

## **Introduction**

Hi \_\_\_\_,

Would you first like to start off by introducing yourself? Please tell us about your name, what degree you have, where you currently work, and where your research interests lie.

Thank you so much for the introduction. So, today's topic will be about how Biased Clinical Research Data Affects Healthcare Practices. Our interview will consist of 6 specific cases where biased clinical data led to AI systems making incorrect decisions, which specifically stemmed from gender or racial biases. The first three cases will focus on gender-biased biases, while the last three cases will focus on racial-based biases.

Overall, the interview is formatted in a way such that at the beginning of each of the 6 cases, I will give you very little information about the case, and try to bias you in some way. The point of these first few questions is to give you limited information and help you understand why the AI system made the biased prediction that it did. Afterwards, I will give you the full details about the case, and will ask you to analyze the impacts of the biased AI decision, and how we move forward to prevent such cases from ever occurring again for future patients.

Finally, at the very end, after all 6 cases, I will ask you some overarching questions about biased clinical data and how it affects marginalized communities, and what you would suggest or include in similar systems if you were tasked to design one.

## **Case 1: Transgender Patient's Emergency Misdiagnosed by Algorithmic Bias**

### **Case Summary**

A 32-year-old transgender man arrived at the emergency room with severe abdominal pain. Although he informed staff that he was transgender, the hospital's electronic health record (EHR) system listed him as male (Ring, 2019). As a result, clinicians failed to consider pregnancy and misattributed his symptoms to factors like obesity. The patient was, in fact, pregnant and experiencing labour complications. Due to the delay in recognizing this, urgent care was postponed, and tragically, the baby was stillborn (Ring, 2019).

This case highlights how rigid, binary classifications in health records — combined with provider assumptions — can lead to critical misdiagnoses. Standard pregnancy-related alerts were never triggered because the system registered the patient as male. This highlights the dangers of algorithmic bias and the need for more inclusive healthcare systems that reflect the realities of transgender and non-binary individuals in today's day and age (Compton, 2019; Cirillo et al., 2020).

### **Interview Questions**

1. Imagine you're a physician treating a patient with severe abdominal pain and no visible injuries. What diagnoses would immediately come to mind, and what additional information would you ask for?
2. What if you were told the patient was a woman?
3. How might your diagnostic thinking change if you were told the patient was a man?
4. How would learning that the patient is a transgender man affect your clinical reasoning, and what challenges might arise in ensuring they receive appropriate care?

Summarize the case.

5. In your opinion, now knowing that the person was transgender and pregnant, what role did algorithmic bias, system design, or any other factors that you can think of, play in the AI's failure to identify pregnancy risk in this case?
6. What changes would you suggest — to electronic health records, triage systems, provider training, etc. — to prevent similar outcomes for transgender patients in the future?

## **Case 2: Symptom Checker Underestimates a Woman's Heart Attack**

### **Case Summary**

An AI-powered symptom-checker app provided dangerously different recommendations for a man and a woman with identical symptoms. Both were 59-year-old smokers reporting sudden chest pain and nausea. The only variable was gender. The male patient was warned of potential heart issues, including unstable angina or a heart attack, and advised to seek emergency care. The female patient, however, was told her symptoms might be due to depression or anxiety, with no urgent care recommended (Trendall, 2024).

This discrepancy highlights how AI tools can replicate and amplify gender bias, potentially leading to delayed treatment or misdiagnosis (Joshi, 2024). In this case, the AI appeared to rely on statistical trends that underrepresented heart attacks in women—overlooking the real and serious risk of cardiac events in female patients (Cirillo et al., 2020). Public concern grew after this example came to light, as it revealed how gender stereotypes, like “women are anxious and men have heart attacks”, can shape algorithmic decisions (Trendall, 2024).

Ultimately, the case highlights the critical importance of designing healthcare AI systems that recognize and adjust for bias. Life-threatening conditions must be considered for all patients—regardless of gender—especially when symptoms are shared.

