Algorithmic Bias in Atrial Fibrillation Therapies

Heather M. Ross

Artificial intelligence (AI) using machine learning and deep learning approaches is often touted as the future promise of personalized medicine across a variety of clinical conditions. While the promise of better personalized health outcomes is appealing, AI also introduces the risk of algorithmic bias and long-term health threats based on AI algorithms that have been "trained" on non-representative data.

The classic example of algorithmic bias is often cited as the mid-twentieth century practice of intentional race-based redlining for home mortgages, creating long-lasting racial segregation and areas of concentrated poverty associated with lower educational attainment and poor health outcomes. The mid-twentieth century also provided a poignant example of unintentional algorithmic bias resulting in disastrous long-term health effects. In the 1950s, many children received radiation therapy to treat an enlarged thymus gland. However, the enlargement was based on "normal" measurements derived from indigent subjects who were malnourished and had abnormally small thymus glands, an error realized in subsequent decades. Therefore, the children who underwent thymus radiation actually had normal anatomy and received unnecessary radiation. Decades later, those children had a five-fold increase in risk of thyroid cancer.

These examples demonstrate that algorithmic bias has the potential to exert negative impacts both in the present and in the long-term future, particularly in the case when basic knowledge about pathophysiology and therapeutic effects is limited. In contemporary medicine, this type of uncertainty is present in the case of atrial fibrillation (AF), the most common heart rhythm disorder affecting more than 5 million Americans with an estimated cost of over 20 Billion USD annually.

Our knowledge of AF is fundamentally incomplete, evidenced by the ongoing clinical challenge associated with achieving heart rhythm control via pharmaceutical or procedural mechanisms. Close examination of the AF literature demonstrates a recursive pattern of knowledge development, with historical data often re-examined and reinterpreted through the lens of emerging technologies that allow scientists and clinicians to better understand the complex anatomical and physiological systems they seek to master. AI may be a critical tool to help scientists move toward that mastery.

However, scientists still fundamentally lack understanding of the ways that external factors impact the heart's physiologic system. The complex "system" extends well beyond the heart's structure to include comorbid physiological conditions, social, environmental, and behavioral determinants of health. Beyond that, there are significant "unknown unknowns" that must not be discounted.

The risk of algorithmic bias due to undersampling is a particular concern in the United States due to race, ethnic, and socioeconomic disparities in access to care in the absence of a universal healthcare system. To this end, the standard of practice for developing and evaluating AI algorithms for AF therapies should include training on datasets that include broad sociodemographic and comoborbidity characteristics. AI research and development guidelines, along with FDA evaluation practices, should employ mechanisms to avoid the risk of undersampling on characteristics that may be underrepresented in elite tertiary care settings that supply a significant proportion of the clinical research data.