David E. Newman-Toker, MD, PhD
Director, Armstrong Institute Center for Diagnostic Excellence
Director, Division of Neuro-Visual and Vestibular Disorders
Johns Hopkins Hospital, Pathology Building 2-221
600 N. Wolfe St., Baltimore, MD 21287-6921
OFFICE: 443-287-9593 | EMAIL: toker@jhu.edu
WEB: hopkinsmedicine.org/david-newman-toker

January 12, 2024

Emma Schultheis
Legislative Clerk, Committee on Energy and Commerce
2125 Rayburn House Office Building
Washington, DC 20515

RE: Newman-Toker Congressional Testimony Follow-up Questions for the Record

Dear Ms. Schultheis:

I greatly appreciated the opportunity to testify before the Subcommittee on Health on November 29, 2023, at the hearing entitled "Understanding How AI is Changing Health Care." My responses to the additional questions the Members submitted are enclosed.

Please do not hesitate to contact me if I can provide additional information.

Sincerely,

David E. Newman-Toker, MD, PhD
Director, Division of Neuro-Visual & Vestibular Disorders
Director, Armstrong Institute Center for Diagnostic Excellence
Past President, Society to Improve Diagnosis in Medicine (2018-2020)

Professor of Neurology, Ophthalmology, Otolaryngology & Emergency Medicine
David Robinson Professor of Vestibular Neurology
Johns Hopkins University School of Medicine

Professor of Epidemiology, Health Policy & Management
Johns Hopkins Bloomberg School of Public Health

The views expressed herein do not necessarily reflect the views of the Johns Hopkins University or Johns Hopkins Medicine.

Answers to Questions for the Record Following the
House Committee on Energy & Commerce
Subcommittee on Health
Hearing on November 29, 2023:
Understanding How AI is Changing Health Care

January 12, 2024

David E. Newman-Toker, MD, PhD
Johns Hopkins University School of Medicine

The Honorable Morgan Griffith

1. Are there any privacy implications involved with using AI learning machines to help diagnose patients? It is my understanding that AI technology is constantly learning by new data being added into the system so what exactly does this AI software do with the diagnoses it finds for a patient when it is sent it to a provider or doctor's office?

In general, AI and other machine learning systems in healthcare are designed to "remember" only the "rules" (called "models"), not the specific examples used for them to create these models. However, there are still certainly privacy implications associated with use of patient data as part of AI systems.

First, large patient data sets are needed to train AI systems. Often this work is done by private entities outside of healthcare that may not be directly bound by the HIPAA privacy rule or other data protection rules[1] or that may not have processes in place for adequate privacy-related security. This creates an added risk of data breaches in the creation of AI systems.

Second, AI systems can be built that will "re-identify" de-identified patient data. Some of this will be done very directly (e.g., from a photograph taken of a skin lesion on someone's face that is then run back through facial recognition software to re-identify the person). However, AI computer vision can already identify individuals from high-resolution imaging of body parts other than the face or fingerprints, since there are many details of human anatomy that are person-specific (e.g., dental

x-rays[2]). However, much of this re-identification will probably be done instead by aggregating data across multiple data sources, only some of which are identifiable (e.g., internet search history for symptoms and doctors, GPS tracking of location while driving to a clinic or hospital, plus medical records from a specific clinic or hospital from a specific date). This is already happening outside of healthcare. Standard digital industry processes are used to track user behaviors (e.g., internet "cookies").[3] These, in turn, may be used by "data brokers" to aggregate user data across digital platforms and public records (including date of birth and Social Security number) to build a more complete user picture than a user may have intended when granting consent to being tracked by cookies.[4] Healthcare data will begin to be added to such profiles using AI systems.

For feedback and improvement, AI systems that use reinforcement learning (rather than being trained "once" on a large, de-identified data set) need to have access to personal health information at more than a single point in time. For example, imagine an AI system that identifies a lesion on a brain scan as a brain cancer. That same AI system would need to be given information in follow-up to let it know whether this diagnosis was correct or not (e.g., a brain biopsy result). The AI system might also be designed to let it know whether the patient's treatment response or ultimate health outcome was good or not, so that future versions of this AI system could not only offer a diagnosis but also a recommended course of treatment and a prognostic estimate of a positive health outcome. For this, the AI system would need to know the patient's treatment and health state. All of these data points make it much easier for AI systems to then "link out" to more identifiable data from "cookies," etc. If the same system is directly responsible for notifying the patient or the patient's doctor about any AI-determined diagnoses, treatment recommendations, etc., then the system will likely have ready access to personal health information that is either identified or easily identifiable. Strong privacy controls would need to be in place to minimize risks of data breach or mitigate the resulting harms from such a breach on patient privacy. Similarly, protections must also be in place to prevent deliberate misuse of such information by healthcare providers or payers.

2. In your testimony, you mentioned how AI technologies use electronic health records to learn and create a uniform data set and how if the initial clinical diagnosis the AI system is learning on is faulty, it could lead to the same mistakes human make. How do you correct this and ensure it does not happen?

To my knowledge, there are only two ways to prevent AI systems from replicating the same mistakes we routinely make with clinical diagnosis in day-to-day medical practice. The first is to create large, iron-clad (often called "gold standard") data sets on which to train AI systems in the first place. The second is to use reinforcement learning (i.e., training the system in ongoing fashion using accurate feedback) to gradually eliminate the errors. The best results will likely come from adopting **both** strategies: gold-standard data sets to have AI systems start with "A" or "A-" performance (rather than "B" or "C" performance), then reinforcement learning to get an "A" or "A-" up to an "A+".

The best method to create gold-standard data sets is to do so as part of large diagnostic research projects that are well resourced and can create rigorous diagnostic results using unbiased methods. Unfortunately, there are structural impediments to federal funding for such studies, especially since NIH is organized by diseases (optimal for treatment-related research) rather than symptoms (optimal for diagnosis-related research) (*see Appendix: Response to Senate RFI for additional details*). Congress should allocate resources to diagnostic research initiatives to close this important gap.

The best method to facilitate high-quality reinforcement learning is to do so as part of clinical care processes by ensuring that health systems (a) capture information on correct final diagnoses after follow-up systematically for every patient, regardless of whether patients return to their institution for ongoing care and (b) assess their institutional diagnostic error rates and any associated patient harms from missed or delayed diagnoses (i.e., diagnostic errors). Unfortunately, there are currently no strong healthcare incentives for either of these or, for that matter, for getting the correct

diagnosis in the first place. Congress should seek regulatory requirements and payment incentives in healthcare that support (a) capturing information on correct final diagnoses after follow-up, (b) ascertaining patient health outcomes from diagnostic errors, and (c) getting the correct diagnosis in the first place.

3. Fraud, Waste, and Abuse costs hundreds of billions of dollars per year in health care. We currently do not have a good system in place to prevent this waste. Do you think this is one area that is ripe for AI to potentially be able to step in and provide a more robust system to analyze waste?

Because waste is generally considered non-criminal,[5] it is more common, more challenging to define in a real-world healthcare context, and, consequently, probably harder to detect than fraud or abuse. Waste is estimated to account for approximately 25% of annual healthcare spending in the US,[6] while outright fraud and abuse likely account for a smaller fraction estimated at about 3-10% of healthcare spending.[7]

Detection of fraud, abuse, or waste using AI systems is plausible and has already been studied,[8] but the constraints on accuracy are the same as with other AI functions—classification is only as accurate as the underlying data used to train systems. This means that waste, which is harder to accurately classify, will be harder to analyze in this way.