### **Interview Questions**

1. Imagine a 59-year-old patient who presents with sudden chest pain and nausea. Without knowing their gender, what would be your top differential diagnoses, and what would your next steps be?
2. Would you take the problem seriously right away, or would you think that chest pain for a 59-year-old is likely not serious?
3. Now imagine learning the patient is a woman, and you're told that statistically, women are less likely to have a heart attack. Would this influence your decisions? Should it?
4. If the patient was a man, would your sense of urgency change, keeping in mind that statistically, men are more likely to have a heart attack?
5. So overall, would your perspective and seriousness stay the same if they were a 59-year-old with chest pain regardless of their gender?
6. Even if the statistical risk is lower in women, how would you balance data like this with the need to consider life-threatening conditions like heart attacks?

Summarize the case.

7. How should an AI triage system handle gender-based statistical differences? Should it present all serious possibilities regardless of probability, especially in potentially fatal scenarios?
8. Was the AI's response defensible, given it followed statistical trends, or should its priority have been patient safety over data-driven averages? Where's the line between data and ethical care? Do you believe it failed in its duty to provide safe recommendations?

### **Case 3: Male Breast Cancer Patient Denied Treatment by Gendered Algorithm**

#### **Case Summary**

Raymond Johnson, a 26-year-old man from South Carolina, was denied Medicaid coverage for breast cancer treatment solely because of his gender. Although diagnosed with breast cancer, he was ruled ineligible for a federal Medicaid program that covered patients screened through certain government programs — programs that only screened women. As a result, the insurance algorithm automatically excluded men, leaving Raymond without coverage for the chemotherapy and surgery he urgently needed (Park, 2022).

This case exemplifies how gender bias embedded in policy and algorithmic systems can have life-threatening consequences. The program failed to account for the fact that men can also develop breast cancer, effectively barring male patients from life-saving care (Joshi, 2024). Advocacy groups, including the ACLU, condemned the exclusion as discriminatory and pushed for reform (Park, 2022). The case prompted broader conversations about the need to design healthcare systems and policies that are inclusive of all genders and reflective of real patient populations.

#### **Interview Questions**

1. When a young patient (let's say 25 years old) presents with a lump in their chest, what key factors guide your differential diagnosis and decision to recommend further testing?
2. How might your diagnostic approach change based on if the patient is female? What would your decision lean more towards in terms of diagnosis?
3. If they were male, would you consider the same things as women's chest pain or would your approach be quite different for male chest pain?
4. Breast cancer is rare in men, less than 1% of all breast cancers occur in men, should that rarity affect your decision to test for it? How do you balance the risk of over-testing versus missing a critical diagnosis?

Summarize the case.

5. What role do you think societal assumptions — like breast cancer being a “women’s disease” — played in the algorithm’s decision to deny coverage to Raymond?
6. What changes would you suggest to healthcare algorithms or policies to prevent gender-based denial of coverage in similar cases? How can we ensure equal access to treatment for all?
7. Should healthcare algorithms prioritize statistical likelihoods and trends or be designed to account for rare but serious conditions — especially when lives are at stake? How should they strike that balance?

## **Case 4: Anthony Randall and Kidney Transplant Algorithm Bias**

### **Case Summary**

Anthony Randall is a Black man from Los Angeles who was on dialysis, waiting for a kidney transplant for over five years (TheGrio, 2023). What he did not know is that an algorithm from the transplant system incorporated a race-based “modifier” that made Black patients’ kidney scores seem better than they were. This modifier caused Randall’s kidney disease to be less severe than it truly was, leading to his placement on the national transplant waiting list being significantly delayed (TheGrio Staff, 2023). In mid-2023, he filed a case against his hospital (Cedars Sinai Medical Center) and the United Network for Organ Sharing alleging that he was unfairly deprived of a fair chance to get the transplant because of the racially biased formula (Penn Medicine, 2024). It was no secret that the algorithm had a bias (Penn Medicine, 2024).

