**RE-SEGREGATING HEALTHCARE:** 

# FINDING THE FLAWS IN A FAMOUS – AND DANGEROUS – STUDY

Dr. Stanley Goldfarb, Chairman, Do No Harm Alexander Raikin, Senior Fellow, Do No Harm



# I. THE OAKLAND STUDY: AN OVERVIEW

It was the medical study heard 'round the world. In June 2018, followed by a revised version in August 2019, Marcella Aslan, Owen Garrick, and Grant Graziani wrote "Does Diversity Matter? Experimental Evidence from Oakland." They posited that Black physicians treat Black patients better than White or Asian physicians, all else being equal. The Oakland study was widely covered in the media<sup>1</sup> and is widely cited in the research community to this day.<sup>2</sup> Its central claim has been used to justify an aggressive push for racial diversity in medical schools and residency and training programs. More broadly, it is the pivotal study for what is now known as "racial concordance" — the idea that patients should generally or even exclusively see physicians of the same race, which is tantamount to segregation.

Its findings are startling. The authors state that racially matching physicians and patients "could lead to a 19 percent reduction in the Black-White male cardiovascular mortality gap and an 8 percent decline in the Black-White male life expectancy gap."<sup>3</sup> This and other findings are now spurring action from the healthcare community. For instance, despite an internal survey finding fierce opposition from physicians and consumers, UnitedHealth Group, the largest health insurance company in America, is in the process of publicly collecting and disseminating the race of physicians to incentivize patients to choose a health care provider of their same race.<sup>4</sup> Black physicians are already being encouraged to treat Black patients over non-Black patients — and vice-versa.

But there is a major problem with the Oakland study: Quite simply, it can't prove what it wants to prove, that Black patients should seek out Black doctors for better health outcomes.

Despite its flaws, the study holds sway, and for a good reason. Medical journals are almost entirely hostile to arguments that push back on the paradigms of anti-racism and diversity, equity, and inclusion, which is the foundation of the Oakland study. For instance, one article that found a weakness in the study was retracted. Its author, Norman Wang, was "canceled" by the cardiology community, in part over his critique of the



<sup>1.</sup> Kolata G. The Secret to Keeping Black Men Health? The New York Times, August 20, 2018. Accessed August 1, 2022. <u>https://www.nytimes.com/2018/08/20/health/black-men-doctors.html</u>.

<sup>2.</sup> Google Scholar. Marcella Alsan. n.d. Accessed August 1, 2022. <u>https://scholar.google.com/citations?user=q1QT4pgAAAA-J&hl=en&oi=sra</u>.

<sup>3.</sup> Alsan M., Garrick O .and Graziani G. Does Diversity Matter for Health? The American Economic Review. 2019; 109 (12):4071-4111.

<sup>4.</sup> United Health Care. UHC Provider Directories. December 2020. Accessed August 1, 2022. <u>https://www.scribd.com/docu-ment/494330436/EXTERNAL-UHC-Provider-Directories-Choice-and-Diversity.</u>

Oakland study.<sup>5</sup> Yet there are at least four major flaws in the design and interpretation of the Oakland study. Taken together, they call into question its conclusion, which rejects the long-standing scientific consensus that a physician's race has no relevant impact on health outcomes.<sup>6</sup>

Before exploring the Oakland study's flaws, it's important to consider the ways in which institutionally segregating physicians by race injures the various parties involved in healthcare. First, it damages the trust between physicians and patients. To Black patients, exaggerated claims from a poorly designed and executed study will wrongly tell them that non-Black physicians are less able to treat their illnesses — even when such a physician might be the most qualified or appropriate one for the situation. For Black physicians, this study implies that even when they perform better than non-Black doctors, it is not because of the quality of their medical expertise, but rather because their work is —to a significant degree—skin-deep. To medical organizations, it gives an imprimatur to inch closer to the slippery slope of race-based medicine. To medical premises in the pursuit of a cause they deem worthy. And to White patients, it raises the ugly possibility that they should seek out White doctors to achieve optimum health outcomes.