Preventing waste in healthcare is a complex problem because the causes of waste are multifactorial, and the system responds dynamically to changing conditions in ways that are sometimes undesirable. For example, if a change is made to lower reimbursements for a particular clinical service, physicians may respond by increasing volumes or other changes to maintain revenue despite the lower reimbursement rate.[9] There is one major form of healthcare waste, however, that should be readily reducible with the effective use of AI systems in healthcare: eliminating waste linked to errors in medical diagnosis.

Improving diagnostic accuracy in the US healthcare system is likely to result in significant waste reduction. This is because we currently overuse more diagnostic test resources (searching for dangerous diseases in those with benign causes) than we underuse (failing to search), so improving diagnostic practices will lead to both higher quality and lower costs, the latter with an emphasis on reducing wasteful overuse of diagnostic tests.[10] For example, implementing evidence-based bedside diagnosis of stroke in the emergency room for just one symptom (dizziness)[11] could potentially result in $1 billion in savings, while simultaneously leading to the elimination of over 50,000 missed strokes each year in the US.[12] If AI-based diagnostic decision support makes such innovations a reality, then such AI systems will likely dramatically reduce waste.

The Honorable Anna Eshoo

1. I introduced the *CREATE AI Act* (H.R. 5077) to give medical researchers in all sectors of society, including academia, small businesses, nonprofits and government agencies equitable access to resources needed to develop AI technology. a) Do you support this bill?

Yes, I support the *CREATE AI Act* (H.R. 5077). I believe it is crucially important for there to be public (as opposed to only private/commercial) resources related to AI development. According to materials about the CREATE AI Act: "The NAIRR will offer the following to researchers, educators, and students at higher education institutions, non-profits, and federally funded agencies: (1) Computational resources, including an open-source software environment and a programming interface providing structured access to AI models; (2) Data, including curated datasets of user interest and an AI data commons; (3) Educational tools and services, including educational materials, technical training, and user support; and (4) AI testbeds, including a catalog of open AI testbeds and a collaborative project with the National Institute of Standards and Technology." For the purposes of AI in healthcare, I believe that the first two of these are especially critical: (1) computational resources/open-source software environment; and (2) curated datasets. These would permit a degree

of independence from the commercial healthcare information technology sector, permitting an alternative avenue for AI/machine learning algorithm development and deployment that is free from the potential biasing influence of commercial goals. This could allow development of AI tools with a specific focus on optimizing healthcare quality (especially minimizing harms from diagnostic errors), healthcare efficiency (particularly reducing societal costs of healthcare), and healthcare equity.

The Honorable Nanette Barragán

1. The human component in the development and deployment of AI is important since humans can observe patterns or biases in AI-generated content. How do you recommend the medical technology workforce is diversified or trained in digital literacy so that AI algorithms do not reinforce racial biases and other biases?

Expert human oversight of AI algorithms is essential to catch errors and bias in such systems. For this oversight function to work, it is crucial that such humans be sufficiently skilled to be able to detect problems that may arise in AI systems. There is legitimate concern that overreliance on AI systems will result in "deskilling" of clinicians such that they may no longer be able to adequately oversee or "fact check" AI systems. There are ways to avoid clinical deskilling by AI systems (e.g., ensuring that clinicians render their own judgments before seeing AI system recommendations). However, it is unclear whether these sorts of approaches can be used successfully to avoid racial or other biases from creeping into clinical practice, since such biases already exist without AI systems.

A more diversified medical workforce (or one trained in bias concepts, especially in relation to how bias can be propagated into AI systems/algorithms) may be more attuned to racial bias, but it is unclear whether this will be sufficient to permit them to identify racial bias in AI systems. While training in bias and digital literacy are certainly reasonable measures, it is likely that other forms of oversight will be needed that capitalize on medical data science experts, rather than general clinicians. For example, there are ways to interrogate algorithms and analyze algorithm results using

"big data" that can identify untoward patterns suggestive of racial or other biases.[13] These approaches must call on those with methods expertise in AI, clinical research, and statistics. Mechanisms supporting oversight by teams with these skills must be established and fostered.

2. Due to existing systemic inequities in our healthcare system, there is a lack of robust medical data for Black, Latino, and Asian communities. I am concerned that the use of AI that relies on this data will not serve all communities in an equitable way. Will increased diversity in clinical trials help improve AI algorithms and ensure health disparities are not worsened?

Reliance on healthcare data that lack adequate representation from minority populations risks perpetuating inequities in our healthcare system by concretizing them as part of AI algorithms. To help reduce (rather than maintain or exacerbate) such demographic biases, it is essential to improve both the quality and quantity of healthcare data about minority populations. Increasing diversity in clinical trials will improve the **quality** of healthcare data about minority populations, at least for the subset of healthcare questions that are subjected to clinical trials. To a lesser extent, the **quantity** of minority data can also be increased in such trials by oversampling from minority populations (as opposed to proportional representation of minority groups, which is often the expectation).

However, there are two major reasons why addressing diversity in clinical trials alone is not likely to produce a major impact on demographic bias in AI systems. First, clinical trials will rarely have enough minority participants to support adequate AI system training. AI systems require thousands (and often tens of thousands) of relevant patients with complete data to be appropriately trained and validated. According to the FDA, phase 3 clinical trials typically have 300 to 3,000 participants.[14] Even if minority populations are oversampled, any given minority population is highly unlikely to make up more than 20% of the clinical trial population (i.e., a maximum of 60 to 600 participants). It is possible to aggregate data across clinical trials, but two sequential clinical trials answering the exact same question are rarely funded, unless the results of the initial trial are believed

to have been equivocal because the trial was too small (in which case even more minority research subjects will be needed to see clear demographic differences). Second, clinical trials evidence is never sought for the vast majority of decisions made in clinical practice. One study found that only 18% of recommended practices are supported by clinical trials evidence.[15] In addition, roughly half of medical care actually delivered on a day-to-day basis is discordant with clinical practice guideline recommendations.[16] In particular, diagnostic decisions are rarely studied in clinical trials where there are large demographic differences—minorities experience lower access to diagnostic testing and higher diagnostic error rates and resulting harms across a range of conditions.[17]

In summary, greater diversity in clinical trials is necessary but not sufficient for ensuring equitable healthcare delivered using AI systems. In addition, there should be a specific emphasis placed on ensuring equitable outcomes from AI delivery in healthcare via mechanisms such as developing large, gold-standard data sets (e.g., for diagnostic accuracy) and ongoing monitoring of patient outcomes. All of this will require combining expertise across AI, clinical research, and statistics.

