitment efforts. Often, researchers do not have familiarity with the communities that have been underrepresented in research, which leads to a lack of knowledge about the needs and desires of those communities with regard to their ability, willingness, and motivation to participate in research studies. Overall, research efforts that reduce participant burden, build trust between the participant and research team, and improve participant knowledge of the research study will enhance research participation.[44] Having greater representation at both the staff and investigator level of members of these communities will greatly enhance the ability to recruit and retain diverse samples.[6,45]

The location and timing of research visits can be a barrier to participation, particularly for those from lower socioeconomic conditions, rural communities, and communities of color. Consideration should be given to reducing participant burden by providing mobile clinic sites, providing in-home visits, and/or locating clinic sites within target communities. Accessibility for those with disabilities should also be considered. Researchers should make efforts to provide flexibility in timing of visits and options outside of traditional working hours. Imaging modalities that are more mobile include electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS) and, most recently, mobile low-field magnetic resonance imaging, MRI (see chapter on "Population neuroscience: Principles and advances"). For studies using non-portable imaging modalities (e.g. 3T MRI), researchers should consider paying for or providing transportation to the study site and offering scan times outside of normal business hours. Analytic approaches (discussed below) can help with addressing the selection that often occurs in imaging studies due to the inconvenience of visits, discomfort with procedures, and safety-related exclusion criteria. Blood biomarkers of brain pathologies such as Aβ, tau, neurodegeneration, and neuroinflammation are also a promising approach to increasing inclusion in future population neuroscience research.

Unnecessarily strict exclusion criteria can also impact the diversity of individuals included in research samples.[15] All inclusion and exclusion criteria should be carefully considered with regard to their impact on external validity and their necessity for maintaining participant safety and internal validity. Communities that are underrepresented in research may be hesitant to participate in studies, lack trust in researchers, and/or lack knowledge about what it means to participate in research. In all of these cases, the burden should be on researchers to work with communities to overcome these barriers. Randomized controlled trial evidence shows that participant compensation increases recruitment.[46,47] Best practices in the Alzheimer's Disease Research Centers' Network have recently been developed and are available via the National Alzheimer's Coordinating Center website (https://files.alz.washington.edu/documentation/remuneration-guidelines.pdf). Finally, giving to the community (at least) twice over is key for engagement, recruitment, and retention: first,

before asking for research participation to help meet community needs (e.g., health fairs or educational talks), and second, after participation to return group and/or individual research results to participants and community members.[6] This return of research results and sharing educational information with participants has recently been conceptualized as return of value (ROV).[48]

To draw relevant inferences about the target population of interest, the study must include diverse participants with all of the characteristics observed in the target population. External validity will be enhanced if the study sample is population representative. There are circumstances, however, where this is neither feasible nor desirable. For example, getting valid estimates of prevalence or of exposure-outcome associations in smaller subgroups may require oversampling of that group relative to their population proportion to achieve sufficient statistical power. We discuss analytic strategies that may be used in such circumstances below. Variables to describe relevant aspects of diversity must be collected by design; this includes race/ethnicity and geographic factors as well as factors which may vary across diverse groups such as health conditions, resilience factors and coping strategies, and sociocultural, political, economic, and environmental factors. Classification of race/ethnicity and geographic factors should provide the appropriate level of specificity to capture the heterogeneity in researchers' outcomes of interest.

Analytic approaches to enhance external validity also warrant discussion for cases where data have already been collected. A key goal of some studies will be to understand how multiple factors jointly contribute to risk and resilience, and such analyses should test for additive and multiplicate interactions. Knowledge of diverse populations is enhanced when contexts (e.g., social, environmental) are considered; such studies may need to employ multilevel analytic approaches, which account for correlated data. Analytic approaches can also be used to enhance both internal and external validity in studies. Approaches such as inverse probability of selection weighting, propensity score matching, and the like can be used to weight subsamples, for example a neuroimaging sub-study, to the full parent study, reducing selection bias and thereby enhancing internal validity.[49,50] Joint modeling can enhance internal validity of longitudinal studies by addressing differential attrition due to causes such as drop out and death.[51] External validity can be improved analytically by using transportability estimators.[52] This can be thought of as an extension of survey weighting methods to weight the study sample to the population. This approach can be useful when taking a population representative sample was not feasible or was not the appropriate design for obtaining accurate estimates within smaller subgroups.