The result, in each of these cases, is that the people most vulnerable—the patients—will be hurt as healthcare schisms along racial lines. This is precisely what the Oakland Study will do. Indeed, it is already doing that.

### **The Oakland Study Model**

The Oakland study was designed to test prospectively whether Black patients would accept preventative recommendations at higher rates if they were made by a Black physician. Black men were recruited from mostly Black barbershops in Oakland, California, with a small financial incentive to come to a medical clinic for a preventative check-up. When they arrived, they were shown a computer tablet with a male doctor's name and photo (Black or White) and were told to select which preventative service they would be willing to accept. They could pick either an invasive test like a cholesterol or diabetes testing (requiring a blood draw) and a flu shot, or some non-invasive tests like measuring blood pressure or body mass index (BMI).



Wang N. Diversity, Inclusion and Equity. Journal of the American Heart Association. 2020; 9(7): <u>https://doi.org/10.1161/JAHA.120.015959</u>.

At first, Wang's paper passed peer review and was published without significant criticism—until months later, a Twitter mob formed and called for its retraction. Citing improper citations, the *Journal of the American Heart Association* relented. The Journal retracted Wang's paper without his approval, while Wang's university removed him from his leadership positions. Wang is currently seeking legal actions against his employer, claiming that it violated his First Amendment rights. The case is ongoing. https://www.medpagetoday.com/special-reports/exclusives/96923

Meghani S. Brooks J. Gipson-Jones T. et al., "Patient–Provider Race–Concordance: Does It Matter in Improving Minority Patients' Health Outcomes?" Ethnicity & Health. 2009 Feb;14(1):107-30. doi: 10.1080/13557850802227031.

The patients subsequently met their assigned physician, either Black or White. The physician was told to convince each patient to accept as many preventative services as possible. The Oakland researchers compared the difference in willingness of the patients to engage in the preventative services before and after they met with their assigned doctors. The key outcome was whether the patients who saw a Black doctor were more willing to undergo the preventative tests than those who saw a White doctor.

The Oakland study reported that Black patients chose more preventative services, especially those that were more invasive, from Black physicians compared to those who saw White physicians. They went on to speculate that the increased willingness to undergo preventative services, if applied to the entire national cohort of Black men, would produce dramatic improvements in morbidity and mortality of Black Americans.

## The Oakland Study's Four Flaws

#### Problem #1: No Control Group

While it is commendable that the authors managed to recruit over 1,300 Black men from 20 Black barbershops and two flea markets from Oakland, that is not sufficient to justify the study's findings. The study design makes it impossible to distinguish the effect of unobserved covariates like the relative effectiveness of the individual Black vs. the non-Black physicians in persuading patients. A control group consisting of White patients experiencing the same test conditions is necessary to assess whether the Black doctors were more effective because they were Black, or because they were better communicators.

The authors realized this problem, but they did so too late to solve it. Instead, they post-facto tried to produce a small study within a study and use the 12 non-Black patients who were recruited incidentally as a pseudo-control group. However, there is a fundamental problem: These non-Black men likely had unobservable *and* observable characteristics that differ from the sample of Black men. They also were not recruited to match the experimental group in important characteristics like age and underlying health status. Given these differences, trying to compare twelve in-sample subjects with 12 out-of-sample subjects could only work if the individuals were closely matched. There is no indication that this happened, or even that the 12 patients themselves were a homogeneous group.

To prove that the race of the physician matters, it is necessary to prove that other factors don't—the most obvious being that a physician is better at their job because of their skills instead of their skin color. Consequently, a control group of non-Black patients is crucial, as it would help determine whether some doctors were simply more



persuasive across all race groups. The authors postulate that the effectiveness of each Black doctor, independent of race, couldn't be an issue. That's because the differences between Black and non-Black doctors manifested most for invasive tasks rather than for non-invasive ones. Yet this is precisely what is to be expected from the design study — the two non-invasive tests, BMI and blood pressure measurements, do not require any substantial effort or discomfort to the patients. Instead, it seems that patients would be more amenable to take a test that would be more relevant for their health — and that a doctor who is a better communicator would be able to convince patients to take these tests first. The assumption that the doctor's race rather than his communication skills and strategies for convincing patients to undergo blood drawing or flu shots is the crucial ingredient for success cannot be justified without a substantial control group of non-Black patients interacting with the same group of doctors.

