NO PREP

AI in Healthcare

Helping or Hurting Patients?

Includes: Student reading, case study, differentiated questions, teacher key

MORE LESSONS LESS PLANNING · MLLP · FREE TODAY · FREE TOMORROW · FREE FOREVER

Name: Period: Date:

AI in Healthcare: Helping or Hurting Patients?

The Case for AI in Medicine

Artificial intelligence is already being used in hospitals, clinics, and medical research facilities around the world, and the results in some areas are genuinely impressive. AI-powered systems can analyze thousands of medical images — X-rays, CT scans, MRIs — faster than any human radiologist, and in certain narrow tasks, such as detecting specific types of tumors or eye diseases, some AI systems have demonstrated accuracy comparable to or exceeding that of trained specialists. Google's DeepMind developed an AI system that can detect over 50 eye diseases from retinal scans with accuracy matching that of expert ophthalmologists. AI tools are also being used in drug discovery, genomics research, and predicting patient readmission rates — all areas where processing large amounts of data quickly could genuinely improve outcomes. In emergency departments and intensive care units, AI monitoring systems can track vital signs continuously and flag deteriorating patients before they crash — potentially catching problems that an overworked nursing staff

Name: **Period:** **Date:**

AI in Healthcare: Helping or Hurting Patients?

The Case for AI in Medicine

Artificial intelligence is already being used in hospitals, clinics, and medical research facilities around the world, and the results in some areas are genuinely impressive. AI-powered systems can analyze thousands of medical images (X-rays, CT scans, MRIs) faster than any human radiologist, and in certain narrow tasks, such as detecting specific types of tumors or eye diseases, some AI systems have demonstrated accuracy comparable to or exceeding that of trained specialists. Google's DeepMind developed an AI system that can detect over 50 eye diseases from retinal scans with accuracy matching that of expert ophthalmologists. AI tools are also being used in drug discovery, genomics research, and predicting patient readmission rates: all areas where processing large amounts of data quickly could genuinely improve outcomes. In emergency departments and intensive care units, AI monitoring systems can track vital signs continuously and flag deteriorating patients before they crash, potentially catching problems that an overworked nursing staff might miss during a busy shift. In radiology, AI can serve as a second reader, helping catch cancers that a fatigued radiologist might overlook. These applications represent real promise: AI as a tool that amplifies human expertise, catches human error, and helps medical professionals do their jobs better.

An Algorithm That Was Wrong About Race

In 2019, researchers published a landmark study in the journal Science documenting a serious bias problem in one of the most widely used AI algorithms in American healthcare. The algorithm (developed by Optum, a subsidiary of UnitedHealth Group) was used by hospitals and health systems across the country to identify high-risk patients who should receive additional care management services. The algorithm was used to make care decisions for approximately 200 million patients per year. The

researchers found that the algorithm was significantly less likely to refer Black patients for additional care than equally sick white patients. In fact, to receive the same level of care referrals as white patients, Black patients had to be considerably sicker. The algorithm did not use race as a variable. Instead, it used healthcare costs as a proxy for health needs; the reasoning was that patients with higher medical costs must have more health needs. The problem was that Black patients in the United States spend less on healthcare than white patients with the same health conditions, largely due to longstanding barriers to healthcare access, including historical discrimination, geographic disparities, and insurance gaps. The algorithm interpreted lower spending as lower need, effectively translating systemic inequality in healthcare access into a direct disadvantage in who received care management.

Skin Tone and Dermatology AI

Dermatology, the branch of medicine dealing with skin conditions, has been a particularly active area for AI development. AI tools that analyze photographs of skin lesions and moles have shown impressive accuracy in detecting melanoma and other skin cancers. However, most of these systems were developed and trained primarily on datasets of images from lighter-skinned patients. Skin conditions look different on darker skin, and symptoms that are easy to detect on light skin may present differently or be harder to photograph accurately on darker skin. A 2021 study published in JAMA Dermatology found that most publicly available dermatology AI datasets were severely lacking in representation of darker skin tones. When AI systems trained on those datasets are deployed in clinical settings, they perform less accurately for patients with darker skin. This means the patients at highest risk of having their diagnoses delayed or missed are the same patients who already face disparities in access to dermatological care. The problem compounds: underrepresentation in data leads to less accurate AI, which leads to worse care, which leads to worse health outcomes for groups already facing health inequity.

The Sepsis Prediction Problem

Sepsis is a life-threatening immune response to infection that kills more than 270,000 Americans per year. Early identification is critical — treating sepsis hours earlier can mean the difference between survival and death. Several hospitals have

deployed AI early-warning systems designed to predict which patients are developing sepsis before obvious symptoms appear. These tools sounded like an obvious win: faster detection, earlier treatment, lives saved. Research on deployed sepsis prediction tools has found mixed results. Some systems generate high numbers of false alarms — alerting medical staff to potential sepsis cases that turn out not to be sepsis. When alerts are too frequent, nurses and doctors begin to ignore them — a phenomenon researchers call "alert fatigue." A 2021 study in BMJ investigated one of the most widely used sepsis prediction algorithms, Epic's Sepsis Model, and found that it missed a substantial portion of actual sepsis cases and triggered a large number of false alarms, with performance that varied significantly across patient subgroups. A single flawed algorithm, deployed at scale across multiple hospitals, can affect care for enormous numbers of patients before problems are identified.