It is critical to train the next generation of researchers for community engagement and participant recruitment as well as fluent study design and analysis to advance diverse and representative research that can serve the population. Unique training approaches, such as training individuals in population neuroscience and recruitment and representation science can help to bring knowledge of representative and diverse population characteristics to research.[6,53–55] Additional funding that specifically requires representative samples can improve diversity in studies. This may be accomplished, for example, by policies such as the recent announcement that the US National Institute on Aging (NIA) will prioritize new grant submissions that represent the diversity of populations experiencing the disease of interest

and that are designed to include participants from groups experiencing health disparities.[56] This is a particularly difficult barrier in LMICs due to limited funding to conduct health research. Greater funding from organizations which do not limit investigator location can help fill the gap. Funding researchers who are themselves from groups historically underrepresented in research can improve relationships with the community and inform scientific hypotheses to test and interpretation of results.

SUMMARY AND CONCLUSIONS

There is a growing awareness of the need to discuss target populations, define them explicitly, and to sample systematically to create representativeness of research samples to those target populations. This awareness is evidenced by the development and adoption of population-neuroscience approaches[49,57–59] and a push from funding agencies to diversify research participants to create more representative samples. Our review of the state of the field shows that diversification across multiple domains of identity and geography has been slow, however. We discussed three key factors which can negatively impact external validity including differential prevalence of effect modifiers across populations, variations in levels of exposures, and differential item functioning and differing psychometric properties of instruments across groups. We discussed approaches to enhance diversity in studies and suggest that a merging of perspectives from recruitment and representation science and population neuroscience can best increase diversity and improve external validity. These methodological improvements will most serve the public's health.

Acknowledgments:

REFERENCES

1. US Department of Health E, and Welfare,. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. In: National Institute of Health Bethesda; 1979.

2. Gilmore-Bykovskyi A, Jackson JD, Wilkins CH. The Urgency of Justice in Research: Beyond COVID-19. Trends in molecular medicine. 2021;27(2):97–100. [PubMed: 33277159]

3. Schwartz AL, Alsan M, Morris AA, Halpern SD. Why Diverse Clinical Trial Participation Matters. The New England journal of medicine. 2023;388(14):1252–1254. [PubMed: 37017480]

4. Alsan M, Durvasula M, Gupta H, Schwartzstein J, Williams H. Representation and Extrapolation: Evidence from Clinical Trials. National Bureau of Economic Research, Inc;2022.

5. Gilmore-Bykovskyi A, Croff R, Glover CM, et al. Traversing the aging research and health equity divide: Toward intersectional frameworks of research justice and participation. The Gerontologist. 2022;62(5):711–720. [PubMed: 34324633]

6. Denny A, Streitz M, Stock K, et al. Perspective on the "African American participation in Alzheimer disease research: Effective strategies" workshop, 2018. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2020;16(12):1734–1744.
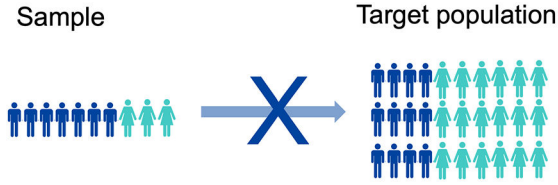
7. World Bank. World Bank Country and Lending Groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Accessed February 17, 2024.

8. Chiao JY, Cheon BK. The weirdest brains in the world. The Behavioral and brain sciences. 2010;33(2–3):88–90. [PubMed: 20546651]

9. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? The Behavioral and brain sciences. 2010;33(2–3):61–83; discussion 83–135. [PubMed: 20550733]

10. Chiao JY. Cultural neuroscience: a once and future discipline. Prog Brain Res. 2009;178:287–304. [PubMed: 19874977]

11. Mooldijk SS, Licher S, Wolters FJ. Characterizing Demographic, Racial, and Geographic Diversity in Dementia Research: A Systematic Review. JAMA neurology. 2021;78(10):1255–1261. [PubMed: 34491284]

12. Mielke MM, Aggarwal NT, Vila-Castelar C, et al. Consideration of sex and gender in Alzheimer's disease and related disorders from a global perspective. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2022;18(12):2707–2724.