The board of the transplant system understood the modifier was resulting in Black patients' illnesses being severely underestimated. By early 2023, all hospitals were directed to stop the usage of race adjustment and Black patients' waiting times were to be changed to reflect the postponement (Penn Medicine, 2024). Randall claims that had these changes come sooner; he could have already had the kidney that he desperately needs. His case highlights how the goal of the clinical algorithm was good, but the execution was not due to the insertion of race which caused Black patients to not receive quality care in a timely manner.

### **Interview Questions**

1. Imagine you're a healthcare provider evaluating a patient with kidney disease for transplant eligibility. What clinical factors would you typically consider when assessing disease severity and readiness for transplant?
2. Have you heard of race-based modifiers in kidney function scoring (such as eGFR)? What do you think was the original intention behind including race in those calculations? Summarize the case.
3. Knowing now that this modifier made Black patients appear healthier than they were — delaying transplant eligibility — how do you think this affected patients like Anthony Randall?
4. Should hospitals or national organizations be held accountable when algorithms known to be biased continue to be used? What obligations do they have once they become aware of harm?
5. In Randall's case, the board eventually mandated the removal of race adjustments, but it came years after the issue was known. Do you think that delay was acceptable? What should have been done differently?
6. How can health systems design clinical algorithms that avoid reinforcing historical or systemic inequities — especially those tied to race?

7. What steps, if any, should be taken to make things right for patients who were harmed by biased algorithms — such as adjusting wait times, issuing apologies, or offering compensation?

### **Case 5: Dr. Noha Aboelata and Pulse Oximeter Bias During COVID-19**

#### **Case Summary**

During the pandemic, race-related biases in medical technology confronted Dr. Noha Aboelata, a family physician and the Chief Executive Officer of Roots Community Health Center based in Oakland. In late 2020, one of her patients was an elderly African American gentleman who suffered from chronic lung illness (UCSF, 2022). One of the checks done previously, the pulse oxygenation check, revealed that his oxygen saturation levels were high. Even though the device showed a relatively normal oxygen level, Dr. Aboelata's clinical instinct indicated the patient was much more distressed. She conducted an arterial blood gas test which confirmed her worst fears, the oxygen content in the patient's blood was too low and he needed oxygen.

Sometime later, she came across an article in the New England Journal of Medicine that confirmed her hunch; the oximeters were unable to register low oxygen levels in dark-skinned patients as compared to white patients, due to differences in how light is absorbed by the skin (Allen, 2024). She and her colleagues were outraged by a device that was supposed to help their patients but was grossly inaccurate for the Black population. Finally, her clinic participated in a class-action lawsuit against easier manufacturers and sellers of pulse oximeters for more detailed warnings and up-to-date devices (UCSF, 2022). She did not stand idle while demanding the FDA take pulse oximeter discrimination towards races very seriously.

#### **Interview Questions**

1. Imagine you're treating a patient with chronic lung disease during a respiratory pandemic. If a pulse oximeter shows normal oxygen saturation, but the patient appears visibly distressed, what would you do next?
2. How much would you completely rely on tools like pulse oximeters in clinical decision-making? Do you think you would ever question or double-check the accuracy of a medical device?

Summarize the case.

3. In this case, Dr. Aboelata's clinical judgment overruled the device's reading. What does this say about the limitations of relying too heavily on technology without understanding its biases?
4. Pulse oximeters were found to overestimate oxygen levels in patients with darker skin. Why do you think this design flaw persisted for so long, despite the risk it posed to patients of colour?
5. Do you believe the FDA and manufacturers have done enough to address this issue? What more should regulatory bodies be doing to ensure devices are accurate across diverse populations?

6. How should future healthcare providers be trained to detect and challenge device-based bias in patient care?
7. For communities that have been historically underserved or harmed by biased tools, how can the medical system begin to rebuild trust and ensure safer, more equitable care?