Regulation, Oversight, and the Path Forward

The U.S. Food and Drug Administration (FDA) regulates AI software used as a medical device, but the regulatory framework for AI in healthcare is still developing. Unlike a drug or a traditional medical device, AI systems can be updated continuously, which makes ensuring consistent safety and performance difficult. The FDA has approved hundreds of AI-based medical devices, primarily in imaging and diagnostics, but independent researchers have raised concerns that approval processes have not always required rigorous real-world performance testing across diverse patient populations. The path forward that most researchers and ethicists advocate involves several elements: mandatory demographic performance reporting for AI medical tools (not just average accuracy, but accuracy broken down by race, sex, age, and socioeconomic status); independent validation studies before wide deployment; ongoing monitoring after deployment; and transparency requirements so that clinicians know when AI is influencing a recommendation and can apply their own judgment. The promise of AI in healthcare is real, but realizing it equitably requires confronting the fact that systems built on unrepresentative data will reproduce existing health disparities at scale. The promise of AI in healthcare is real and should not be dismissed. Early detection of cancer, faster diagnosis of rare diseases, continuous monitoring of high-risk patients — these are genuine contributions that could save lives at scale. The problem is not AI itself but the

conditions under which it is developed and deployed. A healthcare AI system that works well for some patients while systematically underserving others does not represent a neutral technological advance; it represents a choice to distribute benefits unequally. Holding that standard explicitly — requiring that new tools demonstrate comparable performance across demographic groups before wide deployment — would not so much slow progress as redirect it toward progress that actually serves everyone.

Key Information: AI in Healthcare

Detail	Information
DeepMind eye disease tool	Detects 50+ eye diseases from retinal scans with expert-level accuracy
Optum algorithm (2019 study)	Used by ~200M patients/year; found to under-refer Black patients for care management
Why Optum algorithm was biased	Used healthcare costs as proxy for health needs; Black patients spend less due to access barriers
Dermatology AI problem	Most datasets lack darker skin tones; AI less accurate for darker-skinned patients
Sepsis prediction issue	High false alarm rates cause "alert fatigue"; algorithms miss cases across patient subgroups
Epic's Sepsis Model finding	2021 BMJ study: missed many actual cases and triggered many false alarms
FDA role	Regulates AI as medical device, but framework still developing for continuously updated systems
Recommended path forward	Demographic performance reporting, independent validation, ongoing monitoring, transparency

Case Study: Nurse Priya's Alert

Priya is a registered nurse in a hospital that deployed an AI sepsis prediction system eight months ago. In the first two months, the alerts seemed useful — they caught a few patients who were declining. But lately, the alerts fire constantly. On a recent night shift, Priya received 14 sepsis alerts for patients on her unit. She investigated all of them. Eleven were false alarms. Now, when the alert fires, her instinct is to do a quick check and move on. "I used to stop everything when the alarm went off," she tells a colleague. "Now I treat it like the car alarm in the parking lot. You hear it, but you don't really hear it." Last week, a patient in Room 412 got the alert. Priya checked, saw the patient was stable, and moved on. By morning, the patient was in the ICU with a confirmed sepsis diagnosis. Priya cannot stop thinking about it. She does not blame herself entirely — she was managing twelve other patients — but she wonders whether she would have caught the signs earlier if she still trusted the alarm.

1. Based on the reading, what specific problem does Priya's experience illustrate, and what is the technical term used in the reading for this problem?

2. Priya's case raises a question about human judgment and AI tools in medicine. Using evidence from the reading, explain what the reading suggests should happen when AI alerts conflict with a nurse's or doctor's direct observation.

Multiple Choice

1. Why did the Optum healthcare algorithm under-refer Black patients for care management, according to the reading?

a) The algorithm was explicitly programmed to reduce costs for Black patients

b) Black patients scored lower on the self-reported health surveys that the algorithm used to rank medical need

c) It used healthcare spending as a proxy for health need, but Black patients spend less due to access barriers

d) The developers tested the algorithm only on white patient populations

2. What is "alert fatigue" as described in the reading?

a) A technical failure that causes AI systems to stop sending alerts after extended use

b) A condition in which patients become anxious when too many medical devices monitor them all at once

c) A regulatory term for AI systems that produce alerts without FDA approval

d) A phenomenon in which clinicians begin ignoring alerts because they are triggered too frequently

3. What does the reading identify as a key element of responsible AI deployment in healthcare?

a) Requiring all AI healthcare tools to be built only from U.S. patient data

b) Mandatory demographic performance reporting broken down by race, sex, age, and socioeconomic status

c) Replacing all human radiologists with AI systems by 2030

d) Allowing hospitals to self-certify their own AI tools without any independent federal review process

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ Optum algorithm	A) When clinicians begin ignoring alerts because they fire too frequently and are often wrong
2. ___ Alert fatigue	B) A life-threatening immune response to infection that kills over 270,000 Americans per year
3. ___ Proxy variable	C) Google AI division that developed a system matching expert accuracy in detecting eye diseases
4. ___ DeepMind	D) An indirect measure used as a substitute for a variable that is harder to observe directly
5. ___ Sepsis	E) Healthcare AI that under-referred Black patients by using spending as a stand-in for health need

Fill-in-the-Blank

1. The Optum algorithm was used to make care decisions for approximately _____ million patients per year.

2. A 2021 study in JAMA Dermatology found that most publicly available dermatology AI datasets lacked representation of _____ skin tones.

3. The FDA regulates AI software used as a medical _____, but the regulatory framework is still developing.

Short Answer

Answer each question in 2-4 complete sentences.

1. Explain how the Optum algorithm discriminated against Black patients without using race as a variable.

2. Why does underrepresentation in dermatology AI training data compound existing health inequities, according to the reading?

3. The reading says the promise of AI in healthcare is real but must be realized equitably. What does equitable mean in this context, and what would it require?