13. Fiest KM, Roberts JI, Maxwell CJ, et al. The Prevalence and Incidence of Dementia Due to Alzheimer's Disease: a Systematic Review and Meta-Analysis. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques. 2016;43 Suppl 1:S51–82. [PubMed: 27307128]

14. Lim AC, Barnes LL, Weissberger GH, et al. Quantification of race/ethnicity representation in Alzheimer's disease neuroimaging research in the USA: a systematic review. Commun Med (Lond). 2023;3(1):101. [PubMed: 37491471]

15. Franzen S, Smith JE, van den Berg E, et al. Diversity in Alzheimer's disease drug trials: The importance of eligibility criteria. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2022;18(4):810–823.

16. Dotson VM, Duarte A. The importance of diversity in cognitive neuroscience. Annals of the New York Academy of Sciences. 2020;1464(1):181–191. [PubMed: 31663150]

17. Schumacher-Schuh AF, Bieger A, Okunoye O, et al. Underrepresented Populations in Parkinson's Genetics Research: Current Landscape and Future Directions. Movement disorders : official journal of the Movement Disorder Society. 2022;37(8):1593–1604. [PubMed: 35867623]

18. Cirillo A, Diniz E, Gadelha A, et al. Population neuroscience: challenges and opportunities for psychiatric research in low- and middle-income countries. Braz J Psychiatry. 2020;42(4):442–448. [PubMed: 32267341]

19. Gianattasio KZ, Bennett EE, Wei J, et al. Generalizability of findings from a clinical sample to a community-based sample: A comparison of ADNI and ARIC. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2021.

20. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-representativeness. Lancet (London, England). 2019;393(10178):1297.

21. Doré V, Krishnadas N, Bourgeat P, et al. Relationship between amyloid and tau levels and its impact on tau spreading. European journal of nuclear medicine and molecular imaging. 2021;48(7):2225–2232. [PubMed: 33495928]

22. Jack CR Jr., Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet neurology. 2013;12(2):207–216. [PubMed: 23332364]

23. Ossenkoppele R, Smith R, Ohlsson T, et al. Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. Neurology. 2019;92(6):e601–e612. [PubMed: 30626656]

24. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet. 2020;396(10248):413–446.

25. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. Lancet (London, England). 2017;390(10113):2673–2734. [PubMed: 28735855]

26. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. American journal of epidemiology. 2010;172(1):107–115. [PubMed: 20547574]

27. Lima-Costa MF, de Andrade FB, de Souza PRB Jr., et al. The Brazilian Longitudinal Study of Aging (ELSI-Brazil): Objectives and Design. American journal of epidemiology. 2018;187(7):1345–1353. [PubMed: 29394304]

28. Borelli WV, Leotti VB, Strelow MZ, Chaves MLF, Castilhos RM. Preventable risk factors of dementia: Population attributable fractions in a Brazilian population-based study. Lancet Reg Health Am. 2022;11:100256. [PubMed: 36778926]

29. Mukadam N, Sommerlad A, Huntley J, Livingston G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. The Lancet Global Health. 2019;7(5):e596–e603. [PubMed: 31000129]

30. Lee M, Whitsel E, Avery C, et al. Variation in Population Attributable Fraction of Dementia Associated With Potentially Modifiable Risk Factors by Race and Ethnicity in the US. JAMA Netw Open. 2022;5(7):e2219672. [PubMed: 35793088]

31. Jones RN. Differential item functioning and its relevance to epidemiology. Curr Epidemiol Rep. 2019;6:174–183. [PubMed: 31840016]

32. Kato T Measurement Invariance in the Center for Epidemiologic Studies-Depression (CES-D) Scale among English-Speaking Whites and Asians. International journal of environmental research and public health. 2021;18(10).