All the doctors, Black and non-Black, were almost universally rated highly by the subjects. A stunning 99% of the study subjects affirmed that they would recommend the doctor they met to their friends; the average score was 4.8 out of 5. Moreover, when patients were simply shown a picture of a Black or non-Black doctor on a tablet computer, they were equally likely to select the same preventative services irrespective of race of doctor. Even in their own polling of White and Black men, the authors found no difference in the quality of physicians based on race. In other words, the subjects did not rely on racial concordance for these healthcare decisions. Instead, they acted contrary to the authors' claims. It is only when the doctors arrived and tried to persuade the patients that racial concordance became an issue; this could simply be explained by the fact that the Black doctors were more effective doctors on average than the non-Black doctors. But since the Oakland study didn't contain a control group of non-Black patients who might have also favored the Black doctors, one can't make generalizable conclusions about the role of racial concordance.

#### Problem #2: Unrepresentative Samples

The Oakland study is really two separate studies. The first study involves the 637 pre-selected patients who attended the clinic. In its "pre-consultation stage," subjects were randomly assigned a picture of a Black or non-Black doctor on a tablet and told to select which medical interventions they were interested in receiving. The goal was to see if the subjects showed any dislike for physicians, simply because of skin color. Then subjects met a randomly assigned Black or non-Black doctor who attempted to persuade them to consent to the invasive preventative services.

The second study, as it were, surveyed 1,490 White and Black men. This sample group was matched to the initial cohort only in terms of their level of formal education. But the two survey groups differ in more dramatic ways than simply education, which means that sampling methodology is not generalizable to the Black male population writ large.



A convenience sample of Black men from Oakland barbershops will likely not be representative of the population of Black men in America. Moreover, the sample of those Black men who then decide to receive a financial incentive to participate in a later experiment will be even more unrepresentative. Unsurprisingly, among those who showed up to the clinic, participants were 13 percentage points more likely to be unemployed (compared to 18% among non-participants), 19 percentage points more likely to have a high school education or less (compared to 44% among non-participants). They were 3.4 years older than non-participants, on average, and less likely to be married. Participants also had significantly worse self-reported health, were less likely to have a primary physician, and were more likely to have visited the emergency room over the last two years.

The problem of an unrepresentative sample keeps getting worse. Surprisingly, the Oakland study didn't probe a startling disconnect: the patients who chose to participate in this experiment were apparently healthier than the average Black person in America. This doesn't make sense for several reasons, not only because of the differences in characteristics listed above: the patients who came to the clinic were also 60% more likely to be uninsured and 340% more likely to be unemployed. Yet, their blood pressure, BMI, cholesterol levels and blood glucose values, were on average, lower than the typical Black man in America.

How to explain this? The obvious answer is self-selection bias. It seems likely that patients selected the screening tests that they believed they would pass, perhaps either for their own self-image or because they believed that good results would secure the financial incentive. Alternatively, the unhealthiest participants were more likely to have chosen not to participate in the study. For instance, someone who was overweight might have opted for a blood pressure measurement (which he might or might not pass) versus a BMI measurement, which he knows would show that he is overweight.

Most damningly, the doctors involved in the study were not representative of the wider physician population. Only eight non-Black doctors and six Black doctors were used. The authors use percentages to imply a far larger sample, noting in one case that 67% of Black doctors in the study were internal medicine doctors, but that simply means that four out of six doctors were internists. To state the obvious, such a non-representative sample of doctors is not generalizable to the medical field writ large.