### **Case 6: Alex Morales and Smartwatch Blood-Oxygen Reading Bias**

#### **Case Summary**

Alex Morales, a New York resident, brought attention to a case of possible racial bias in the consumer health device, Apple Watch (Stempel, 2023). Morales, who has a darker complexion, bought an Apple Watch with the expectation that its blood oxygen sensor would accurately log his oxygen levels for fitness and health purposes. To his surprise, he later found out that the device's oximeter may not work well with people from his demographic (Stempel, 2023).

In late 2022, Morales initiated a class action lawsuit against Apple for allegedly containing a blood oxygen app that was racially discriminatory and did not function as promised for non-white customers. Supported in part by complaints of other studies claiming that more advanced pulse oximetry devices are “massively less useful on people with darker skin”, Morales asserted that Apple owed the public an explanation (Stempel, 2023; Nicholls, 2022). After all, paying smartphone users assumed that the device would be equal for all users, which is not the case. While the judge dismissed the case in 2023, it did start an important discussion regarding Apple products. Their case showed the world that there are, in fact, indirect biases in medical-grade equipment. This shows us that Alex Morales and the rest of the community are still subjected to discrimination based on race even in the technology they choose to use. Moreover, it demonstrates the keen eye for responsibility the tech industry has in these situations.

#### **Interview Questions**

1. If you purchase a health-monitoring device like a smartwatch, what level of accuracy and reliability would you expect — especially when it comes to tracking critical health metrics like blood oxygen levels?
2. How would you feel if you discovered that your device didn't work as accurately for people with your skin tone or demographic?

Summarize the case.

3. Do you think companies like Apple have an obligation to disclose limitations in their health sensors, especially if those limitations affect certain racial or skin tone groups?
4. Why do you think skin tone bias in oximeters and wearable health devices hasn't received widespread attention until recently? What might be some of the barriers to addressing it?
5. The court dismissed Morales's case, but do you think tech companies should still be held accountable in other ways? What kind of responsibility should they carry when their products are shown to underperform for marginalized users?

6. Should consumer tech devices with health features (like smartwatches) be subject to the same regulatory scrutiny as medical devices? Why or why not?
7. How can the tech industry ensure that innovations in health and wellness tools are tested fairly across diverse populations? Who should be involved in that process?
8. What does this case reveal about the intersection between technology, race, and access to accurate health information in the digital age? How can companies rebuild or maintain trust with users who may feel excluded or misled by biased product performance?

#### **Overall Final Reflecting Questions**

1. Have you ever encountered or learned about implicit bias in clinical settings or in your own work? How do you think it affects healthcare decisions?
2. How can we ensure that AI systems in healthcare reflect the diversity and complexity of human identity, including transgender and non-binary experiences?
3. If you were designing this AI tool, how would you balance giving statistically accurate information with not downplaying rare but deadly possibilities?
4. What kind of training or human oversight should go into building and testing these AI symptom checkers to avoid embedded gender bias?
5. What do you think it feels like to be a patient denied care not because of medical necessity, but due to how an algorithm classifies you? How can future healthcare professionals be trained to recognize and challenge biases like these in medical systems or software?
6. If you were on a team creating a new clinical algorithm, what safeguards would you recommend to ensure it doesn't harm any demographic group, even unintentionally?

That's all my questions for you for today. I really appreciate and thank you for taking the time to share your opinion on this topic of how Biased Clinical Research Data Affects Healthcare Practices. Do you have any questions for me, or any closing remarks you would like to make before we finish?

Sounds great, that's all for today, thank you again for your time, and have a great rest of your day!