33. Assari S, Moazen-Zadeh E. Confirmatory Factor Analysis of the 12-Item Center for Epidemiologic Studies Depression Scale among Blacks and Whites. Frontiers in psychiatry. 2016;7:178. [PubMed: 27872599]

34. Macintosh RC, Strickland OJ. Differential item responses on CES-D inventory: A comparison of elderly Hispanics and non-Hispanic Whites in the United States and item usage by elderly Hispanics across time. Aging & mental health. 2010;14(5):556–564. [PubMed: 20480413]

35. Van Lieshout RJ, Cleverley K, Jenkins JM, Georgiades K. Assessing the measurement invariance of the Center for Epidemiologic Studies Depression Scale across immigrant and non-immigrant women in the postpartum period. Arch Womens Ment Health. 2011;14(5):413–423. [PubMed: 21932024]

36. Bergenfeld I, Kaslow NJ, Yount KM, Cheong YF, Johnson ER, Clark CJ. Measurement invariance of the Center for Epidemiologic Studies Scale-Depression within and across six diverse intervention trials. Psychol Assess. 2023;35(10):805–820. [PubMed: 37616094]

37. Gomez R, McLaren S. The Center for Epidemiologic Studies Depression Scale: Invariance across heterosexual men, heterosexual women, gay men, and lesbians. Psychol Assess. 2017;29(4):361–371. [PubMed: 27362464]

38. Belcher HL, Uglik-Marucha N, Vitoratou S, Ford RM, Morein-Zamir S. Gender bias in autism screening: measurement invariance of different model frameworks of the Autism Spectrum Quotient. BJPsych Open. 2023;9(5):e173. [PubMed: 37781848]

39. Chee ZJ, Scheeren AM, De Vries M. The factor structure and measurement invariance of the Autism Spectrum Quotient-28: A cross-cultural comparison between Malaysia and the Netherlands. Autism. 2023:13623613221147395.

40. McClure K, Bell KA, Jacobucci R, Ammerman BA. Measurement invariance and response consistency of single-item assessments for suicidal thoughts and behaviors. Psychol Assess. 2023;35(10):830–841. [PubMed: 37668583]

41. Demidenko MI, Mumford JA, Ram N, Poldrack RA. A multi-sample evaluation of the measurement structure and function of the modified monetary incentive delay task in adolescents. Dev Cogn Neurosci. 2024;65:101337. [PubMed: 38160517]

42. Dilworth-Anderson P Introduction to the science of recruitment and retention among ethnically diverse populations. The Gerontologist. 2011;51 Suppl 1(Suppl 1):S1–4. [PubMed: 21565811]

43. Gilmore-Bykovskyi AL, Jin Y, Gleason C, et al. Recruitment and retention of underrepresented populations in Alzheimer's disease research: A systematic review. Alzheimer's & dementia (New York, N Y). 2019;5:751–770.

44. Gabel M, Bollinger RM, Knox M, et al. Perceptions of Research Burden and Retention Among Participants in ADRC Cohorts. Alzheimer disease and associated disorders. 2022;36(4):281–287. [PubMed: 35796752]

45. Alsan M, Campbell RA, Leister L, Ojo A. Investigator Racial Diversity and Clinical Trial Participation. National Bureau of Economic Research;2023.

46. Abdelazeem B, Abbas KS, Amin MA, et al. The effectiveness of incentives for research participation: A systematic review and meta-analysis of randomized controlled trials. PloS one. 2022;17(4):e0267534. [PubMed: 35452488]

47. Halpern SD, Chowdhury M, Bayes B, et al. Effectiveness and Ethics of Incentives for Research Participation: 2 Randomized Clinical Trials. JAMA internal medicine. 2021;181(11):1479–1488. [PubMed: 34542553]

48. Wilkins CH, Mapes BM, Jerome RN, Villalta-Gil V, Pulley JM, Harris PA. Understanding what information is valued by research participants, and why. Health Affairs. 2019;38(3):399–407. [PubMed: 30830824]

49. Ganguli M, Lee CW, Hughes T, et al. Who wants a free brain scan? Assessing and correcting for recruitment biases in a population-based sMRI pilot study. Brain imaging and behavior. 2014.

50. Babulal GM, Zhu Y, Trani JF. Racial and ethnic differences in neuropsychiatric symptoms and progression to incident cognitive impairment among community-dwelling participants. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2023;19(8):3635–3643.