The Oakland study's authors appear to have recognized these problems, yet their attempts to fix it are not persuasive. For instance, they make what they call "corrections for correlated error structures within doctor" and add various permutations of the study doctors, including dropping the "best" Black doctor and "worst" non-Black doctor. This, however, cannot compensate for the biased data set. The problem lies with their sampling choices, not the competency of any one doctor. Similarly, balancing doctors "on observables in age, experience and medical school rank" is not a proper substitute



for the unobserved variables of the physician skills in communication, persuasion, and diagnostic knowledge. The Oakland study showed that some Black doctors convinced some Black patients to take more preventative healthcare services—any further argument is unfounded.

#### Problem #3: Unproven Logical Leaps

The researchers rely on a dubious and unproven projection of the value of and subjects' long-term adherence to preventative healthcare services. They argue that "part of the mortality disparity [between non-Blacks and Blacks] is related to underutilized preventative healthcare services" and these preventative services can be represented by five one-time tests (flu vaccine, blood pressure measurement, BMI measurement, cholesterol testing, and diabetes testing). It is from this instrumentalization that the researchers extrapolate huge increases in Black life expectancy from physician-patient racial concordance.

There are many reasons to question this view. Preventative screening is only effective if used in a patient who has a high likelihood of being at risk for a particular clinical condition and who is committed to following any recommended treatments. A healthy young, non-obese man does not need to be screened for diabetes; similarly, someone who had a normal cholesterol level six months ago would not necessarily need another test. Doctors selected for this experiment were told that they needed to encourage every patient to take every test, even when a test might have minimal benefits for a given patient.

More importantly, not only is the one-time use of these five tests not necessarily part of proper preventative services, but it does not define a well-rounded set of preventative services. Primary interventions, like the tests administered, are only a minor component of such services. They need to be combined with lifestyle changes and, possibly, compliance with a consistent medication regime to be effective.<sup>7</sup> All the health outcomes of the procedures performed in this experiment, apart from the flu shot, rely on modifying what patients do in the future: alter their activity level, reduce their weight, and possibly, employ anti-hypertensive medication. Yet, there was no attempt to record the impact of these one-time interventions by using a follow-up survey.

A conclusion of this study — especially a dramatic claim against the contested academic literature — that hiring more Black doctors will lower cardiovascular deaths among Black men by 19% is a wild and unjustified claim. The researchers themselves acknowledge that meta-analysis studies are inconclusive about the patient outcomes of increasing the frequency of racial and gender matches between doctors and their patients. Other meta studies they did not cite claimed that racial concordance between patients and



Levine S, Malone E, Lekiachvili A, Briss P. Health Care Industry Insights: Why the Use of Preventive Services Is Still Low. Preventing Chronic Disease. 2019;16:180625. DOI: <u>http://dx.doi.org/10.5888/pcd16.180625</u>.

doctors could lead to perceived better communication, but — this is the key point — it has no effect on quality of medical outcomes.<sup>8</sup> Instead, the Oakland study researchers prefer to base their healthcare estimates from what they describe as back-of-the-envelope calculations — a far cry from a scientific projection.

Dramatic claims require dramatic evidence to be credible. But the evidence presented by the Oakland study does not support its conclusions.

#### Problem #4: Missing mechanism for racial concordance

This is the final and most uncomfortable problem: It is unclear how the mechanism of Black doctors interacting with Black patients increases health outcomes. The authors try their hardest to avoid claiming that either the patients or the doctors are racist (in academese, they call it "taste-based discrimination"). Patients, for instance, had no preference for doctors with their own racial background before they met their assigned doctor; all doctors were rated highly at the end of the study. But why is it then that Black patients would trust Black doctors more than non-Black doctors, if both sets of doctors were regarded as trustworthy? The proposed mechanism of easier communication presupposes the racial essentialism that a Black man must have something in common with another Black man solely because of his race. If the assertions for the Oakland paper are true, then it seems that "taste-based discrimination" is the guiding phenomenon for health outcomes — which the researchers deny.