References

1. Business Associates. Office for Civil Rights, US Department of Health and Human Services. 2003. Accessed January 10, 2024. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html
2. Heinrich A, Güttler FV, Schenkl S, Wagner R, Teichgräber UKM. Automatic human identification based on dental X-ray radiographs using computer vision. *Sci Rep*. 2020;10(1). doi:10.1038/S41598-020-60817-6
3. What Personal Data Do Companies Track? McAfee Blog. Accessed January 10, 2024. https://www.mcafee.com/blogs/tips-tricks/what-personal-data-do-companies-track/
4. How to Remove Personal Information From Data Broker Sites. McAfee Blog. Accessed January 10, 2024. https://www.mcafee.com/blogs/tips-tricks/how-to-remove-personal-information-from-data-broker-sites/
5. Fraud, Waste, and Abuse for Health Care Providers. Office of Inspector General, US Department of Health and Human Services. Accessed January 10, 2024. https://oig.hhs.gov/reports-and-publications/featured-topics/ihs/training/fraud-waste-and-abuse-for-health-care-providers/content/#/
6. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System: Estimated Costs and Potential for Savings. *JAMA*. 2019;322(15):1501-1509. doi:10.1001/JAMA.2019.13978
7. The Challenge of Health Care Fraud. National Health Care Anti-Fraud Association. Accessed January 10, 2024. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/
8. Kumaraswamy N, Markey MK, Ekin T, Barner JC, Rascati K. Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead. *Perspect Health Inf Manag*. 2022;19(1):1i.
9. Devlin AM, McCormack G. Physician responses to Medicare reimbursement rates. *J Health Econ*. 2023;92. doi:10.1016/J.JHEALECO.2023.102816
10. Newman-Toker DE, McDonald KM, Meltzer DO. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf*. 2013;22 Suppl 2(Suppl 2). doi:10.1136/BMJQS-2012-001616

11. Edlow JA, Carpenter C, Akhter M, et al. Guidelines for reasonable and appropriate care in the emergency department 3 (GRACE-3): Acute dizziness and vertigo in the emergency department. *Acad Emerg Med*. 2023;30(5):442-486. doi:10.1111/ACEM.14728

12. Newman-Toker DE. Missed stroke in acute vertigo and dizziness: It is time for action, not debate. *Ann Neurol*. 2016;79(1):27-31. doi:10.1002/ANA.24532

13. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/SCIENCE.AAX2342

14. Step 3: Clinical Research. US Food and Drug Administration. Accessed January 10, 2024. https://www.fda.gov/patients/drug-development-process/step-3-clinical-research

15. Ebell MH, Sokol R, Lee A, Simons C, Early J. How good is the evidence to support primary care practice? *Evid Based Med*. 2017;22(3):88-92. doi:10.1136/EBMED-2017-110704

16. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Promoting Adoption of Clinical Practice Guidelines. In: Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, eds. *Clinical Practice Guidelines We Can Trust*. National Academies Press (US); 2011. Accessed January 10, 2024. https://www.ncbi.nlm.nih.gov/books/NBK209543/

17. Newman-Toker DE, Peterson SM, Badihian S, et al. *Diagnostic Errors in the Emergency Department: A Systematic Review*. Agency for Healthcare Research and Quality (US); 2022. doi:10.23970/AHRQEPCCER258

SOCIETY to IMPROVE DIAGNOSIS in MEDICINE

October 27, 2023

The Honorable Bill Cassidy, MD
Ranking Member
Senate Committee on Health, Education, Labor, and Pensions (HELP)
428 Senate Dirksen Office Building, Washington, DC 20510

RE: Diagnostic Research Excellence Responses to Senator Cassidy RFI on NIH Modernization

Dear Senator Cassidy and HELP Committee:

Thank you for your willingness to consider modernizing aspects of NIH structures and processes to enhance research and achieve better health outcomes for all Americans. **Our responses to the request for information are thematically tied to the issue of diagnostic research** (as opposed to that focused on therapies). The introductory remarks below provide additional context for those responses.

A 19-year-old champion athlete whose stroke was missed in the emergency room… he can now only communicate by blinking; a 12-year-old girl whose fatal infection was mistaken for the flu and not caught in time; a new father whose malignant cancer pathology results were never communicated to the family, leading to his untimely death; a healthy 75-year-old woman who went blind from a rare but easily treatable disease that could have been diagnosed with an inexpensive blood test—these are among the thousands of patients who suffer serious, preventable harms from diagnostic errors.

The US National Academy of Medicine (NAM) (formerly Institute of Medicine [IOM]), in its 2015 report, *Improving Diagnosis in Health Care*, called diagnostic errors, defined as the failure to make accurate and timely medical diagnoses or communicate these to patients, a "blind spot" for healthcare.[1] The report articulated that diagnostic errors are failures of our healthcare delivery system, rather than individuals[1] and called improving the diagnostic process a "moral, professional, and public health imperative."[1] The NAM report emphasized that lack of public funding for diagnostic quality and safety research is a critical and significant barrier to improving diagnosis for patients; it called for dedicated funding for research to develop, refine and fully implement solutions designed to reduce harms from diagnostic error.

Diagnostic errors were recently estimated to account for roughly 800,000 permanent disabilities or deaths each year in the US.[2] Worse still, the burden of these serious harms from diagnostic error falls unfairly on historically disadvantaged or otherwise marginalized populations—women and minorities are approximately 20-30% more likely to suffer diagnostic errors than their white male counterparts, and both the very young and the very old are at disproportionate risk relative to their numbers.[3,4]

It has been estimated that the societal costs of diagnostic error amount to more than $100 billion annually.[5] Despite this, federal funding for diagnostic research lags well behind funding for therapeutic research. Only about 0.5% of federal funding is diagnosis-focused, and roughly 0.02% of federally sponsored medical research funding focuses on eliminating diagnostic error and harm.[6]

Serious harm from diagnostic error is the most underfunded public health crisis in America today. Although research funding focused on diagnostic errors has increased more than three-fold since 2016, it remains grossly inadequate to its public health impact. Estimated 2024 categorical spending for HIV/AIDS (~8,219 annual deaths) is $3.3 billion, while funding for smallpox, eradicated in 1980 (0 annual deaths worldwide), is $39 million.[7] By comparison, funding to improve diagnostic excellence and reduce diagnostic errors (~370,000 US deaths[2]) currently amounts to $20-30 million, mostly funded by Agency for Healthcare Research and Quality (AHRQ).[8] This translates to over $400,000/year per death for HIV/AIDS, over $39M/year per death for smallpox, and $50 to $80/year per death for diagnostic errors.

NIH does conduct diagnostic research, but relatively little of this research is deployed to address common, ongoing pitfalls in medical diagnosis that lead to serious patient harms on a daily basis. The National Institute of Biomedical Imaging and Bioengineering focuses mainly on fundamental advances in imaging techniques. The NIH Undiagnosed Diseases Program,[9] part of the Undiagnosed Diseases Network, focuses primarily on identifying rare genetic diseases. Individual institutes and some newer programs (e.g., Pandemic Preparedness) fund research into disease-specific laboratory diagnostic tests for patients suspected of having specific infections, cancers, or other disorders. **However, no NIH institutes or programs routinely fund diagnostic research focused on improving bedside diagnosis of common medical symptoms such as fever, abdominal pain, headaches, back pain, or dizziness—symptoms that may be caused by dozens of different diseases (cutting across NIH disease-specific institute mission lines) and where missed, delayed, or wrong diagnoses most commonly occur**.

