William T. Grant
FOUNDATION

Opening the Black Box: Overcoming Obstacles to Studying Causal Mechanisms in Research on Reducing Inequality

Margarita Alegría
Massachussets General Hospital
Harvard Medical School

Isabel Shaheen O'Malley
Massachussets General Hospital

Opening the Black Box: Overcoming Obstacles to Studying Causal Mechanisms in Research on Reducing Inequality

Margarita Alegría
Massachussets General Hospital
Harvard Medical School

Isabel Shaheen O'Malley
Massachussets General Hospital

IDENTIFYING CAUSAL MECHANISMS is an essential aspect of disparities research. Rigorous investigations of mechanisms can open the black box and clarify linkages in the causal chain, indicating how effects occur. Studies of mechanisms can also illuminate novel ways to reduce inequality when intervention at the root cause of unequal outcomes is not feasible.

Still, research on mechanisms can present challenges for researchers and funders alike. Designing and carrying out studies of mechanisms can be difficult in part because of the conceptual issues involved, the complex role of the environment in shaping outcomes, and the confounding factors present in complex systems. Studies of mechanisms can also be expensive, requiring large samples for sufficient power, especially when systems are the unit of analysis. Finding the right balance of scope, innovation, and feasibility, then, is a central concern for all involved.

In this essay, we summarize the necessary trade-offs and inherent tensions of research on causal mechanisms, and we discuss challenges and opportunities for future studies. By explicating the complexities of disparities mechanisms research and exploring ways to address these challenges, we hope to help researchers propose work that is both rigorous and novel. We also aim to help research funders understand how best to support this work in ways that align with existing interests and resources.

Why Mechanisms?

Importantly, the mechanisms that *contribute* to a specific disparity are not necessarily the same mechanisms that are capable of, or best suited for, addressing it.

For instance, extensive empirical evidence demonstrates that structural racism—i.e., "laws, rules, or official policies in a society that result in and support a continued unfair advantage to some people and unfair or harmful treatment of others based on race" (Cambridge Dictionary, n.d.)—is a fundamental cause of observed differences in health outcomes between racial and ethnic groups (Du Bois, 1899; Du Bois, 2003 [1906]; White, Thornton, & Greene, 2021; D. R. Williams, Lawrence, & Davis, 2019). The understanding of structural racism as a root cause of health disparities could suggest that strategies to reduce disparities must also be structural, at the level of the institutions, policies, and laws that create and perpetuate inequality. But this need not be the case. For example, the water supplied to an economically disadvantaged community may be of poor quality because of years of discriminatory housing policies, but more effective water treatment could constitute an effective mechanism for ameliorating negative outcomes in the community today, even as high-level policy change remains ahead of us.

In fact, most models of disparities hypothesize multiple mechanisms leading to disparate outcomes. For example, Bailey and colleagues (Bailey et al., 2017) cite literature demonstrating numerous pathways between structural racism and health, including economic injustice and social deprivation, environmental and occupational health inequities, psychosocial trauma, targeting marketing of health-harming substances, inadequate health care, state-sanctioned violence and alienation from property and traditional lands, political exclusion, maladaptive coping behaviors, and stereotype threats. Bailey and colleagues also describe three inter-related domains through which structural racism in the U.S. has harmed population and individual health, specifically: redlining and racialized residential segregation, police violence and mass incarceration, and unequal medical care (Bailey, Feldman, & Bassett, 2021).

Accounting for multiple mechanisms, then, often requires complex studies that address various levels of influence (e. g. macro, meso, and micro). Sometimes the research may introduce substitute mechanisms; for example, removing police from schools and ending out-of-school suspensions can reduce disparities in discipline rates, but we cannot dismantle the school to prison pipeline without also investing in counselors, training for administrators and school staff, and other evidence-based supports for students. In some cases, we may need support from higher levels of decision-making; for instance, cities and states are primarily responsible for police departments, but we have seen that state initiatives have been necessary to spur investments in alternatives to policing. And in some cases, the disparity has ripple effects across other areas of life and over generations (e.g., genocide and displacement of American Indian tribes has yielded increased rates of inter-generational trauma), so correcting past injustices may require both restorative justice measures related to the causes (e.g., desires and needs of specific tribes regarding occupation of land) and effects (e.g., cultural practices to support positive mental health and wellbeing).

Obstacles to Identifying Mechanisms in Disparities Research

What obstacles prevent researchers from identifying pathways underlying inequality and mechanisms of successful interventions to reduce it? For one, in scientific evaluation of processes in real-world settings, as opposed to controlled clinical experiments, we face the challenge of confidently establishing causal relationships.