51. Griswold ME, Talluri R, Zhu X, et al. Reflection on modern methods: shared-parameter models for longitudinal studies with missing data. International journal of epidemiology. 2021;50(4):1384–1393. [PubMed: 34113988]

52. Hayes-Larson E, Mobley TM, Mungas D, et al. Accounting for lack of representation in dementia research: Generalizing KHANDLE study findings on the prevalence of cognitive impairment to the California older population. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2022;18(11):2209–2217.

53. Rosano C A training program for researchers in population neuroimaging: Early experiences. Front Neuroimaging. 2022;1:896350. [PubMed: 37555144]

54. Falk EB, Hyde LW, Mitchell C, et al. What is a representative brain? Neuroscience meets population science. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(44):17615–17622. [PubMed: 24151336]

55. Butler J 3rd, Fryer CS, Garza MA, Quinn SC, Thomas SB. Commentary: Critical Race Theory Training to Eliminate Racial and Ethnic Health Disparities: The Public Health Critical Race Praxis Institute. Ethn Dis. 2018;28(Suppl 1):279–284. [PubMed: 30116099]

56. National Institute on Aging. NIA's commitment to inclusion https://www.nia.nih.gov/research/blog/2023/11/nias-commitment-inclusion. Published 2023. Accessed December 10, 2023.

57. Ganguli M, Albanese E, Seshadri S, et al. Population Neuroscience: Dementia Epidemiology Serving Precision Medicine and Population Health. Alzheimer disease and associated disorders. 2018;32(1):1–9. [PubMed: 29319603]

58. Jorgensen DR, Shaaban CE, Wiley CA, Gianaros PJ, Mettenburg J, Rosano C. A population neuroscience approach to the study of cerebral small vessel disease in midlife and late life: an invited review. American journal of physiology Heart and circulatory physiology. 2018;314(6):H1117–H1136. [PubMed: 29393657]

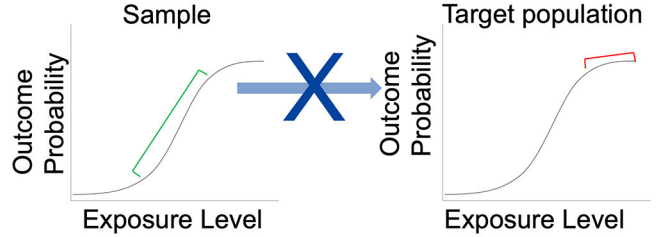59. Paus T Population neuroscience: why and how. Human brain mapping. 2010;31(6):891–903. [PubMed: 20496380]

Problems arising from lack of diversity

Differential prevalence of effect-measure modifiers

Variation in levels of exposure

Non-linearities and range effects

Sample → Target population

Sample (Outcome Probability vs Exposure Level) → Target population (Outcome Probability vs Exposure Level)