Although there are several factors contributing to this gap, **this is primarily due to a fundamental structural flaw in NIH organization: the organ-system orientation of the NIH institutes**. Organ system institutes are optimally aligned with a disease focus and development of new disease-specific therapies. However, they are fundamentally misaligned with the general task of medical diagnosis, whose object is to differentiate disease A from diseases B, C, D, or E among patients with symptoms X, Y, or Z. This is reflected ultimately in an extreme paucity of clinical, symptom-oriented diagnostic research.

In the pages that follow, we offer answers to some of Sen. Cassidy's questions about NIH modernization. **We believe these suggestions have the most potential for enhancing diagnostic research in pursuit of diagnostic excellence, thereby supporting the greatest possible health benefit for every American.**

The Armstrong Institute Center for Diagnostic Excellence and the Society to Improve Diagnosis in Medicine appreciate the opportunity to respond to the RFI.

Sincerely,

David E. Newman-Toker, MD PhD
David Robinson Professor of Vestibular Neurology
Director, Armstrong Institute Center for Diagnostic Excellence
The Johns Hopkins University School of Medicine
Former President, Society to Improve Diagnosis in Medicine (2018-2020)

Jennie Ward-Robinson, PhD
Chief Executive Officer
Society to Improve Diagnosis in Medicine

The views expressed herein do not necessarily reflect the views of the Johns Hopkins University or Johns Hopkins Medicine.

Increasing the Pace of Science

Overarching Questions

3. In your view, what would be the appropriate balance between basic, translational, and clinical research at NIH? How can NIH continue to prioritize truly fundamental research while improving outcomes for translational and clinical projects?

ANSWER: While fundamental research remains crucially important, the current NIH research portfolio is too heavily weighted toward basic research, which only rarely produces a transformational impact on patient health (and, when it does, usually only after decades of work). NIH should more strongly factor into consideration the anticipated short-term public health impact of medical research (rather than just its long-term effect). This includes health outcomes for patients as well as system outcomes such as cost or value. Invariably, later stage research has a far greater potential to directly impact outcomes quickly.

For diagnostic research, it is critical to fund many more translational and clinical projects via NIH programs, including the individual NIH Institutes and Centers, as well as the NIH Common Fund. **AHRQ plays a crucial role in developing and disseminating systems approaches to improving diagnosis, but its budget is far too small to address the full palette of diagnostic research, and many diagnostic innovations that are "early-stage" or "discovery"-oriented do not fit well within AHRQ's portfolio.** The relative paucity of diagnostic research at NIH leads to an inadequate downstream research pipeline for other agencies, focused heavily on implementation research, to ultimately impact health (**Figure 1**).

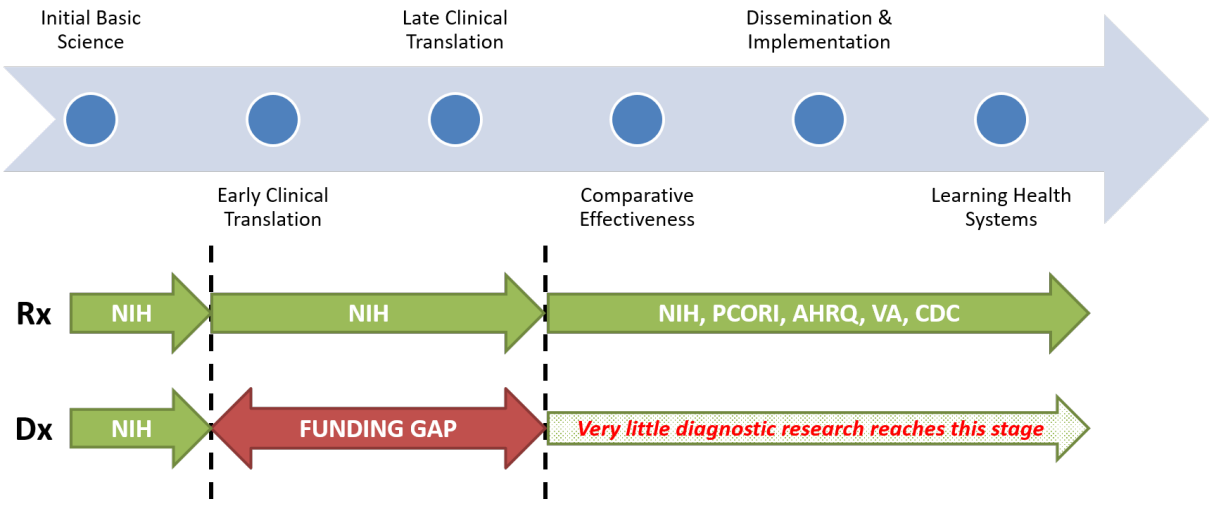

Figure 1. A gap in NIH translational diagnostic research leads to an anemic pipeline for subsequent implementation research. Abbreviations: Dx, diagnostic research; Rx, treatment research.

Extramural Research Program

4. Do you see opportunities to improve the current process for structuring peer review committees? What attributes does NIH tend to prioritize when selecting both chartered and ad hoc reviewers?

6. Could peer review committees be organized differently to improve NIH's evaluation of interdisciplinary work?

ANSWER: There are no peer review committees whose primary focus is diagnostic research, and most committees at the Center for Scientific Review or within individual NIH institutes lack members with expertise in diagnostic research and associated methods. This, in turn, is another structural impediment to funding diagnostic research at NIH. Diagnostic research (especially that focused on diagnosis of undifferentiated clinical symptoms in real-world practice settings) is usually considerably more complex

than therapeutic research. Applications for diagnostic research projects cannot be fairly assessed by those lacking the expertise to do so. As a result, these diagnostic projects tend to fare poorly in study sections, substantially reducing the odds of successful funding. This should be resolved by (a) ensuring that all grant review sections have at least 1-2 members who are diagnostic researchers and (b) constructing a diagnosis-specific study section to which diagnostic proposals can be directed.

The disproportionate presence of reviewers or program officers without MDs is also an issue. This tends to result in a primary focus on basic research and a relative lack of appreciation of the critical need for innovative clinical research. Although most PhDs have a solid foundation for understanding the importance of developing disease-specific treatments or cures, non-MDs sometimes struggle to understand the primacy of the diagnostic process—correct diagnosis is a prerequisite for correct care.

Organizing NIH for Success

Statutory Structure and Functions

1. Does NIH's current organ- and disease-based structure effectively facilitate the conduct of research? If yes, how? If no, what alternative structure would be more effective in your view? What barriers prevent Congress or the administration from implementing this structure, aside from NIH's statutory authorization and appropriations?

ANSWER: NIH's organ- and disease-based structure is superb for therapeutic research for specific diseases but generally antithetical to most diagnostic research (especially symptom-oriented clinical or translational diagnostic research). The general task of medical diagnosis is to differentiate disease A from diseases B, C, D, or E among patients with symptoms X, Y, or Z. At a disease-oriented institute, diagnostic tests intended to confirm that disease may be "within the institute's mission" but diagnostic research focused on bedside processes, tests, or technologies seeking to differentiate a "within mission" disease from ones "owned" by another institute are rarely considered mission-aligned.

