Artificial Intelligence in Medical Diagnosis

Co-hosted by the National Academy of Medicine and the U.S. Government Accountability Office

❖

October 6, 2022 11:00 AM – 12:30 PM ET
Meeting Highlights

MEETING FOCUS: *This public webinar, hosted by the National Academy of Medicine (NAM) and the U.S. Government Accountability Office (GAO), explored the promise and issues associated with using artificial intelligence (AI) in medical diagnosis, building off a GAO technology assessment titled Artificial Intelligence in Health Care: Benefits and Challenges of Machine Learning Technologies for Medical Diagnostics and a NAM Perspectives Discussion Paper titled Meeting the Moment: Addressing Barriers and Facilitating Clinical Adoption of Artificial Intelligence in Medical Diagnosis, which is included as a part of the GAO report.*

GAO Technology Assessment Overview

Karen Howard, Director of Science, Technology Assessment, and Analytics (STAA) at the Government Accountability Office presented the report as the final one in a three-part series of reports on AI in health care initiated by congressional request. The scope of this report includes AI in medical diagnostics as it pertains to current and emerging machine learning (ML) technologies for key diseases (i.e., certain cancers, diabetic retinopathy, Alzheimer's disease, heart disease COVID-19). In meeting with experts and stakeholders, several potential benefits of ML in medical diagnostics were identified. For example, the ability to detect inconsistencies in images earlier than humans and allow non-specialists to quickly screen and flag at-risk patients, in addition to greater consistency in diagnosis compared to humans and increased access to care in various settings (e.g., in-home settings, smaller clinical settings). Challenges in their development and adoption, and policy options to address these challenges were also presented. For example, limited data on real-world performance of these tools might affect clinician adoption, but testing these tools in multiple settings with various patient demographics could be time-consuming and expensive for technology developers. Potential policy options to address this challenge could be creating incentives, guidance, or policies to encourage or require evaluation of ML diagnostic technologies across a wide range of conditions and demographics, which could help identify biases and enhance trust. However, the dollars involved in implementing this policy and the potential delay in movement to get to the patient are factors to consider. Other challenges include a lack of clarity around how these tools are developed and make decisions and whether they would increase clinical productivity. Ultimately, increased collaboration among developers, policymakers, providers, health systems, and regulators could encourage better integration and improve the creation of and access to ML-ready data, but there are other considerations of which policymakers need to be mindful.

NAM Discussion Paper Overview

Steven Lin, family medicine physician at Stanford University and current fellow at the National Academy of Medicine, presented the NAM piece, developed as a companion to the GAO report. This paper drills down on key factors related to the successful adoption and acceptance of AI-diagnostic decision support (AI-DDS) tools within inpatient and outpatient settings. The issues are organized into four domains that comprise a framework derived from the authors' understanding of human and system behaviors in response to introducing new technologies. The reason to use domain explores the alignment of incentives, market forces, and reimbursement policies that drive health care, considering that these tools require substantial financial investment for both deployment and maintenance and must address pressing clinical needs. Establishing the affordability of and a clear impetus for using novel AI-DDS tools will be important for ensuring clinical utility and cost-effectiveness. Alignment with the quintuple aim of health care could be a useful framework for evaluating the necessity and utility of AI-DDS tools. The means to use domain is about ensuring that clinical environments are properly equipped and have the resources (e.g., human capital) and infrastructure to support and sustain AI-DDS implementation. Investment and reimbursement mechanisms are needed to obtain and maintain systems that can support a robust AI pipeline. Method to use discusses the workflow considerations and training requirements to support clinicians. Action priorities here include increasing funding opportunities and resources to advance innovation and the

implementation of AI and integrating AI competencies in training opportunities and certification processes. The last domain, desire to use, discusses the importance of attending to how these tools can facilitate professional fulfillment among providers and promote trust. Action priorities within this domain include aligning AI technologies with work-life balance goals, but also, clarity around regulation and oversight of these tools from the FDA. Finally, being mindful of equity implications throughout the life cycle of these tools, and addressing past current and potential future equity issues are critical to preventing healthcare disparities.

Panel Discussion

Connie Lehman, Harvard University and Massachusetts General Hospital, expanded on three themes from the presentations, providing her perspective as a radiologist and scientist in the AI domain. The first theme Connie amplified was real-world performance post-market assessment. With traditional computer data detection in mammography in the 90s, there was initial excitement around AI, but the dramatic results promised in studies weren't evident in the clinic. A lesson learned was that while technological advancements can go far, human interaction with that technology is the only factor that can be tested carefully once implemented into clinical practice. Factors of implementation are crucial to explore, and there are best practices that can be leveraged in certain use cases to ensure integration, continuous evaluation, and accountability. The second theme was the importance of focused attention to equity, and the opportunity with AI to correct inequities in health care systems. Connie highlighted that, although it is often said that models not trained in diverse populations cannot be used in diverse populations, her lab found that they didn't necessarily need to train their model in the same setting as their intended use, but it certainly needed to be tested. Last, the NAM and GAO have set an example by bringing key multidisciplinary stakeholders together and encourage a culture around removing siloes and cross-sector alignment.