Establishing Causality

While randomization is ideal for establishing causality, much of disparities research is dependent upon observational data. This is due to several limitations of randomization. First, many processes and variables are not possible to randomize,

either logistically (e.g., randomizing race) or ethically (e.g., exposure to racism) (Jeffries et al., 2019). Second, randomization does not yield well to generalizability, because it does not account for self-selection into a study (i.e., bias introduced when a methodology, respondent sample, or analysis is biased toward a specific subset of a target population) or for selecting into other interventions which could work better than the randomly assigned one (e.g., bias of youth who might enroll into pharmacologic treatments when some of the population of youth participating might be different if offered psychotherapeutic treatments) (Deaton & Cartwright, 2018).

Another challenge to establishing causality in research on reducing inequality, as in other types of research, is the high correlation between causal factors. Factors may exist at multiple ecological levels (e.g., economic marginalization experienced at the individual, family, neighborhood, city, state, and national levels) that need to be conceptualized and measured. Another concern is that traditional statistical modeling has not adequately captured complex processes such as feedback loops and variables with dynamic properties (Diez-Roux, 2018).

An additional obstacle to establishing cause-effect relationships with potential mechanisms is that variables can be simultaneously confounders and mediators. Take economic disadvantage and neighborhood health effects, for example. Where one lives can affect income (e.g., access to well-paying jobs). Income could also affect residential location (e.g., what neighborhood one can afford to live in). And there could be independent or compounding effects on health from income and neighborhood; for instance, low income from one job often requires working multiple jobs and leaves no time for spending time with their children, and lack of parks nearby also limits options for destressing with your children. So, the compounding effect of disparities effects might require addressing these divergent pathways, not only tackling one path. But tackling diverse paths might require more costs, diverse expertise, and maybe additional data sources.

Data analysis presents its own set of issues. Methods for analyzing disparities within observational data must be appropriate for the sample and all subgroups. When a minority group sample is small, one may have little choice but to use regression analyses that fit the regression to the majority group. However, an issue with always comparing to the majority group is that it reproduces inaccurate narratives that posit the historically majority group as the default or standard, rather than setting the benchmark based on whatever group has the best health outcome. Another method of disparities analysis for observational data is to compare the relative difference between two groups' change on a given outcome when one group receives an intervention, and the other group does not. This method, called "difference in difference" analysis, has the benefit of not requiring the two groups to be identical to begin with. But to show causal inference this analysis relies on assumptions that are rarely possible in real-world scenarios, such as the condition that there are no other changes that occur in one group but not the other during the time of the intervention, which might have influenced the outcome of interest.

Addressing Multiple Levels of Causality

Causes of disparities can come from multiple levels (e.g., individual level, community level, institution level), and not addressing this appropriately within the analysis design can lead to errors in estimates and inferences unless structural models are specified correctly or randomization is effective in matching on outside factors influencing outcomes. What's more, individual-level cognitive processes can also influence macro-level phenomena and contribute to production and reproduction of inequality.

For example, "cultural processes" (Lamont et al., 2014) involve shared conventions and inter-subjective meaning making. Lamont discusses the importance of shared groupings and classification systems by which individuals perceive and make sense of the world around them. One aspect that is not sufficiently addressed in disparities studies is how the subjective measures of certain constructs seem more strongly related with health outcomes than the objective measures of those constructs. For example, subjective socioeconomic status has a stronger association with mental health conditions than more objective indicators like annual household income or education level (Chen, 2020; Sakurai, 2010). Subjective and objective social support are also unique predictors of mental health outcomes and have different areas for intervention to improve mental health (Ma et al., 2020) but subjective measures appear as better predictors than objective ones at the individual level. Lamont recommends social scientists pay more attention to cultural capital and symbolic domination (e.g., language)—social psychological schemas that we apply to social categories whereby we perceive and evaluate different groups—and establish shared frameworks of reality (Lamont et al., 2014).

Overcoming the Obstacles for Identifying Mechanisms in Disparities Research

The urgency of identifying causal mechanisms through which inequality can be reduced demands new research. But given what we know about the challenges involved, how might we approach this work?

Reducing Threats to Causality

There is no single research design that is best suited to identifying mechanisms of disparities but rather trade-offs. A balance of randomized-controlled trials and other natural experimental research designs might be options to consider, but other approaches such as instrumental variable analyses, propensity score weighting, and regression-discontinuity can also serve the purpose if well-conceptualized.