Take, for example, impactful, practice-changing diagnostic research among patients with dizziness and vertigo using eye movements to diagnose stroke more accurately than our current "gold standard" (MRI of the brain).[10-12] Dizziness and vertigo affect tens of millions of Americans annually, leading to 18 million healthcare visits each year, including 5 million to the emergency department. The main clinical goal in diagnosing patients experiencing dizziness is to differentiate common, self-limited inner ear diseases from uncommon, dangerous strokes. This can be done by experts through careful bedside examination of eye movements, and may be done in the future by technologies such as "stroke goggles" or mobile phones that provide access to such expertise (https://giving.jhu.edu/story/dizzy-stroke-research/).

Unfortunately, ear diseases "belong" to the National Institute on Deafness and Other Communication Disorders (NIDCD), while strokes "belong" to the National Institute of Neurological Disorders and Stroke (NINDS), making it hard for either one to prioritize the research. When emergency department-based research focused on technology development as a solution to this important problem was proposed to NINDS, it was rejected on the grounds that most of the patients under study would have ear diseases, not stroke—NINDS suggested NIDCD. When the same research was proposed to NIDCD, it was initially rejected because the real emergency is to diagnose stroke, rather than inner ear diseases, which are not life-threatening—NIDCD suggested NINDS. In other words, the disease-specific orientation of the two institutes "trumped" the public health impact of the problem. In this way, most symptom-oriented diagnostic research, no matter how important from a public health perspective, falls between the cracks of different institutes' missions, never aligning fully with any single disease-specific mission. This is reflected ultimately in an extreme paucity of clinical, symptom-oriented diagnostic research.

This would be best addressed by creating a new institute devoted to diagnostic research (e.g., **National Institute for Diagnostic Research**). Short of this, a well-resourced Common Fund program devoted to

cross-disciplinary diagnostic research could be a reasonable alternative. Another option might be a program that required each institute to commit a meaningful percentage of their funding to diagnostic research (e.g., a minimum of 10%). However, the latter two options would require more oversight to ensure the resources were being deployed to serve the public health need around improving diagnosis through innovation and translation. Barriers are mainly sociopolitical (e.g., Congressional mandate for no new institutes; competition from other groups that wish to pursue new institutes but have previously been denied; or entrenched, historical interests within institutes that do not wish to change).

2. How might NIH's mission, strategic goals, and objectives be refined to better reflect and enable its core functions?

ANSWER: NIH's stated mission is: "NIH is the steward of medical and behavioral research for the Nation. Its mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability." The words "fundamental knowledge" (or similar words such as "discovery research") are code words for "basic science." This framing tends to exclude clinical diagnostic research and should ideally change. Eliminating the word "fundamental" would help, but even better would be to also add the words "improve diagnosis," "increase diagnostic accuracy," or "pursue diagnostic excellence" immediately before the words "enhance health" (e.g., "…application of that knowledge to pursue diagnostic excellence, enhance health, lengthen life, and reduce illness and disability."

On the About the NIH web page (https://www.nih.gov/about-nih/what-we-do/nih-almanac/about-nih) the following is the only place where diagnosis or diagnostic research is mentioned: "In realizing these goals, the NIH provides leadership and direction to programs designed to improve the health of the Nation by conducting and supporting research: in the causes, diagnosis, prevention, and cure of human diseases…" (*underline added for emphasis*). Note that diagnosis is only mentioned in a disease-oriented framework (which is reflected in the way that most diagnostic research is supported today). It is crucial that a statement about symptom-oriented diagnostic research be added to the NIH mission, goals, or associated explanatory text. For example, adding a bullet in this section that has words such as, "in symptom-oriented diagnostic research to optimize medical diagnosis."

3. In your view, could NIH research dollars be better allocated within the agency's portfolio? Are there certain areas of research that are over-funded or under-funded? What strategy should Congress and NIH take in allocating resources to specific areas?

ANSWER: There is little doubt that addressing failures in medical diagnosis (through innovation, translation, and dissemination) is the most underfunded public health crisis in America today. Correct diagnosis is a prerequisite for correct treatment of disease. Despite leading to death or permanent disability for an estimated ~800,000 Americans annually[2] (more than the estimated 608,570 who die of cancer annually in the US[13]), addressing diagnostic error receives just $20-30 million in research funding each year—just 0.06% of the annual research budget (**Figure 2**). Cancer receives $8.8 billion.

AHRQ plays a crucial role in advancing health services and safety research to improve diagnosis, but this problem cannot simply be the purview of AHRQ, as much of the work to be done involves what is generally considered translational or clinical research. This means that NIH must play its critical role.

Congress and NIH should adopt a public health-oriented stance in terms of return on investment. Using a simple metric such as "dollars per death" or "dollars per serious harm" (with serious harms representing either death or permanent disability[2]) would be one way to gauge (or set) the appropriate allocation of resources from a public health impact perspective. Estimated fiscal year 2024 (FY24) categorical spending for HIV/AIDS (~8,219 annual deaths) is $3.3 billion while funding for smallpox,

eradicated in 1980 (0 annual deaths), is $39 million.[7] By comparison, funding to improve diagnostic excellence and reduce diagnostic errors (~370,000 deaths[2]) currently amounts to $20-30 million, mostly funded by Agency for Healthcare Research and Quality (AHRQ).[8] This translates to over $400,000/year per death for HIV/AIDS, over $39M/year per death for smallpox, and $50 to $80/year per death for diagnostic errors. Research spending per serious harm (i.e., death *or* permanent disability, defined as equivalent or worse to losing a limb) from diagnostic errors is even lower at $25 to $40/year.

Congress should act to ensure dedicated diagnostic research funding that proportionally matches the public health impact of serious misdiagnosis-related harms, targeting at least $1 billion per year for 10 years across federal agencies. Eliminating harms from medical misdiagnosis of just 15 well-known, dangerous diseases could cut the total number of permanent disabilities and deaths by 50%.[2] Sustained funding in the hundreds of millions or billions will be required to effect critical changes in diagnosis. Funding might partly come from an institute-level mandate to allocate no less than 10% to diagnosis-related research. Currently this is just 0.5% (**Figure 2**), which is far too low for an aspect of clinical care that represents roughly half of what we do in medicine—i.e., we diagnose and treat medical illnesses.