Barbara Evans, University of Florida, focused on two issues affecting the desire to use these tools – reimbursement and privacy concerns. Advanced diagnostic technologies have the potential to better target treatments, improve outcomes, and reduce treatment costs at this time. A value-based reimbursement framework that can grasp these benefits is needed. Limited data on real-world performance presents a road block – these technologies cannot enter the clinic until they are reimbursed, but evidence of clinical effectiveness and utility to support reimbursement cannot be gathered until they are used in the clinic. A mechanism where to move tools to the clinic in order to study them further and enhance the evidence base is needed. She also cautioned that current reimbursement frameworks have not improved equity – a key consideration driving acceptability and desire to use this technology. The shift to AI-enabled health care presents an opportunity to rethink equity and health rather than expediting the addition of new technology into an existing payment framework and replicating payment processes that have not achieved equity and disregarding the uninsured. For privacy concerns, it is important to consider that the use of data in AI-DDS is clinical, not research-based. Thus, the privacy framework is different. The law tends to increase the confidentiality duties of data users (clinicians and health care systems) and layer more restrictions on what can be done to the data to protect patients. However, AI-DDS brings new players into clinical healthcare encounters who are not subjected to existing confidentiality norms (e.g., IT professionals, software developers). Therefore, professional societies and accreditation bodies need to be involved to help think through new requirements, and laws should evolve to add these stakeholders and protect privacy.

Adewole "Ade" Adamson, University of Texas at Austin, provided an equity perspective from the field of dermatology. While a lack of diverse data hasn't precluded using some technology among a larger population, training algorithms to recognize certain diseases on different skin types requires being mindful about applying algorithms trained in non-diverse settings. Measuring real-world impact of the use of these algorithms is important to understand scalability. If an algorithm that presents systematic bias is used, there is a potential risk of causing harm. One situation in which AI-based technology could improve equity is in the prescription of opioids. Many studies have shown how opioids are disproportionately prescribed to different groups due to clinician biases on pain tolerance, and an algorithm related to these decisions could perhaps reduce bias.

Dalia Powers, Humana, explained that, from a payer perspective, advancements and the leverage of AI and ML will help improve provider experiences and care, lower costs, and ultimately improve health outcomes. For example, at the start of

the COVID-19 pandemic, Humana applied AI and, using the social determinants of health, identified members that might be impacted by the pandemic and ensure their well-being. Crucial next steps for payers include working with regulators and CMS, a U.S. federal agency and the nation's largest healthcare payer, to establish a structure for using this new technology, especially in these beginning phases of adoption for assistive AI. Dalia noted that AI models could also help remove unnecessary bureaucratic steps as their providers would rather spend time attending to patients than completing processes that could be automated.

Q&A

- **When asked how to guarantee that AI-based tools work effectively in diverse settings**, panelists discussed the importance of including key stakeholders, such as clinicians and patients, in deciding what questions are important, what areas of need AI could fill, and how AI could be successfully implemented. Once the technology is tested and evaluated within restricted environments and made available in clinical practice, the next steps would be to create an analytical framework to measure performance. A process that would allow continuous model training and improvement is needed, and when issues, such as biases, are detected, people can easily intervene. Also necessary is implementing the right mix of incentives. Rather than favoring medical conditions that tend to be well-insured, policymakers, granting agencies, and public funding could help emphasize the need to address ones that may not be covered. Beyond, incentives, creating soft and hard governance mechanisms could also be of use. For example, recent FDA guidance on increasing patient diversity in clinical trials could encourage funders and healthcare systems to sponsor studies aligned with this guidance.
- **In response to a question on what is being done to recruit/educate medical students on AI,** a point was made that there is already a lot of material involved in the medical school curriculum and that not all doctors will also need to be trained AI specialists. Care teams might benefit more from integrating IT specialists and software developers, and creating reimbursement mechanisms to pay new professionals could help create more value.
- **Participants shared real-world examples of AI in practice**. One example entailed using AI to assess breast density on a mammogram, which can greatly impact care decisions. AI dramatically reduced human variation in breast density measurement in settings such as community practice sites. With more accurate measurements, the impact on patient quality of life in communities was significant as it limited additional imaging and anxiety among women. Another example was the over diagnosis of cancers, an issue in which AI has the possibility of magnifying if patients are unnecessarily labeled with the disease when they do not have it. Having patient and family advisory councils to encourage their voice in the research and design of these health technologies is important, especially for those who are often marginalized (e.g., patients with disabilities).
- **In asking what makes datasets "AI-ready," and how they could be improved,** data was described as "the new oil" – a source of wealth that is not useful until it is refined. Data quality varies drastically depending on its source (e.g., EHRs), and data quality needs to be improved at the source to enhance AI and ML. However, biases can be inadvertently introduced if data is overly processed and not applicable in practice across populations.
- **Regarding liability issues and their impact on adoption,** people are not for liability, although they want more accountability with AI systems. Liability is how U.S. court systems assign accountability, but physicians tend to carry a majority of the liability in a world where AI has multiple contributors. Liability can be a way to move technology forward but also hold us back if the burden is not shared and physicians' concerns are not relieved.
- **One question asked was how oversight and patient safeguards need to change to enable the shift from locked AI systems to continuously learning models to realize the full promise of AI**. This goes back to how assessment can continue after post-market implementation – one cannot monitor what is not measured. Performance measures for AI technology need to be implemented to support continuous learning and evolution of AI models. Also, even if an AI model is locked, the patient population and disease of focus may change over time, so continuous monitoring is essential regardless.
- **There was a discussion on what to do when AI outputs recommendations contradict clinician intuition**. In situations where patients were asked for their opinion when there was a difference between the AI and clinicians recommendations, the patient has usually gone with the clinician's opinion, which speaks volumes to the importance of their role. Even with technology completing various tasks, it's still a person that communicates with the patient as certain processes are not conducive to automation like others.