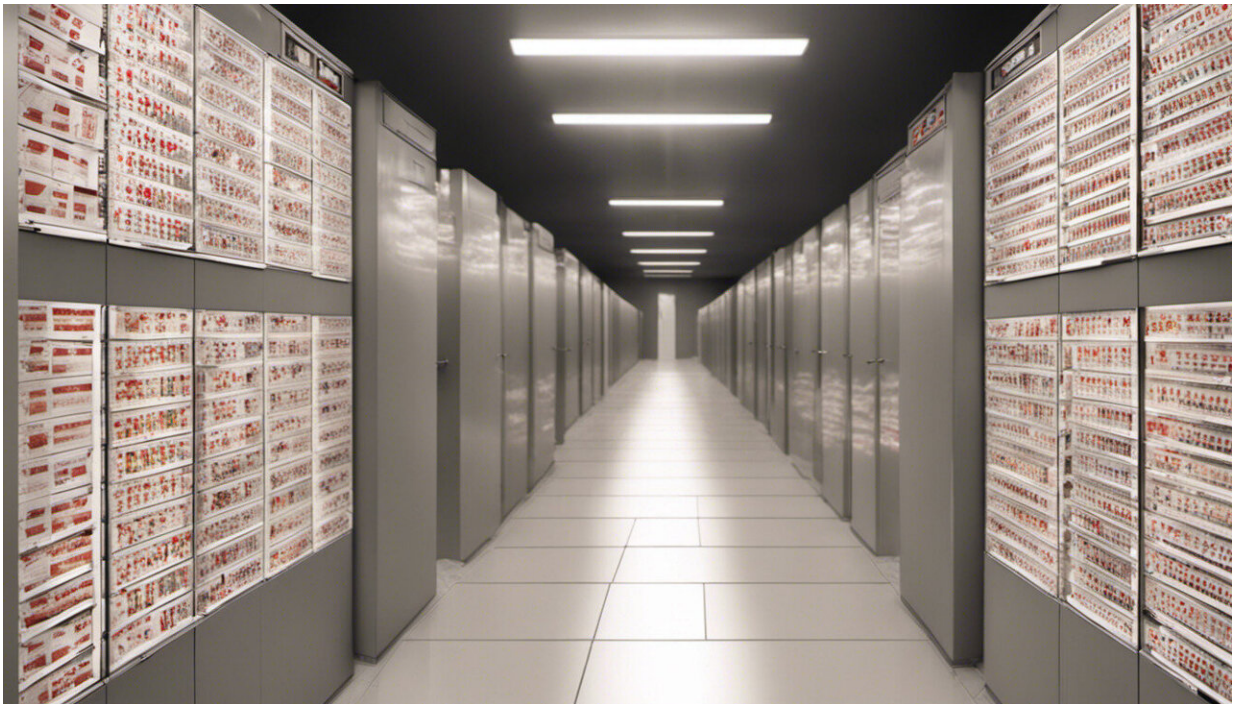# How algorithms can create inequality in health care, and how to fix it

December 5 2018, by Matt Wood

Credit: AI-generated image ([disclaimer](#))

Machine-learning algorithms and artificial intelligence software help organizations analyze large amounts of data to improve decision-making, and these tools are increasingly used in hospitals to guide treatment decisions and improve efficiency. The algorithms "learn" by identifying patterns in data collected over many years. So, what happens when the

data being analyzed reflects historical bias against vulnerable populations? Is it possible for these algorithms to promote further bias, leading to inequality in health care?

Marshall Chin, the Richard Parrillo Family Professor of Healthcare Ethics at University of Chicago Medicine, is working to ensure equity across all areas of the healthcare system, including data analysis. He has worked for three decades to examine and develop solutions addressing health disparities. Chin recently teamed up with a group of data scientists from Google to write an article in the Annals of Internal Medicine that discusses how [health care providers](link) can make these powerful new algorithms fairer and more equitable. We spoke to him about the use of machine learning in health care, and how doctors and patients can build fairness into every step of the decision-making process.

How widely are these kinds of algorithms used in health care now?

It varies across different settings, but they're increasingly being used for clinical care, like reading X-rays and images to diagnose conditions like eye disease or skin cancer. They're also being used from a business perspective to analyze medical records and insurance claims to increase the efficiency of the organization and lower costs.

Do you think people realize how software and algorithms are used to determine their care?

The phrase "[big data](link)" is popular because there's so much data collected on all of us, whether it's health data or on the internet. Big data are a powerful tool, but we need to clearly and explicitly discuss the ethical implications of how software can analyze and use big data. Those issues

are still hidden to most people.