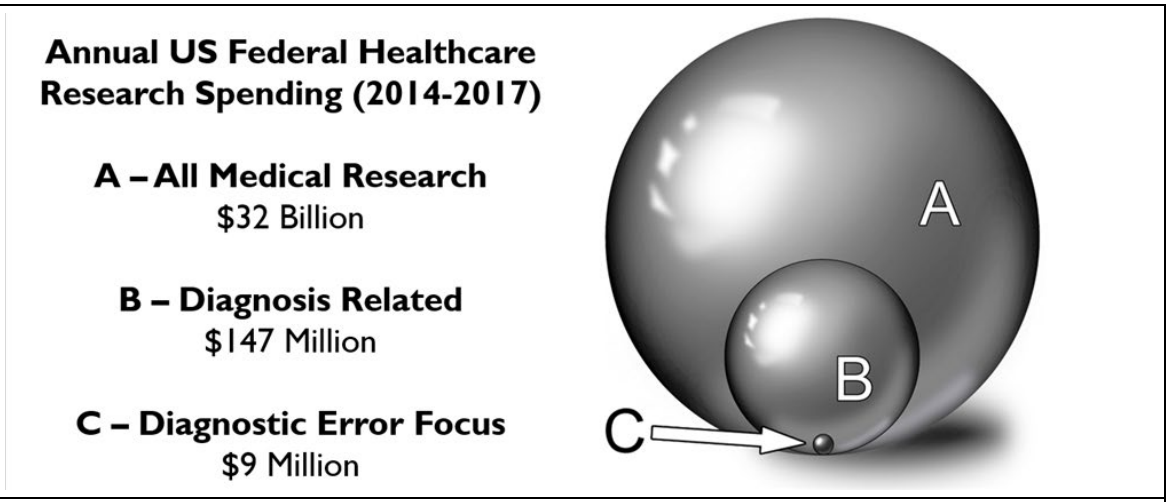

Figure 2. Distribution of funding for all medical research and the subset that is diagnosis-related or specifically focused on diagnostic error. The illustration shows spheres that are approximately proportional in size to the level of funding provided. While this illustration is based on federal spending in fiscal year (FY) 2014-2017, the overall picture has not changed appreciably since then. All medical research for FY 2024 is estimated at more than $50 billion. Although diagnostic error-related research has increased roughly three-fold to $20-30 million, it is still only ~0.06% of the overall research funding portfolio and receives a smaller share than we spend on smallpox research ($39M per year),[7] a disease eradicated nearly a half century ago. With ~800,000 Americans permanently disabled or dying as a result of failures in medical diagnosis,[2] this is an urgent public health crisis, and clearly the most underfunded ($50 to $80 per death) public health problem we face in America today. Some diseases, such as HIV/AIDS, receive over $400,000 per death for 45-fold fewer deaths.[2,7]

6. Could the process for selecting Common Fund projects be changed to improve the impact of Common Fund dollars and more easily integrate, and eventually transition, Common Fund projects into IC-level programs?

ANSWER: Congress should create a new **National Institute for Diagnostic Research**. However, if a new institute devoted to diagnostic research is not a realistic consideration, then the Common Fund is the next logical place to house a large diagnosis-focused program (e.g., $1 billion per year for 10 years).

Current cycle times from proposal submission to actual funding initiation is no less than a year and typically two or more years (since most proposals must be resubmitted at least once to be funded). In

general, the process of proposal selection (and especially for high-risk, high-reward programs) should be faster and more interactive. Applications should be staged, with a short initial application focused mainly on significance/public health impact of solving the problem, and the basic contours of the proposed research. In a subsequent stage, finalist candidates should be given an opportunity to present directly to the selection committee, where they can answer questions on the spot, rather than taking another year to reapply after getting written feedback from the selection committee. If questions are addressed adequately, a post-review update to the proposal could be made, if necessary.

For successful diagnostic research programs, the Common Fund could then transition them to a single institute or center where the diseases are fully aligned with the institute's mission or, alternatively, broker a deal with multiple institutes or centers if the diagnostic issues do not align well with just one (e.g., stroke diagnosis in patients with dizziness or vertigo might be co-funded by NINDS and NIDCD).

8. How, if at all, should NIH's high-risk, high-reward research portfolio be adjusted now that the Advanced Research Projects Agency for Health (ARPA-H) has launched? What high-risk research should be retained at NIH, and what types of projects are a better fit for ARPA-H?

ANSWER: For diagnostic research, high-risk, high-reward projects that lack clear industry ties or commercial incentives (such as high-value intellectual property, devices, or algorithms) should be retained by NIH. There are numerous "orphaned" diagnostic research projects that may not be tied to existing industry and offer little commercial incentive, yet still are both high risk and high reward.

For example, strokes and transient ischemic attacks affect over a million Americans each year and are among the leading causes of death and long-term disability in the US.[14] The cost to the US healthcare system is estimated at more than $56 billion per year.[14] Disability prevention requires rapid stroke treatment (within hours), which, in turn, requires rapid stroke diagnosis, yet we miss over 17% of strokes at initial medical contact[2] (which is more than 10-fold more than we miss heart attacks). The global burden of stroke is massive—in 2019, there were 6.6 million stroke deaths and 143 million lost years of full health from disabling strokes.[15] Mobile phone-based stroke diagnosis could be truly transformational worldwide,[16] yet it is unlikely to lead to massive commercial success.

This is also a more general problem for diagnostic research. In therapeutic research (particularly drug treatments), there is a clear and established pathway for industry-sponsored clinical trials. For FDA approval, pharmaceutical companies are required to demonstrate that their product is both safe and *effective* (meaning that it offers some tangible health benefit to patients). However, for diagnosis, the FDA standard is that the test or device is safe and *measures accurately what it purports to measure*. In other words, there is no requirement that diagnostic test manufacturers demonstrate that their products actually yield a health benefit. This means that they do not need to sponsor large, definitive (phase 3) clinical trials prior to marketing their product. As a result, there are very few diagnostic tests that have ever been subjected to rigorous testing of their value to human health.[17] There is little market incentive for a diagnostic test manufacturer to invest millions in a large-scale clinical trial that could (and likely often would) undermine the perceived utility of their approved device or test. NIH must be the main funder of such research, whether via usual clinical trials or high-risk, high-reward mechanisms.

Nevertheless, it is crucial that ARPA-H also play a role. As far as we are aware, none of the ARPA-H funded projects or programs is focused on diagnostic innovation or transformation. This should be rectified by recruiting a dedicated Program Manager (https://arpa-h.gov/careers/program-managers/) to lead a **Diagnostic Transformation Program** that will create the new future of diagnostic medicine.

Administrative Opportunities and Challenges

1. Regarding NIH's interagency collaborations, what currently works well and what could be improved? How can NIH better leverage capabilities that exist within the interagency, particularly for technologies and disciplines outside NIH's traditional scope?

4. What opportunities exist to harmonize funding applications for research awards across ICs and the interagency?

ANSWER: Diagnostic research is ripe for both cross-institute/center and interagency collaboration.

By Congressional mandate, AHRQ currently leads a Federal Interagency Workgroup on Improving Diagnostic Safety and Quality in Healthcare.[18] However, this program lacks dedicated funding and has very limited representation from NIH institutes.[18] As a result, its outputs have been limited.

From 2015-2017, the Office of Emergency Care Research (OECR) funded roughly a dozen multi-disciplinary studies related to diagnosis in emergency care settings (e.g., emergency departments).[19] Some were symptom-oriented diagnostic research studies that are very difficult to get funded in the current NIH structure. However, the visibility of this office is limited, both because it is housed in NINDS (rather than the Office of the Director) and because it has received extremely limited funding.

A large Common Fund investment (with matching funds required from relevant institutes/centers) could issue diagnostic research requests for applications (RFAs) and coordinate cross-institute/center collaborations around diagnostic research efforts. For example, the Common Fund could broker a deal with multiple institutes or centers if the diagnostic issues do not align well with just one (e.g., stroke diagnosis research in patients with dizziness or vertigo might be co-funded by NINDS and NIDCD).

Improving Transparency and Oversight

1. What specific policy recommendations do you have to improve the transparency of NIH's work, including its accountability to the American people and Congress? Are you aware of any specific mechanisms that have effectively achieved this goal for other federal agencies, including outside of the Department of Health and Human Services (HHS)?

