

How analytics, AI tools can overlook multiracial patients

JESSICA KIM COHEN

Data-driven tools that incorporate race have to figure out how to account for patients who identify with more than one race, as multiracial people—today, 10% of Americans—become larger segment of the U.S. population. Hospitals and health systems are rolling out more tools that analyze and crunch data to try to improve patient care—raising questions about when and how it's appropriate to integrate race and ethnicity data.

Racial data has grown more complicated as the U.S. becomes increasingly diverse, with a growing number of Americans identifying with more than one race or ethnicity.

The number of Americans who identify with at least two races has doubled over the past decade, according to last year's U.S. census, which takes place every 10 years. The Census Bureau started letting people identify as more than one race in 2000, according to the [New York Times](#). It's now the fastest-growing racial and ethnic category.

That's a demographic shift that executives should keep top-of-mind as the healthcare industry moves toward being more data-driven. If an analytics or artificial-intelligence tool incorporates whether a patient is Black, white or another race into its prediction, that could lead to confusion for a patient who's Black and white, for example.

Multiracial patients represent a growing population that needs to be accounted for in AI and other data-driven tools, said Tina Hernandez-Boussard, an associate professor of medicine in biomedical informatics, biomedical data science and surgery at Stanford University.

If health systems and software developers aren't considering ways to ensure multiracial patients are accounted for when using algorithms or protocols that rely on race, such models may not be reliable for that patient population, she said. That could erode trust that patients have in the health system.

"It is highly complicated," Hernandez-Boussard said. "By developing algorithms that are not particularly tailored for this growing population, we lose the trust of that community."

Predicting risk

Healthcare organizations in recent years have been investing in tools that assess data to flag patients in need of additional care those at risk for poor outcomes and who may have other needs. More than three-quarters of acute- and ambulatory-care organizations are using advanced analytics for population health, according to a [survey](#) from the College of Healthcare Information Management Executives.

Some of those tools—everything from basic risk equations to advanced AI—incorporate race, but not always in ways that account for the U.S.'s growing multiracial population.

"How should we care best for individuals that identify as multiple races?" Dr. Michael Simonov, director of clinical informatics at hospital-backed data company Truvena, said of risk calculators and predictive models that incorporate race and ethnicity data. "That's an open question and a very active area of research."

Several [risk prediction algorithms](#), which have been used in medicine for years, ask clinicians to report whether a patient is Black or white as part of their calculation.

A tool that estimates a patient's 10-year risk of atherosclerotic cardiovascular disease requires a user to select a patient's race as "white," "African American" or "other," which could leave uncertainty for a patient who is Black and white—particularly if the patient only selected one race on their intake forms or if a doctor assumes race based on the patient's appearance. This year the National Kidney Foundation and the American Society of Nephrology [released](#) an equation to estimate kidney function that doesn't include race—replacing an existing version that asked whether a patient was Black. A [calculator](#) used to predict the risk to a patient if they have a vaginal delivery after a C-section in a previous pregnancy removed race this year, too. "If a physician has been trained to view race as a risk factor and they're encountering a patient who doesn't fit into a clean category of race, then it's very difficult for them to make the assessment that they've been trained to do," said Dr. Megan Mahoney, chief of staff at Stanford Health Care and

clinical professor in the department of medicine at Stanford University. I don't fit into any clean category for the use of their calculator," added Mahoney, who is Black and white.

Mahoney said she said she wants to see more data tools and calculators follow in the footsteps of the equation to estimate kidney function, moving away from incorporating race at all.

Next generation medicine

AI, which for years has been touted as the future of healthcare, could pose an opportunity for incorporating multiracial and multiethnic data—if developers have the right data to work from.

Unlike other analytics or modeling approaches, which tend to rigidly collect specific types of data to calculate an outcome, advanced AI is more flexible—able to ingest more variables as well as complex and multilayered data that it hasn't been explicitly programmed to handle, said Dr. Russ Cucina, chief health information officer at UCSF Health.

But [good algorithms start with good data](#).

For an AI tool to be able to produce generalizable insights, it needs to analyze a massive amount of data that's reflective of the population the tool will be used with.

To create an AI system, developers feed the AI reams of training data, from which they can learn to identify features and draw out patterns. But if that dataset isn't diverse and lacks information on some subpopulations, the predictions and recommendations from the system might not be as accurate for those patient groups.

[Healthcare providers](#) and [advocacy groups](#) have increasingly been challenging whether to even incorporate race data into algorithms, arguing race has inappropriately been used as a proxy for other variables linked with risk of illnesses, like ancestry, genetics, socioeconomic status, or the environment in which a patient lives.

Using that data, instead of race, would be more appropriate, they say. But even if race isn't included as a variable in an algorithm, it's important to have a diverse dataset available to validate AI tools—so that organizations

can test the product against specific subpopulations and ensure it performs well across demographics.

"We see a lot of examples of the problems that can result when we don't have good representative samples of data when we're developing these algorithms," said Dr. Peter

Embi, president and CEO of the Regenstrief Institute. Embi joins Vanderbilt University Medical Center as chair of the biomedical informatics department in January.

In dermatology, for example, [researchers have said](#) skin-cancer detection AI tools primarily trained on images of light-skinned patients may not be as accurate for dark-skinned patients.

More research is needed to figure out in what cases noting that a patient has multiple races or ethnicities would improve accuracy of a predictive tool, said Suchi Saria, professor and director of the Machine Learning and Healthcare Lab at Johns Hopkins University and CEO of Bayesian Health, a company that develops clinical decision-support AI.

Getting the right data

But even accumulating enough data on multiracial patients to train or validate an AI system is challenging.

Only about 10% of Americans are multiracial. That's a diverse label in and of itself, encompassing people who could be white and Black, Black and Asian, Asian and Native American, to name a few examples—and not to mention patients who would select more than two races.

Patient data often isn't captured granularly enough in medical records to identify multiracial patients.

Based on Bayesian Health's experience working with hospital customers' EHR data, Saria said she suspects multiracial patients are undercounted in medical records.

Only about 1% of patients in the data the company's worked with were recorded as having multiple races, she said.

That could be because multiracial patients are often grouped into an "other" category or might select just one of the races they identify with.

Gathering enough data for research, development and validating of analytics, AI and other data-driven tools will be key to ensuring they work effectively for patients with diverse backgrounds.

"If we did have the data, then yes, an algorithm would be able to appropriately deal with those issues," Hernandez-Boussard said. "But the problem is we don't have data to train the[algorithms] appropriately."

Inline PlaySource URL: <https://www.modernhealthcare.com/information-technology/how-analytics-ai-tools-can-overlook-multiracial-patients>