What's an example of how algorithms can create inequality in health care?

I'll give you an example we included in our article. There is an outstanding data analytics group at the University of Chicago Medicine, and one of the things they do is create algorithms to analyze data in the electronic medical records. One of the projects they're working on is to help decrease the length of stay for patients, because it's in everyone's best interest to have patients go home as soon as they're ready to leave. The thought was if we can identify patients who are most likely to be discharged early, we can assign a case manager to make sure there are no further blockages or barriers that could prevent them from leaving the hospital in a timely manner.

The data analytics group initially developed the algorithms based on clinical data, and then they found that adding the zip code where the patient lives improved the accuracy of the model identifying those people who would have shorter lengths of stay. The problem is when you add a zip code, if you live in a poor neighborhood or a predominantly African-American neighborhood, you were more likely to have the longer length of stay. So, the algorithm would have led to the paradoxical result of the hospital providing additional case management resources to a predominantly white, more educated, more affluent population to get them out of the hospital earlier, instead of to a more socially at-risk population who really should be the ones that receive more help.

What happened after they discovered this?

To their credit, when the data analysis team saw this, they were horrified by the implications of what would've happened. They ended up working

with James Williams, the director of Diversity, Inclusion and Equity, and our Diversity and Equity Committee at UCM. They became champions of developing formal systems to make sure these equity issues are explicitly thought about as they develop algorithms, as well as how they could proactively use machine learning to improve equity.

So with that local example, it became clear that this is not just an abstract, theoretical issue. It's happening here and probably a lot of other places under the radar. I think a lot of health care organizations aren't intentionally trying to do things that are going to worsen disparities, but clearly there are a lot of unintentional bad things that can happen. Probably very few organizations are proactively using these tools to improve outcomes for everyone.

How does inequality creep into algorithms? Is it because the input data aren't representative of different groups of people? Or does it reflect historical bias?

You might think of three buckets that cause the problems. One is the data itself. The second one is the algorithms, and the third is how the algorithms are used.

You mentioned that the data may be biased. For example, poor populations can have more scattered, fragmented care. They may come here for some hospital admissions and go to Stroger Cook County Hospital for something else, as opposed to some of the more affluent patients who have a continuous source of care here. So if you're building algorithms based upon only UCM data, you'd have a more complete data set for the more affluent patients. It's likely then that whatever predictions you develop on the incomplete data aren't as accurate.

There's also the issue of perpetuating historical biases. For example, racial and ethnic minority patients can present differently than the textbook, psychiatric definitions of mental illness. So, if you're using flawed criteria to begin with, you're perpetuating incorrect diagnoses. Another example would be something like coronary artery disease, where women tend to be under diagnosed compared to men. If you're using criteria that under diagnose women, you may be building into your formula a perpetually biased underdiagnosis of women.

How do you prevent these unfair outcomes from happening?

Think back to those three buckets of the data, the formulas themselves and how the formulas are used. First, be cognizant that data could be a problem. Are you working with valid data sets? Do you have incomplete data for the at-risk population? Are you using data based upon faulty diagnoses and faulty labels? That's an important first step.

The second step would be examining how the models are actually developed. Here, there are technical ways to design the algorithms to advance specific principles of ethical justice. You can create algorithms that will ensure equal outcomes across two populations, and you can make sure that the technical performance of the model is fair. So, if there's a problem where algorithms are under diagnosing African Americans for some condition, you can alter the parameters of formulas to make them more accurate.

Another way to promote justice is to adjust formulas so you have equal allocation of resources. The previous example about assigning case managers to help people go home from the hospital sooner is a good example. You can alter the thresholds for who qualifies in these formulas to equalize the allocation of actual resources to different

groups.

How do you know it's working?

You can do your best trying to develop a good formula, but you still have to monitor what happens in real life, the third bucket for preventing unfair outcomes. That involves monitoring the data for inequalities and also talking to the health care providers, the patients and the administrators to determine if they see any fairness problems. One thing we recommend is to have patients at the table as we design these algorithms that will ultimately affect their lives.

Software reflects the people who write it, so it's important to have these fairness issues in mind. At every step of the way, we should check if the algorithm is going to lead to an unfair result. Which bad things unintentionally could happen and how can we proactively bake in ways to benefit everyone? And that requires careful attention to each step: picking the data, developing the formula, and then deploying the algorithm and monitoring how it is used.

We must have equity and improving the health of everyone as explicit goals and then build the systems toward those goals. We must build in specific steps where there's a chance for self-reflection: Is what we're doing advancing equity for everyone, or have we unintentionally worsened things?