ANSWER: Transparency for diagnosis-focused research funding is low. A first crucial step in increasing this transparency is for a category of "Diagnostic Research" to be added to the NIH list of categorical spending at the NIH RePORT website (https://report.nih.gov/funding/categorical-spending#/). The category should then be divided into subcategories (in the way "cancer" is both recorded by specific cancer and rolled up for all cancer research taken together). The subcategories should include "basic" diagnostic research (development of novel biomarkers or molecular tests, new imaging modalities, etc.) and "applied" diagnostic research (test accuracy, testing impact on management, testing impact on health outcomes, etc.). Ideally, all other disease-specific research categories would also report the diagnostic component subset separated out and accounted for (e.g., for "cancer" research, how much of the $8.8 billion estimated for fiscal year 2024 will be for diagnostic research?). On the website's table (https://report.nih.gov/funding/categorical-spending#/) the first request above is for the addition of several rows to account for diagnostic research spending; the second is for the addition of a column.

5. How, if at all, should the Office of the Inspector General for the Department of Health and Human Services' oversight of NIH be enhanced?

6. Congress established the Scientific Management Review Board to advise on NIH's structure and management, and the Research Policy Board to advise the Office of Management and Budget (OMB) and review administrative requirements for extramural research. However, to date, the Executive Branch has not effectively leveraged these entities. What steps could Congress take to encourage full

implementation of these statutory requirements, and how could Congress maximize the boards' ability to increase transparency and provide outside recommendations to NIH and OMB?

ANSWER: An incorrect diagnosis is almost guaranteed to produce incorrect treatment or failed management, so correct diagnosis is generally a "first step" prerequisite for providing optimized, disease-specific treatment and care. Despite this, there is an entrenched structural and cultural bias at NIH against funding diagnosis-related research. This is evidenced by the fact that <0.5% of all federal research spending is devoted to a process that represents roughly half of the medical endeavor—i.e., to diagnose and treat disease in order to alleviate suffering and prolong high-quality life.

Oversight is absolutely crucial to ensure that these entrenched tendencies adapt and change in the face of Congressional mandates to realign resource allocations with public health priorities. Using a simple metric such as "dollars per death" or "dollars per serious harm" (with serious harms representing either death or permanent disability[2]) would be one way to gauge (or set) the appropriate allocation of resources from a public health impact perspective. More complex calculations such as "dollars per quality-adjusted life year gained" (or "dollars per disability-adjusted life year averted") or even "return on investment" might be calculated by an agency such as the OMB, which may be well positioned to conduct the economic analyses required, including accounting for the financial burden of different disease groupings or medical problems. While the OMB might serve an "accounting" function, the Office of the Inspector General (OIG) might serve an "accountability" function. The OIG could use data generated by the OMB combined with other information about NIH structures and processes, noting the extent to which NIH practices have been appropriately modified to most effectively address diagnostic research (e.g., whether peer review groups have appropriate diagnostic research expertise; if NIH institutes/centers are devoting at least 10% of their funding to diagnostic research; and ensuring interagency workgroups are effectively coordinating cross-agency efforts).

Given the massive toll of diagnostic failures on Americans in the form of both health losses (~800,000 permanently disabled or dying each year[2]) and financial costs (societal cost of over $100 billion[5]), it is essential that the OIG be able to hold NIH accountable to Congressional statutes and appropriations as well as the American people. The overall goal is the successful development, conduct, translation, and dissemination of diagnostic research findings as part of a broader "pipeline" that ultimately improves health outcomes and decreases waste via more accurate and efficient diagnostic methods.

Other Issues

1. What other policies or issues should Congress consider, aside from those mentioned above?

ANSWER: There are three issues that fall outside of NIH but directly impact diagnostic research.

The first relates to funding for AHRQ. AHRQ plays a crucial role in the field of Diagnostic Excellence by working with stakeholders to develop, test, and move evidence-based tools, processes, and guidelines that enhance diagnostic safety into clinical practice. This includes work to improve "systems" reliability in the diagnostic process. We are thankful that Congress has begun to recognize the importance of this work with annual funding increases. The next essential step is a fully authorized and robustly funded program-level commitment in Diagnostic Safety and Quality at AHRQ, at a level at least on par with the Healthcare Associated Infections (HAI) program. In addition to its health services research focus, part of the mandate for such a program would be to rapidly accelerate dissemination and implementation of novel diagnostic technologies, discovered via NIH-funded diagnostic research, into doctors' offices.

The second relates to data availability for diagnostic researchers. Diagnostic research to measure diagnostic accuracy or error (or develop real-time surveillance by artificial intelligence methods) requires connecting the dots between initial symptoms and underlying diseases.[20] Currently, the

tracking of medical symptoms in relation to current diagnoses is inadequate to the task. Relatively simple policy steps could be taken to address this deficiency. The first (and likely most impactful) would be to have CMS mandate that providers and provider organizations report both symptom and disease when submitting clinical billing requests. The ability to do this already exists (coding systems for symptoms, spaces on the billing forms, etc.), but instead of initial symptoms carrying forward in the billing record once diagnoses are determined (or later revised), these symptoms are discarded (by CMS billing rule) in favor of a more "specific" diagnosis. This hinders progress in diagnostic research. These problems (i.e., a lack of symptom data) are propagated into other regional or national datasets, including those within HCUP (SID [state inpatient] and SEDD [state emergency department]). Similarly, large clinical data research networks (e.g., PCORnet) do not include symptom data in their data models. Finally, Medicare or other actors have the potential to provide researchers with anonymized datasets that support tracking patients across institutions or state lines, but many times key details are deliberately obscured (e.g., symptoms or institutional attributes such as location). Addressing these sorts of administrative barriers with policy changes could make an enormous difference to supporting effective diagnostic research and routine operational tracking of its impact on diagnostic outcomes.

The third relates to regulatory oversight for new diagnostic tests or technologies. The FDA offers a very clear set of guidelines, processes, and pathways for drug development, including the essential steps required (https://cacmap.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process). A similar sort of guidance is offered for device development (which, loosely speaking, encompasses many diagnostic tests; https://www.fda.gov/patients/learn-about-drug-and-device-approvals/device-development-process), but this guidance is generic to both diagnostic and treatment devices, and some diagnostic tests do not readily fit into the "device" category. The result is non-specific regulatory advice and a lack of clarity regarding the pathways for developing new diagnostic technologies. Establishing and formalizing a set of guidelines specific to diagnostic tests (laboratory or device-based), technologies (e.g., machine learning diagnostic algorithms), and other novel processes (e.g., diagnostic pathways or electronic health record-based decision support tools) would go a long way to clarifying what the "pipeline" for diagnostic research should look like. One example of a structural way to visualize the development of novel diagnostic techniques is the hierarchical diagnostic test assessment model shown below (**Figure 3**). Creating clear steps and guidance would also serve the function of identifying where funding gaps exist in the pipeline.