Why would Black Americans not use preventative medical care? A suggested reason is that current African American mistrust in medical care is caused by racist historical events, notably the Tuskegee syphilis experiments.<sup>9</sup> (The authors say this is potentially the reason why a subject rejected a flu shot because he feared "being experimented on," which they say was "a possible reference to the syphilis experiment in Tuskegee." But his suspicion could have, instead, reflected contemporary anti-vaccine dialogue.) To test this claim that American Blacks are most influenced by the history of events like the Tuskegee experiments, one can compare the health and economic status of Blacks in the United States with a similar country without an extended history of slavery and segregation, such as Canada. This involves comparing the health of Black immigrants to Canada and the United States. (Canada accepts a relatively more highly skilled group of immigrants, so it is necessary to make statistical adjustments for education and other covariates.) The finding of one such promising study is that the rates of social and eco-



Shen MJ, Peterson EB, Costas-Muñiz R, et al. The Effects of Race and Racial Concordance on Patient-Physician Communication: A Systematic Review of the Literature. Journal of Racial Ethnic Health Disparities. 2018 Feb;5(1):117-140. doi: 10.1007/s40615-017-0350-4.

<sup>9.</sup> Alsan M. Wanamaker M. Hardeman RR. The Tuskegee Study of Untreated Syphilis: A Case Study in Peripheral Trauma with Implications for Health Professionals. Journal of General Internal Medicine. 2020 Jan;35(1):322-325. doi: 10.1007/s11606-019-05309-8.

nomic integration "are virtually identical" in almost all supposed social determinants of health, ranging from social equality to income.<sup>10</sup> Consequently, it seems unlikely that the current mistrust Black Americans have derives from past historical events. Instead, any mistrust would have to be based on current phenomena, and the researchers did not attempt to make or support this argument.

# II. HOW WERE INACCURATE CLAIMS SPREAD?

What happens when loose scientific claims, like those in the Oakland Study, are released in the general scientific literature? Common sense implies that the academic medical community would not merely praise a paper but would offer an even-handed approach. Without skepticism, the scientific method is impossible: the pursuit of truth requires well-founded criticism to probe the weaknesses and strengths of any argument, even more if we believe the claim in question. If something sounds true, there must be evidence that it is true. Yet the Oakland study has been all but unchallenged, even as its unproven assumptions and faulty design have formed the foundation of subsequent medical research.

We conducted two analyses showing the spread of the Oakland study in the academic literature. The first is a citation tree diagram, counting the number of times a study is cited in another paper. The second follows the methodology of Connected Papers, examining the similarities in a given number of frequently cited papers based on their mutually overlapping citations and references. For the former, we placed each study in at least one of several categories: it mentioned the results of the Oakland study; it interpreted the results of the Oakland study; it included any potential limitations of the study; it accurately portrayed the results of the Oakland study and if cited it favorably. Not surprisingly, every paper save one agreed with the Oakland study and mentioned it favorably. There was no critical discussion.

Worse, there was virtually no consideration of the study's limits. Only Norman Wang's study discussed potential problems with the Oakland study, stating:



Reitz J. Multiculturalism Policies and Popular Multiculturalism in the Development of Canadian Immigration. September 2013. Accessed August 1, 2022. <u>https://www.yumpu.com/en/document/view/51633638/multiculturalism-policies-and-popular-mul-ticulturalism-university-of-</u>

Yet residual confounding in patient-physician racial and ethnic concordance studies is impossible to eliminate as physicians of the same race and ethnicity are not interchangeable. The results have little external validity as the study only involved 14 physicians (8 non-Black and 6 Black). Moreover, mortality estimates were extrapolated from single patient-physician encounters using methods so unscientific that the investigators themselves described them as "back-of-the-envelope calculations."<sup>11</sup>

But here too, the story takes a twist. This paper by Norman Wang was retracted almost six months after the American Heart Association published it. The AHA said the article "contains many misconceptions and misquotes and that together those inaccuracies, misstatements, and selective misreading of source materials void the paper of its scientific validity."<sup>12</sup> Wang rejects this characterization. He believes that his essay was retracted because he opposes affirmative action in cardiology programs, which lead to him being fired from his leadership positions and disciplined by his academic institution.<sup>13</sup> It has become academic suicide to claim like Wong that "all who aspire to a profession in medicine and cardiology must be assessed as individuals on the basis of their personal merits, not their racial and ethnic identities."