Differential item functioning / psychometric properties

Population attributable fraction

Sample → Target population

Fig 1. Examples of problems arising from lack of diversity.
Top left: The exposure-outcome relationship varies by sex assigned at birth, and the sample has a greater proportion of male participants than the target population. Bottom left: An activities of daily living measure works differently in men vs. women. Top right: The exposure-outcome association is strong in the sample, but not in the target population due to differing ranges of exposure values in the sample vs. target population. Bottom right: The sample has a greater proportion of participants with high education (aqua) than the target population.
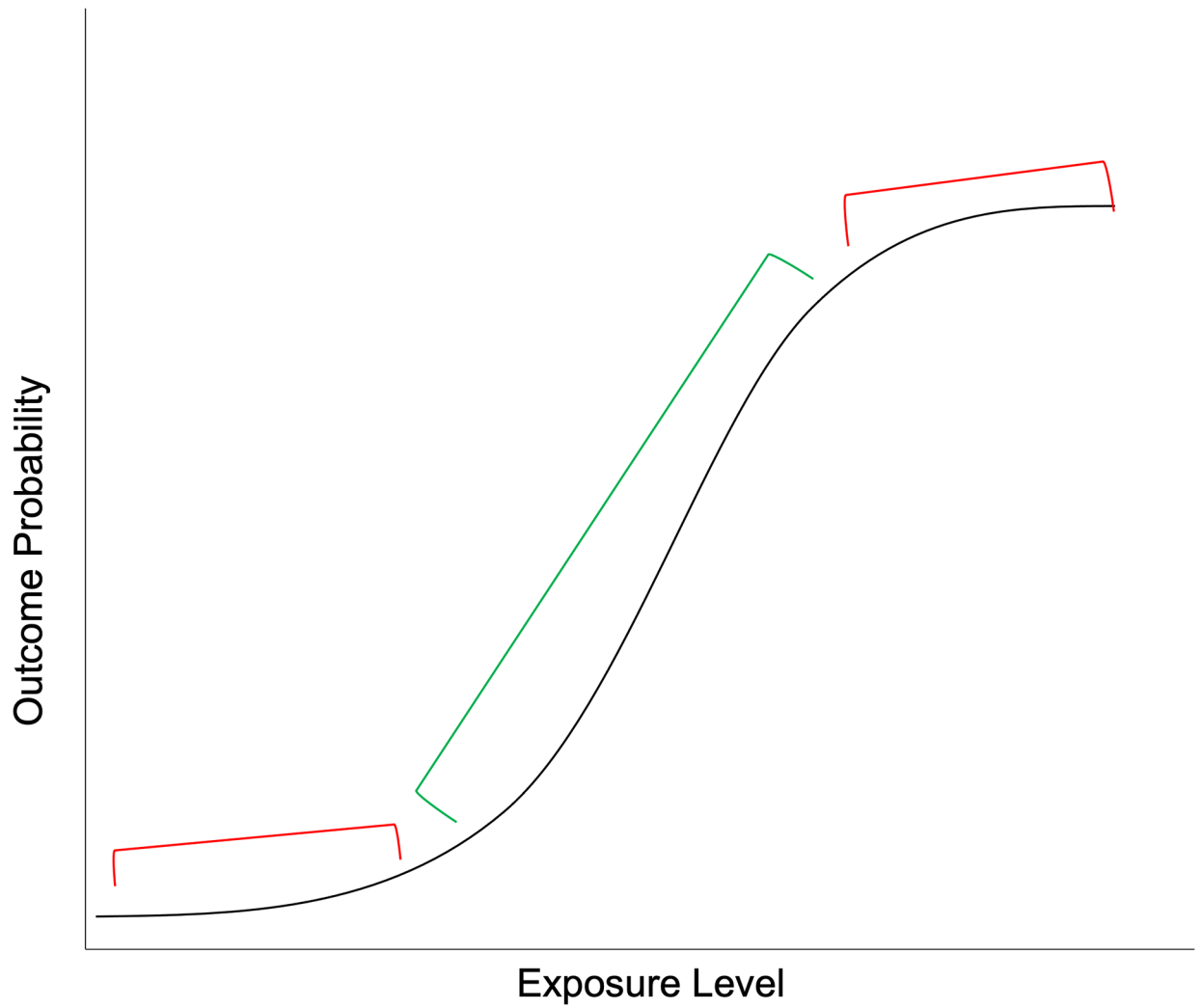
Fig 2. Non-linearities and range effects.
Samples drawn from populations in the red areas (top and bottom of the sigmoid) would have no association between exposure and outcome whereas samples drawn from populations in the green area (middle of the sigmoid) would have a strong positive correlation between exposure and outcome.

Table 1.

Population attributable fractions (PAFs) of modifiable dementia risk factors across various study populations.