## References

Chung, H., Park, C., Kang, W. S., & Lee, J. (2021). Gender bias in artificial intelligence: Severity prediction at an early stage of COVID-19. *Frontiers in Physiology*, 12, Article 778720. <https://doi.org/10.3389/fphys.2021.778720>

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Santuccione Chadha, A., & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digital Medicine*, 3, Article 81. <https://doi.org/10.1038/s41746-020-0288-5>

Compton, J. (2019, May 22). Trans dads tell doctors: “you can be a man and have a baby.” NBC News. <https://www.nbcnews.com/feature/nbc-out/trans-dads-tell-doctors-you-can-be-man-have-baby-n1006906>

Joshi, A. (2024). Big data and AI for gender equality in health: Bias is a big challenge. *Frontiers in Big Data*, 7, Article 1436019. <https://doi.org/10.3389/fdata.2024.1436019>

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10), 100347. <https://doi.org/10.1016/j.patter.2021.100347> accuray.com

Park, S. (2022, September 8). Breast cancer doesn't discriminate against men. American Civil Liberties Union. <https://www.aclu.org/news/womens-rights/breast-cancer-doesnt-discriminate-against-men>

Ring, T. (2019, May 16). A pregnant trans man's misdiagnosis leads to Stillbirth. Advocate. <https://www.advocate.com/transgender/2019/5/16/pregnant-trans-mans-misdiagnosis-leads-stillbirth>

Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: A sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ Health & Care Informatics*, 29(1), e100457. <https://doi.org/10.1136/bmjhci-2021-100457>

Trendall, S. (2024, July 29). Gender bias concerns raised over GP app. Public Technology. <https://www.publictechnology.net/2019/09/13/health-and-social-care/gender-bias-concerns-raised-over-gp-app/>

University College London. (2022, July 11). Gender bias revealed in AI tools screening for liver disease. UCL News. <https://www.ucl.ac.uk/news/2022/jul/gender-bias-revealed-ai-tools-screening-liver-diseases-ucl.ac.uk>

Baten, R. B. A. (2024, October). How are US hospitals adopting artificial intelligence? Early evidence from 2022. *Health Affairs Scholar*, 2(10), qxae123.  
<https://pubmed.ncbi.nlm.nih.gov/39403132/>

Nicholls, M. (2022, January 10). AI in skin cancer detection: Darker skin, inferior results? *Healthcare in Europe*.  
<https://healthcare-in-europe.com/en/news/ai-in-skin-cancer-detection-darker-skin-inferior-results.html>

Manke, K. (2019, October 24). Widely used health care prediction algorithm biased against Black people. *Berkeley News*.  
<https://news.berkeley.edu/2019/10/24/widely-used-health-care-prediction-algorithm-biased-against-black-people/>.

TheGrio Staff. (2023, April 11). Black man awaiting kidney transplant alleges racial bias. TheGrio.  
<https://www.washingtonpost.com/health/2023/04/10/lawsuit-unos-kidney-transplant-race-discrimination/>.

Allen, A. (2024, October 7). FDA's promised guidance on pulse oximeters unlikely to end decades of racial bias. *KFF Health News*.  
<https://kffhealthnews.org/news/article/pulse-oximeters-racial-bias-fda-rules/>.

UCSF (Department of Epidemiology & Biostatistics). (2022, September 29). Pulse oximeters don't work as well on darker skin, leading to flawed COVID care. *UCSF News (Epi/Biostatistics)*.  
<https://epibiostat.ucsf.edu/news/pulse-oximeters-dont-work-well-darker-skin-leading-flawed-covid-care>

Stempel, J. (2023, August 21). Lawsuit claiming Apple Watch sensor exhibits "racial bias" is dismissed. *Reuters*.  
<https://www.reuters.com/legal/lawsuit-claiming-apple-watch-sensor-exhibits-racial-bias-is-dismissed-2023-08-21>

Penn Medicine. (2024, March 25). Friend or foe: A closer look at the role of health care algorithms in racial and ethnic disparities. *Penn Medicine News Release*.  
<https://www.pennmedicine.org/news/news-releases/2024/march/the-role-of-health-care-algorithms-in-racial-bias-inequity>