Political scientists Imai et al. (2011) propose an approach for establishing causality in experimental and observational studies. First, they suggest a minimum standard set

of assumptions required under standard designs of experimental and observational studies. They present a general algorithm for estimating causal mediation effects. Second, they offer a method for assessing the sensitivity of conclusions to potential violations of a key assumption. Third, they identify alternative research designs for identifying causal mechanisms under weaker assumptions. When Imai et al (2011) use this three-step method to re-analyze data from a prior study (Brader, Valentino, and Suhay, 2008) their method produces more appropriate estimates of the average causal mediation effects because they use nonlinear models. Their method also clarifies what the necessary assumptions are to identify the mediation effects and provides a sensitivity analysis. There are also new approaches such as for imitating or approaching a target trial to inform decision-making when evidence from a randomized trial is not available (Hernán & Robins, 2016).

Natural experiments, or quasi-experimental designs used in disparities research, involve an intervention or exposure that is not randomized but rather determined by another force which is assumed to be unrelated to the potential confounding factors. Examples include differences in availability of health services by geography, or the phase-in of a policy like Medicaid expansion. Unfortunately, the geographic variation and local context makes it hard to have true control groups (Layton, 2019). Two types of structural modeling can reduce bias in natural experiments. Marginal structural models (Robins et al., 2000) reduce bias by creating pseudo-populations, where the exposure's effect is not confounded with the covariates that are used for adjustment. Fixed-effects modeling involves an indicator variable for an individual or group (depending on which is the unit of analysis) which represents all the time-invariant characteristics (observed and unobserved) (Allison, 2009). This allows us to estimate the effects of an individual-level intervention that varies with time conditioned on individual-level characteristics that do not vary with times. This method is only possible when at least some individual-level or group-level confounders can be held constant (e.g., in longitudinal studies), and when there is substantial observable within-person variation. With fixed effects models there is still the possibility of unmeasured covariates that do vary with time, and they are not good for studying causal processes with long-term lags.

No matter the design, social sciences best practices still apply, particularly those that support unobtrusive measurement and naturalistic interventions, that pay attention to realistic outcomes and consequential behaviors, and that have application to diverse samples and settings (Baldassarri & Abascal, 2017). The QuaSIC approach (Harris et al., 2016), for example, is one way of conducting a longitudinal evaluation of a multi-level intervention through semi-structured interviews, focus groups, field notes, and questionnaires. It is proposed by Jeffries and colleagues (2019) as an alternative because it reduces bias arising from time-varying confounders that are usual in health disparities research. Each method provides unique data from different sources and levels in a way that enables researchers to test mechanisms by analyzing interactions between the levels. Still, there are concerns about whether we

need newer approaches and diverse ways of knowing (Buchanan et al., 2021; Tuck et al., 2019), like storytelling, writing, or photovoice.

In studies that test hypotheses, designing research with sufficient statistical power is crucial. Power and sample size analysis help define and justify the research costs. Studies that are underpowered prevent researchers from drawing meaningful conclusions, and thus add to inefficiency and waste of participants' efforts and risks (Hey, 2017), and money used for research activities (Vadillo, Konstantinidis, & Shanks, 2016). A realistic research proposal may need to promise less, such as a study of a pilot project that will not have the power to significantly find the relationship to detect the hypothesized effects but might provide a signal or otherwise improve the types of relations that might be considered. Kraemer et. al (2006) note that one should be careful in making determinations of mechanisms when pilot testing in potentially underpowered samples to not discredit a mechanism that might actually work. Gluek and Mullens have worked on solutions to this challenge, identifying what one can do with small budgets to test disparities if the study is suspected of being under-powered (samplesizeshop.org). For example, Gluek and Mullens created a free and easy to use online software that can calculate the power and sample size analysis needed for specific study designs (Guo et al., 2013; Kreidler et al., 2013).

Finally, large samples are typically necessary for identifying mediators and moderators in research of how policies, programs, or practices reduce inequality. In disparities research one needs to compare (i.e., "stratify") by racial/ethnic group and, ideally, by sub-ethnic group (B. A. Williams, Brooks, & Shmargad, 2018). Aggregating sub-ethnic groups (e.g., Asian Americans" might make the aggregate grouping mask differences across these diverse subgroups (e.g., Vietnamese, Chinese, Filipinos). But issues of differential language and culture and need for varied expertise typically jeopardize the feasibility of such challenging studies.

Working Within the Limits of Feasibility

It is very difficult to calculate the resources (e.g., funding source and amount, time, personnel, types of studies) required to establish, test, refine theoretical models. For example, in 1996 García Coll and colleagues introduced the "integrative model for the study of developmental competencies in minority children," which outlined domains and pathways (e.g., racial discrimination) that result in worse outcomes for racial/ethnic minority youth development (Garcia-Coll, 1996). The Boricua Youth Study is one of many longitudinal studies that expanded on this model (Alegría et al., 2019). Future research directions expanding on the integrative model continue to be proposed (Seaton, Gee, Neblett, & Spanierman, 2018), such as applying it to white youth populations and expanding segregation to include mass incarceration and deportations.