The Oakland study has turned into a cottage industry for those who write and publish similar academic articles, reinforcing a particular bias. Consider how it is graphically related to similar papers (Figure one ).<sup>14</sup> The black circle is the original study; the blue circle contains mostly diversity papers; the red circle contains papers on anti-Blackness studies, an offshoot of critical race theory ; the yellow circle is the literature on racial and gender concordance between physicians and patients. What is clear from Figure 4 is that a significant cohort of earlier studies, like that of Jerant et al.<sup>15</sup> and Smith et al.<sup>16</sup> which found no statistically significant relationship between racial and sexual concordance of physician and patient, are ignored in favor of the highly speculative conclusions derived from the Oakland study. Yet the literature changed. Newer studies in the red circle took the supposed benefit of racial concordance and imported it to the field of anti-Blackness studies; similarly, in the blue circle, it was applied to the academic literature on diversity. The major question that remains: will acting on the Oakland study benefit Black patients?



<sup>11.</sup> Wang, "Diversity, Inclusion, and Equity."

Retraction to: Diversity, Inclusion and Equity. Journal of the American Heart Association. 2020; 9(20). <u>https://doi.org/10.1161/JAHA.119.014602</u>.

D'Ambrosio. Cardiologist Who Criticized Affirmative Action Proceeds With Free Speech Battle. MedPage Today. <u>https://www.medpagetoday.com/special-reports/exclusives/96923</u>. January 28, 2022. Accessed August 1, 2022.

<sup>14.</sup> Connected Paper. Does Diversity Matter? Experimental Evidence from Oakland. Accessed August 1, 2022. <u>https://www.con-nectedpapers.com/main/10ff9c24d738219fc9b306b13ba7917c5eff2ff3/Does-Diversity-Matter-for-Health%3F-Experimental-Evidence-from-Oakland/graph</u>.

<sup>15.</sup> Meghani S. Brooks J. Gipson-Jones T. et al. Patient-Provider Race-Concordance: Does it Matter in Improving Minority Patients' Health Outcomes?

Street R. O'Malley K. Cooper L. et al. Understanding concordance in patient-physician relationships: personal and ethnic dimensions of shared identity. Annals of Family Medicine. 2008; 6(3):198–205. doi: 10.1370/afm.821.

# III. CONCLUSION

The Oakland study is not a landmark paper that proves the United States must train more Black physicians to achieve better health outcomes for Black patients, much less push patients to rely primarily or solely on physicians of their own race or gender. Rather, it is an unproven and unpersuasive effort to convince policymakers of a dubious plan to make racial concordance a reality in American healthcare — which in essence means a return to racial segregation of medical care. The Oakland study should have been challenged by an army of researchers — people who typically demand rigor in design and interpretation before they accept studies as definitive. Instead, it was treated with broad and unthinking acceptance. The mounting campaign for a 21<sup>st</sup> century segregation of healthcare is dangerous and baseless. As this analysis shows, the Oakland study does not change that fact.





### Dr. Stanley Goldfarb Chairman

#### Stanley Goldfarb, MD, is the Chairman of the Board for Do No Harm. He is a former associate dean of curriculum at the University of Pennsylvania's Perelman School of Medicine and lives in Philadelphia.

### Alexander Raikin Senior Fellow

Alexander Raikin is a Senior Fellow with Do No Harm. He is an independent researcher and writer based in Washington D.C.



# DONOHARMMEDICINE.ORG