Study	Measure	↓ education	↓ hearing	tbi	htn	↑ alcohol	obesity	smoking	depressin	social isolation	physical inactivity	diabetes	air pollution	Total Weighted PAF
RR's for all studies are from: Livingston, et al., 2017 & 2020	RR	1.6 (1.3–2.0)	1.9 (1.4–2.7)	1.8 (1.5–2.2)	1.6 (1.2–2.2)	1.2 (1.1–1.3)	1.6 (1.3–1.9)	1.6 (1.2–2.2)	1.9 (1.6–2.3)	1.6 (1.3–1.9)	1.4 (1.2–1.7)	1.5 (1.3–1.8)	1.1 (1.1–1.1)	
Livingston, et al. 2020. Lancet Commission Report[24] 12 exposures	P_e	40.0%	31.7%	12.1%	8.9%	11.8%	3.4%	27.4%	13.2%	11.0%	17.7%	6.4%	75.0%	
	PAF	7.1%	8.2%	3.4%	1.9%	0.8%	0.7%	5.2%	3.9%	3.5%	1.6%	1.1%	2.3%	39.7%
Livingston, et al., 2017. Lancet Commission Report[25] 9 exposures	P_e	40.0%	31.7%		8.9%		3.4%	27.4%	13.2%	11.0%	17.7%	6.4%		
	PAF	7.5%	9.1%		2.0%		0.8%	5.5%	4.0%	2.3%	2.6%	1.2%		35.0%
Borelli, et al, 2022.[28] Brazil 10 exposures	P_e	51.1%	29.4%		52.0%	4.3%	29.2%	17.1%	18.1%	5.1%	47.2%	15.7%		
	PAF	9.5%	14.2%		10.4%	0.2%	6.3%	2.9%	6.9%	1.5%	11.3%	3.6%		50.5%
Mukadam, et al., 2019.[29] 6 Latin American LMICs 9 exposures	P_e	68.8%	28.8%		55.6%		44.8%	30.0%	23.9%	0.5%	34.2%	18.5%		
	PAF	10.9%	7.7%		9.3%		7.9%	5.7%	6.6%	0.1%	4.5%	3.2%		55.8%
Mukadam, et al., 2019.[29] India 9 exposures	P_e	92.2%	22.3%		19.3%		13.7%	39.9%	5.2%	10.4%	15.3%	9.3%		
	PAF	13.6%	6.4%		4.0%		2.9%	6.4%	1.7%	2.3%	2.2%	1.7%		41.2%
Mukadam, et al., 2019.[29] China 9 exposures	P_e	75.9%	14.3%		38.1%		32.0%	23.0%	1.5%	3.4%	50.7%	9.4%		
	PAF	10.8%	3.9%		6.4%		5.6%	4.2%	0.5%	0.7%	5.8%	1.6%		39.5%
Lee, et al., 2022.[30] US PAFs overall	P_e	10.7%	10.8%	17.1%	42.2%	3.6%	44.0%	8.5%	7.4%	11.9%	62.8%	28.6%	22.8%	
	PAF	2.0%	3.0%	4.1%	6.8%	0.2%	7.1%	1.6%	2.1%	2.3%	6.8%	4.2%	0.8%	41.0%
Lee, et al., 2022.[30] US PAFs, Hispanic 12 exposures	P_e	27.1%	13.1%	10.3%	38.5%	2.0%	48.3%	6.9%	10.7%	24.0%	68.6%	41.0%	44.4%	
	PAF*	4.6%	3.5%	2.5%	6.2%	0.1%	7.4%	1.3%	2.9%	4.1%	7.1%	5.6%	1.4%	46.7%
Lee, et al., 2022.[30] US PAFs, Non-Hispanic Asian 12 exposures	P_e	6.4%	6.9%	6.0%	38.5%	0.7%	14.6%	4.9%	4.3%	8.0%	56.6%	44.1%	55.2%	
	PAF*	1.4%	2.2%	1.7%	7.1%	0.1%	3.1%	1.1%	1.4%	1.7%	7.0%	6.9%	2.0%	35.6%

Study	Measure	education	hearing	tbi	htn	alcohol	obesity	smoking	depressin	social isolation	physical inactivity	diabetes	air pollution	Total Weighted PAF
Lee, et al., 2022.[30] US PAFs, Non-Hispanic Black 12 exposures	P_e	10.6%	6.5%	9.2%	61.0%	2.7%	54.3%	11.7%	6.6%	12.1%	73.2%	37.2%	41.3%	
	PAF*	2.1%	1.9%	2.4%	9.3%	0.2%	8.5%	2.3%	1.9%	2.3%	7.8%	5.4%	1.4%	45.6%
Lee, et al., 2022.[30] US PAFs, Non-Hispanic White 12 exposures	P_e	5.5%	10.6%	20.1%	39.8%	4.2%	43.5%	8.4%	7.2%	10.8%	61.3%	25.4%	17.2%	
	PAF*	1.1%	3.0%	4.7%	6.5%	0.3%	7.0%	1.6%	2.1%	2.1%	6.7%	3.8%	0.6%	39.4%

Note: tbi=traumatic brain injury; htn=hypertension. All PAFS are weighted.

*
Weighted PAFs obtained via direct correspondence with the authors.