Often, funding agencies have maximum grant levels that make it impossible to test the full causal chain from intervention to mechanisms of disparities reduction to improved outcomes. This is especially true for private funders, including the William T. Grant Foundation, which generally funds studies on reducing inequality up to $600,000. In such cases, it may be possible to assess a portion of the causal chain and make the case that information on a proximate outcome will inform efforts to reduce inequality. Alternatively, conceptually rich qualitative research may be highly informative about mechanisms and fit within the available budget, even though it does not aim to test all the potential hypotheses of a conceptual model. Either of these strategies may be preferred to a "black box" study that gathers no information about mechanisms, unless knowledge of mechanisms is already well established.

Exploiting Qualitative and Mixed Methods

Incorporating qualitative research is critical for identifying causal mechanisms in research on reducing inequality. Qualitative methods are useful at all stages of disparities research, including identifying potential causal factors and processes, generating, and refining conceptual models and hypotheses, and explaining relationships among factors observed in quantitative studies (Louie, 2015). Qualitative research also provides a holistic and realistic view of the people you're studying: It centers the perspectives of participants, allowing them to use their own language (identifying themselves, terminology), set the agenda, and determine priorities. Importantly, qualitative methods can introduce ideas and explanations that researchers may not think of. Of course, researchers' interpretations of participants' perspectives need to be validated, through methods like member checking and inter-coder agreement checks, and by presenting the results to the participants and community forums (Creswell & Báez, 2020).

Mixed-methods research, which combines and integrates quantitative and qualitative methods, can lead to additional insights about mechanisms of disparities that are not possible from qualitative or quantitative approaches alone. There are at least seven distinct types of mixed-methods designs, all of which are appropriate for uncovering potential causal factors and evaluating disparities interventions in complex environments (Creswell & Clark, 2017). Threats to validity and strategies to minimize validity threats vary by design type (see Creswell & Clark, 2017, pp. 251-253). In addition, using mixed-methods requires that researchers have expertise and follow rigorous methods when adapting a mixed-methods design for a particular research question. The American Psychological Association has recently set standards for publishing qualitative and mixed methods projects (Levitt et al., 2018), which are useful for investigators to review when preparing a grant application to make sure they budget the time and resources for the necessary procedures.

Pursuing Innovative Questions

Innovation is important for disparities research because the status quo of research has not solved inequities fast enough, with bias and harm in past literature driven by majority culture (Buchanan, Perez, Prinstein, & Thurston, 2021). Much of the research on the mechanisms of health inequities have identified and focused on the same problems: poverty in disadvantaged neighborhoods, poor healthcare access and quality, and insufficient opportunities for housing, employment, and education. Importantly, researchers of color have different lived experience and may look for different mechanisms within an intervention. Black, Indigenous, and Latinx investigators have also demonstrated new research methods for analyzing problems and proposing potential ways of addressing disparities. For instance, Eve Tuck has written extensively on settler colonialism; instead of focusing only on the relationship between settler colonies and Indigenous societies, Tuck directs attention to each group's relationship with place and land (Tuck, 2011), which has not been a common approach among White scholars.

The extent to which a study identifies novel or innovative approaches to reducing inequality is an increasingly important consideration for funders. For example, the National Institutes of Health rewards innovation through its Common Fund, which supports novel ideas or approaches and prioritizes the scope of a study's potential impact without requiring preliminary data or experimental details (e.g., the "Transformative Research to Address Health Disparities and Advance Health Equity" initiative). In addition, the William T. Grant Foundation explicitly calls out innovation in its review criteria for research grants, noting "Where appropriate, we value projects that … demonstrate significant creativity and the potential to advance the field" (William T. Grant Foundation, n.d.).

What Types of Studies Might Researchers and Funders Pursue?

To better help researchers identify the opportunities and budgets that match their research, funders could create a three-tier system of research spanning exploratory and developmental studies, testing of disparities mechanisms, and implementation research.

Tier 1: Exploratory or Developmental Studies

In exploratory or developmental studies, which could constitute a first tier of research, researchers may develop theoretical frameworks or conceptual models of mechanisms (untested or newly proposed), or identify steps in a theory of change, including interventions, polices, or programs that are innovative and in early-development stages. This work may also include pilot testing or new ways of thinking about a given problem.

Mesmin Destin's work on social mobility and identity exemplifies strong research at this stage. Focusing on college students of color, Destin first developed a framework of the novel concept "status-based identity," defined as an individual's subjective experience of their SES through the lens of personal identity (i.e., narrative, social, and future identity) (Destin, 2017), and collected a sample to better capture the underlying phenomena of interest. Destin also developed the *Status-Based Identity Uncertainty Scale* (SBIU), established measurement validity and reliability, then pilot tested the study with undergraduates to determine specific relationships within the framework. Early findings indicated that lower family income was linked with greater status-based identity uncertainty, which uniquely predicts lower self-esteem and lower satisfaction with life. Destin then expanded this study (Destin, 2019) by collecting survey data from a large number of college students and adding biological data to test the theoretical model with objective measures of stress.

Tier 2: Studies that Test Mechanisms

A second tier may comprise studies that test mechanisms for reducing inequality, either through policies, programs, or practices. This level of study is appropriate when preliminary work has identified some signal of the theorized mechanisms in the predicted directions.

Depending on how large the study is, getting funding from multiple sources, or co-funding, is a great option. One example of a study in this category is Sean Reardon's (2019) study of educational opportunity in early and middle childhood, which analyzed test scores from 45 million students across eleven thousand school districts in the U.S, receiving support from the Institute of Education Sciences, the Spencer Foundation, the William T. Grant Foundation, the Bill and Melinda Gates Foundation, and the Overdeck Family Foundation. Reardon measured educational opportunity through two distinct measures: average test scores in 3rd grade across school districts (reflecting high levels of early childhood education like HeadStart) and growth in average test score rates from grades 3 to 8 (reflecting the average extent of educational opportunities available to children ages 9-14 in a given school district). Reardon found the two measures largely uncorrelated and concluded that early and middle childhood opportunities varied across school districts and seem to be two different dimensions of educational opportunity.

Tier 3: Implementation Research

Finally, a third tier may encompass implementation research, defined as "the scientific study of methods and strategies that facilitate the uptake of evidence-based practice and research into regular use by practitioners and policymakers" (The UW Implementation Science Resource Hub, 2021).

Implementation research can be used to evaluate policies and strategies that are widely practiced but not evidence-based, or strategies that may have evidence but require certain conditions and capacities that are unreasonable for many settings and populations. For example, online courses in high schools had become common in K-12 schools even before the COVID-19 pandemic, in part due to their promise for expanding opportunities for individualized instruction. Yet the evidence is still under-developed, and findings are mixed on student achievement and wide variation across schools and by student sub-groups (Pane et al., 2017). One of the few experiments conducted on online courses found that students randomized to an online course had lower test scores and credit recovery rates compared with similar students randomized to the face-to-face course (Heppen et al., 2017). Heinrich and colleagues (2019) conducted an implementation study involving an analysis of records from a large urban school district's online classes, students' grades, and classroom observations to see how online courses are implemented and to evaluate whether they improve student academic outcomes. The team found that although upperclassmen did raise their grade point averages, participation in online classes overall was associated with lower math and reading scores compared to in-person instruction, and students who were less prepared at the start of the semester performed worse academically and were set back by online courses (Heinrich et al., 2019). The study was able to identify that the implementation of district-recommended practices and supports (e.g., individualized instruction and live-teacher instruction) was constrained by the limited resources, likely contributing to these inequitable outcomes.

Tapping into the potential of implementation research also means greater attention to the collaboration necessary for this work to bear fruit, and a broader recognition of the communities experiencing the disparities we seek to address. For instance, research funders including federal agencies, state and local governments, and philanthropies have historically under-represented many BIPOC populations within implementation studies that tackle disparities (Baumann & Cabassa, 2020). This must be addressed through an expansion of funding opportunities in this area. Furthermore, Baumann and Cabassa also recommend that grants and contracts for implementation research plan for time in the initial stage of the project for researchers and community partners to strengthen their partnerships. There is also a need to design implementation outcomes (e.g., cost, fidelity, sustainability) that have equity built in. Baumann & Cabassa (2020) suggest several research questions to aid with this, such as "what community-, organization-, provider-, and client-level factors contribute to inequities in implementation outcomes between organizations delivering the same evidence-based intervention to different populations?" and "Which implementation strategies produce equitable implementation outcomes between organizations delivering the same EBI to different populations?" (p. 6).

Conclusion

Today more than ever, we see the importance of conducting state-of-the art studies on programs, policies, and practices to reduce inequality in youth outcomes. We also see the need for funders to be both flexible and critical in considering the research questions and designs of prospective grantees and how the proposed work aligns with grantmaking aims and limitations.

For all involved, the complexities involved in disparities mechanisms research demand careful attention, and are less amenable to short term pressures for publication or financial reward (Conrad, 2013). Trade-offs will be necessary. For instance, funders and researchers might need to consider incremental studies that enable us to view the fuller causal chain of mechanisms over time. This approach adds pieces to the puzzle and starts formulating answers to research questions but does not promise definitive answers. Reardon's study, cited earlier, was a massive undertaking, but other studies can test the mechanisms for reducing inequality with smaller samples or with qualitative or mixed-methods studies that evaluate the mitigation of disparities. The challenge, of course, is finding the right alignment of research question, study design, budget, and expertise.

Evaluating the feasibility of data collection, qualitative or quantitative, is difficult for new scholars, particularly those using new sources of data or adopting unfamiliar study designs. Having funders help investigators connect to resources and co-funders, and possibly funding multiple PIs to jointly conduct studies, might be helpful. Funders might also question whether it is better to support fewer studies with more funding than to put the burden of sustaining the research enterprise mostly on the researcher's shoulders, with scarce resources.

Finally, funders want innovation, but at the same time they want results. While no-strings-attached funding, such as that of the MacArthur Foundation, is not an option for all funders, more research funders can incorporate grantmaking strategies that engender autonomy and flexibility (Conrad, 2014). Importantly, real innovation in research means considering who is involved in the conduct, reporting, reviewing, and dissemination of the work. Participatory methods and decolonizing research processes are increasingly important for research involving minoritized communities (Simonds & Christopher, 2013).

Although many of today's disparities have their roots in policies or actions initiated generations ago, there is so much we can do today. And new research on causal mechanisms can point the way forward. Studies of mechanisms that can reduce inequality can shine a light on ways to improve people's lives today, without waiting for higher-level policy reform or large-scale structural change. As we have discussed, this research can introduce new responses to old challenges. It can illuminate how and why larger efforts succeed or fail and identify more effective means of bringing

about improved outcomes for our nation's youth. Conducting and supporting this work will not be without challenges, but by acknowledging the inherent obstacles and working together to overcome them, researchers and funders can generate empirical evidence that increases understanding and guides decisions where and when it matters most.

References

Alvidrez, J., Castille, D., Laude-Sharp, M., Rosario, A., & Tabor, D. (2019). The National Institute on Minority Health and Health Disparities Research Framework. *American Journal of Public Health, 109*(S1), S16_S20. https://doi.org/10.2105/ajph.2018.304883

Bailey, Z. D., Feldman, J. M., & Bassett, M. T. (2021). How structural racism works — Racist policies as a root cause of U.S. racial health inequities. *New England Journal of Medicine, 384*(8), 768-773. https://doi.org/10.1056/NEJMms2025396

Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., & Bassett, M. T. (2017). Structural racism and health inequities in the USA: Evidence and interventions. *The Lancet, 389*(10077), 1453-1463. https://doi.org/https://doi.org/10.1016/S0140-6736(17)30569-X

Baldassarri, D., & Abascal, M. (2017). Field experiments across the social sciences. *Annual Review of Sociology, 43*(1), 41-73. https://doi.org/10.1146/annurev-soc-073014-112445

Baumann, A. A., & Cabassa, L. J. (2020). Reframing implementation science to address inequities in healthcare delivery. *BMC Health Services Research*, *20*(1), 190. https://doi.org/10.1186/s12913-020-4975-3

Boyd, R. W., Lindo, E. G., Weeks, L. D., & McLemore, M. R. (2020). On racism: A new standard for publishing on racial health inequities. *Health Affairs Blog, 10*(10.1377).

Buchanan, N. T., Perez, M., Prinstein, M., & Thurston, I. (2021). Upending racism in psychological science: Strategies to change how our science is conducted, reported, reviewed & disseminated. *American Psychologist*, *76*(7), 1097–1112. https://doi.org/10.1037/amp0000905

Cambridge Dictionary. (n.d.). Structural racism. In *Cambridge Advanced Learner's Dictionary & Thesaurus*. Cambridge University Press. Retrieved January 12, 2021, from https://dictionary.cambridge.org/us/dictionary/english/structural-racism

Chen, R., Kessler, R. C., Sadikova, E., NeMoyer, A., Sampson, N. A., Alvarez, K., ... & Williams, D. R. (2019). Racial and ethnic differences in individual-level and area-based socioeconomic status and 12-month DSM-IV mental disorders. *Journal of Psychiatric Research, 119*, 48-59.

Cleary, P. D., Gross, C. P., Zaslavsky, A. M., & Taplin, S. H. (2012). Multilevel interventions: Study design and analysis issues. *Journal of the National Cancer Institute. Monographs, 2012*(44), 49-55. https://doi.org/10.1093/jncimonographs/lgs010

Conrad, C. (2013). Five myths about the MacArthur "Genius Grants". *The Washington Post*. https://www.macfound.org/media/files/macarthur_fellows_program_article_reprint_-_five_myths_about_the_macarthur_genius_grants.pdf

Conrad, C. (2014). Our society discourages innovation. *The New York Times*. https://www.macfound.org/media/files/macarthur_fellows_program_article_reprint_-_our_society_discourages_innovation.pdf

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2-21. https://doi.org/10.1016/j.socscimed.2017.12.005

Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology, 28*(2), 115-138. https://doi.org/10.1080/02691728.2013.782585

Du Bois, W. (1899). *The Philadelphia negro: A social study*. University of Pennsylvania Press.

DuBois, W. E. B. (2003). The health and physique of the Negro American. 1906. *American Journal of Public Health, 93*(2), 272-276. https://doi.org/10.2105/ajph.93.2.272

Gee, G. C., & Ford, C. L. (2011). Structural racism and health inequities: Old issues, new directions. *Du Bois Rev, 8*(1), 115-132. https://doi.org/10.1017/s1742058x11000130

Gee, G. C., Walsemann, K. M., & Brondolo, E. (2012). A life course perspective on how racism may be related to health inequities. *American Journal of Public Health, 102*(5), 967-974. https://doi.org/10.2105/AJPH.2012.300666

Gelinas, L., Lynch, H. F., Bierer, B. E., & Cohen, I. G. (2017). When clinical trials compete: Prioritising study recruitment. *J Med Ethics, 43*(12), 803-809. https://doi.org/10.1136/medethics-2016-103680

Gorin, S. S., Badr, H., Krebs, P., & Prabhu Das, I. (2012). Multilevel interventions and racial/ethnic health disparities. *Journal of the National Cancer Institute Monographs, 2012*(44), 100-111. https://doi.org/10.1093/jncimonographs/lgs015

Hansen, H., & Metzl, J. M. (2019). *Structural competency in mental health and medicine: A case-based approach to treating the social determinants of health*. Springer.

Harris, F. M., Maxwell, M., O'Connor, R., Coyne, J. C., Arensman, E., Coffey, C., . . . Hegerl, U. (2016). Exploring synergistic interactions and catalysts in complex interventions: Longitudinal, mixed methods case studies of an optimised multi-level suicide prevention intervention in four European countries (Ospi-Europe). *BMC Public Health, 16*(1), 268. https://doi.org/10.1186/s12889-016-2942-z

Heinrich, C. J., Darling-Aduana, J., Good, A., & Cheng, H. (2019). A look inside online educational settings in high school: Promise and pitfalls for improving educational opportunities and outcomes. *American Educational Research Journal, 56*(6), 2147–2188. https://doi.org/10.3102/0002831219838776

Heppen, J. B., Sorensen, N., Allensworth, E., Walters, K., Rickles, J., Taylor, S. S., & Michelman, V. (2017). The struggle to pass algebra: Online vs. face-to-face credit recovery for at-risk urban students. *Journal of Research on Educational Effectiveness, 10*(2), 272–296. https://doi.org/10.1080/19345747.2016.1168500

Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology, 183*(8), 758–764. https://doi.org/10.1093/aje/kwv254

Hey, S. P. (2017). Priority, prediction and the ethical research enterprise. *Journal of Medical Ethics, 43*(12), 812-813. https://doi.org/10.1136/medethics-2016-103999

Jones, C. P. (2000). Levels of racism: A theoretic framework and a gardener's tale. *American Journal of Public Health, 90*(8), 1212-1215. https://doi.org/10.2105/ajph.90.8.1212

Kaufman, M. R., Cornish, F., Zimmerman, R. S., & Johnson, B. T. (2014). Health behavior change models for HIV prevention and AIDS care: Practical recommendations for a multi-level approach. *J Acquir Immune Defic Syndr, 66 Suppl 3*(Suppl 3), S250-258. https://doi.org/10.1097/qai.0000000000000236

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*(5), 484–489. https://doi.org/10.1001/archpsyc.63.5.484

Krieger, N. (2014). Discrimination and health inequities. *Int J Health Serv, 44*(4), 643-710. doi:10.2190/HS.44.4.b

Louie, V. (2016). *Identifying responses to inequality: The potential of qualitative and mixed-methods research*. William T. Grant Foundation. http://wtgrantfoundation.org/library/uploads/2016/06/The-Value-of-Qualitative-and-Mixed-Methods-Research.pdf

Ma, R., Mann, F., Wang, J., Lloyd-Evans, B., Terhune, J., Al-Shihabi, A., & Johnson, S. (2020). The effectiveness of interventions for reducing subjective and objective social isolation among people with mental health problems: A systematic review. *Social Psychiatry and Psychiatric Epidemiology, 55*(7), 839-876.

Metzl, J. M., & Hansen, H. (2014). Structural competency: Theorizing a new medical engagement with stigma and inequality. *Social Science & Medicine, 103*, 126-133. https://doi.org/10.1016/j.socscimed.2013.06.032

Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing progress: Insights on personalized learning implementation and effects*. Research report. RR-2042-BMGF. RAND Corporation. https://doi.org/10.7249/RR2042

Paskett, E., Thompson, B., Ammerman, A. S., Ortega, A. N., Marsteller, J., & Richardson, D. (2016). Multilevel interventions to address health disparities show promise in improving population health. *Health Affairs (Project Hope), 35*(8), 1429-1434. https://doi.org/10.1377/hlthaff.2015.1360

Phelan, J. C., & Link, B. G. (2015). Is racism a fundamental cause of inequalities in health? *Annual Review of Sociology, 41*(1), 311-330. https://doi.org/10.1146/annurev-soc-073014-112305

Ramey, S. L., Schafer, P., DeClerque, J. L., Lanzi, R. G., Hobel, C., Shalowitz, M., . . . the Community Child Health Network. (2015). The preconception stress and resiliency pathways model: A multi-level framework on maternal, paternal, and child health disparities derived by community-based participatory research. *Maternal and Child Health Journal, 19*(4), 707-719. https://doi.org/10.1007/s10995-014-1581-1

Reskin, B. (2012). The race discrimination system. *Annual Review of Sociology, 38*(1), 17-35. https://doi.org/10.1146/annurev-soc-071811-145508

Rosnow, R. L., & Rosenthal, R. (1992). Statistical procedures and the justification of knowledge in psychological science. In *Methodological Issues & Strategies in Clinical Research.* (295-314). American Psychological Association.

Sakurai, K., Kawakami, N., Yamaoka, K., Ishikawa, H., & Hashimoto, H. (2010). The impact of subjective and objective social status on psychological distress among men and women in Japan. *Social Science & Medicine, 70*(11), 1832-1839.

Seaton, E. K., Gee, G. C., Neblett, E., & Spanierman, L. (2018). New directions for racial discrimination research as inspired by the integrative model. *American Psychologist, 73*(6), 768.

Simonds, V. W., & Christopher, S. (2013). Adapting Western research methods to Indigenous ways of knowing. *American Journal of Public Health, 103*(12), 2185-2192. https://doi.org/10.2105/ajph.2012.301157

The UW Implementation Science Resource Hub. (2021). *What is Implementation Science?* University of Washington. https://impsciuw.org/implementation-science/learn/implementation-science-overview/

Trent, M., Dooley, D. G., Dougé, J., AAP Section on Adolescent Health, AAP Council on Community Pediatrics, AAP Committee on Adolescence. (2019). The impact of racism on child and adolescent health. *Pediatrics, 144*(2), e20191765. https://doi.org/10.1542/peds.2019-1765

Trickett, E. J., & Beehler, S. (2013). The ecology of multilevel interventions to reduce social inequalities in health. *American Behavioral Scientist, 57*(8), 1227-1246. https://doi.org/10.1177/0002764213487342

Tuck, E. (2011). Rematriating curriculum studies. *Journal of Curriculum and Pedagogy, 8*(1), 34–37. https://doi.org/10.1080/15505170.2011.572521

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23*(1), 87-102. https://doi.org/10.3758/s13423-015-0892-6

White, A., Thornton, R. L. J., & Greene, J. A. (2021). Remembering past lessons about structural racism — Recentering black theorists of health and society. *New England Journal of Medicine, 385*(9), 850-855. https://doi.org/10.1056/NEJMms2035550

William T. Grant Foundation. (n.d.). *Research grants on reducing inequality*. Review criteria. http://wtgrantfoundation.org/grants/research-grants-reducing-inequality#review-criteria

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy, 8*, 78-115.

Williams, D. R., Lawrence, J. A., & Davis, B. A. (2019). Racism and health: Evidence and needed research. *Annual Review of Public Health, 40*, 105-125. https://doi.org/10.1146/annurev-publhealth-040218-043750

William T. Grant
F O U N D A T I O N

60 East 42nd Street, 43rd Floor
New York, NY 10165
212.752.0071

wtgrantfoundation.org
info@wtgrantfdn.org
@wtgrantfdn