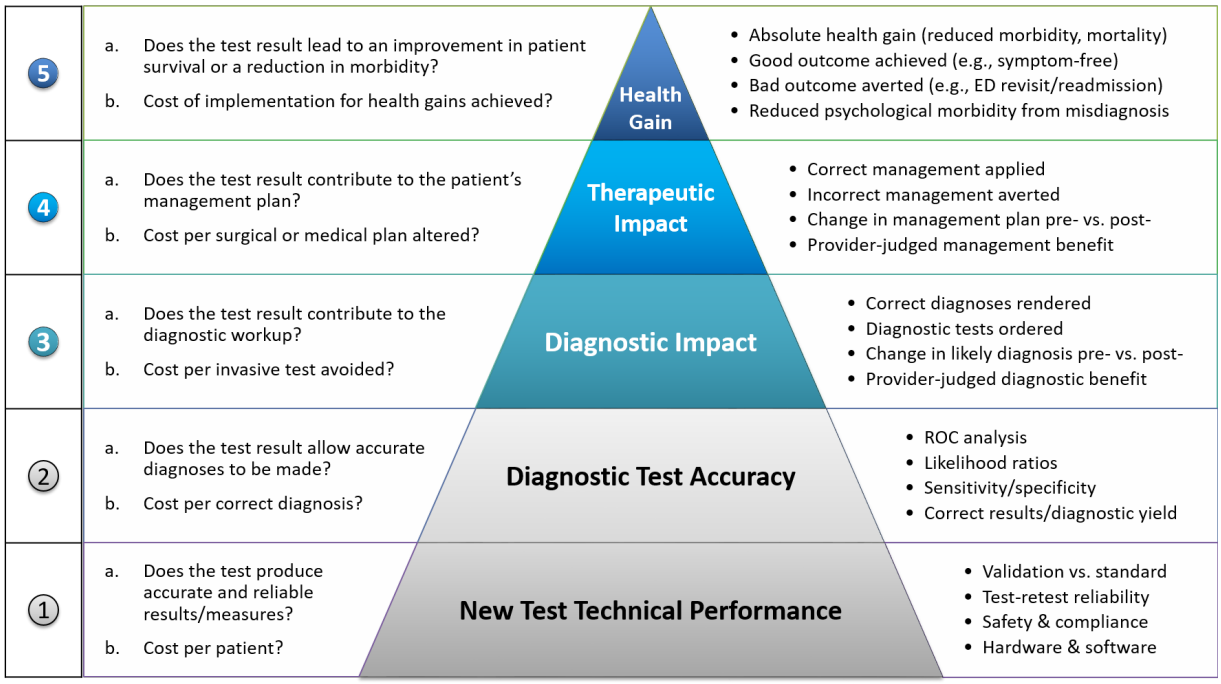

Figure 3. Hierarchical diagnostic test assessment model. This model uses the term "test" broadly to refer to any tool that informs diagnosis, whether it is a typical laboratory or imaging diagnostic test, a bedside clinical examination finding, or an artificial intelligence algorithm. The diagnostic research pipeline should have technologies maturing up the pyramid from more basic diagnostic test attributes such as safety, reliability, and accuracy to therapeutic impact and ultimately health gains. Abbreviations: ED, emergency department; ROC, receiver-operating characteristic. *Figure adapted by David Newman-Toker from Hollingworth & Jarvik, 2007.*[21]

REFERENCES

1. Improving Diagnosis in Healthcare. Institute of Medicine; 2015 [updated 2015; cited 2015 October 28]. Available from: http://www.nationalacademies.org/hmd/Reports/2015/Improving-Diagnosis-in-Healthcare.aspx.
2. Newman-Toker DE, Nassery N, Schaffer AC, Yu-Moe CW, Clemens GD, Wang Z, Zhu Y, Saber Tehrani AS, Fanai M, Hassoon A, Siegal D. Burden of serious harms from diagnostic error in the USA. BMJ Qual Saf. 2023 Jul 17:bmjqs-2021-014130. DOI: 10.1136/bmjqs-2021-014130.
3. Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, Wilson LM, Jia Y, Omron R, Tharmarajah S, Guerin L, Bastani PB, Fracica EA, Kotwal S, Robinson KA. Diagnostic Errors in the Emergency Department: A Systematic Review. Comparative Effectiveness Review No. 258. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 75Q80120D00003.) AHRQ Publication No. 22(23)-EHC043. Rockville, MD: Agency for Healthcare Research and Quality; December 2022 (Updated August 2023). DOI: 10.23970/AHRQEPCCER258. Updated report available from: https://effectivehealthcare.ahrq.gov/products/diagnostic-errors-emergency-updated/research.
4. Newman-Toker DE, Peterson SM, Robinson KA. Frequently Asked Questions for "Diagnostic Errors in the Emergency Department: A Systematic Review." Open Science Framework, 2023. DOI: 10.17605/OSF.IO/B7XVM. Available from: https://doi.org/10.17605/OSF.IO/B7XVM.
5. Newman-Toker DE, Keita M, Nassery N, Schaffer AC, Yu-Moe CW, Saber Tehrani AS, Clemens GD, Wang Z, Fanai M, Siegal D, Padula WV. Total US societal costs of harms from diagnostic error estimated from malpractice and population-based data [abstract]. Diagnostic Error in Medicine 2018; November 4-6, 2018; New Orleans, LA 2018.
6. Saltzman AB, Keita M, Saber Tehrani AS, Hassoon A, Hough DE, Newman-Toker DE. US federal research funding on diagnostic error substantially lags its public health burden. Diagnostic Error in Medicine 2017; October 8-10, 2017; Boston, MA 2017.
7. Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC). National Institutes of Health [March 31, 2023; cited 2023 October 25]. Available from: https://report.nih.gov/funding/categorical-spending#/.
8. Umscheid CA. Diagnostic Safety Portfolio. Armstrong Institute Grand Rounds at the Johns Hopkins University School of Medicine, October 17, 2023 [Lecture].
9. Undiagnosed Diseases Program. National Institutes of Health; [cited 2013 September 15]. Available from: https://www.genome.gov/Current-NHGRI-Clinical-Studies/Undiagnosed-Diseases-Program-UDN.
10. Newman-Toker DE, Curthoys IS, Halmagyi GM. Diagnosing Stroke in Acute Vertigo: The HINTS Family of Eye Movement Tests and the Future of the "Eye ECG." Seminars in Neurology. 2015;35(5):506-21.
11. Tarnutzer AA, Gold D, Wang Z, Robinson KA, Kattah JC, Mantokoudis G, Saber Tehrani AS, Zee DS, Edlow JA, Newman-Toker DE. Impact of Clinician Training Background and Stroke Location on Bedside Diagnostic Test Accuracy in the Acute Vestibular Syndrome - A Meta-Analysis. Ann Neurol. 2023;94(2):295-308.
12. Edlow JA, Carpenter C, Akhter M, Khoujah D, Marcolini E, Meurer WJ, Morrill D, Naples JG, Ohle R, Omron R, Sharif S, Siket M, Upadhye S, LOJ ES, Sundberg E, Tartt K, Vanni S, Newman-Toker DE,

Bellolio F. Guidelines for reasonable and appropriate care in the emergency department 3 (GRACE-3): Acute dizziness and vertigo in the emergency department. Acad Emerg Med. 2023;30(5):442-86.

13. Cancer Facts & Figures 2021. Atlanta: American Cancer Society; 2021 [updated 2021; cited 2023 October 26]. Available from: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html.
14. Stroke Fact Sheet. 2021 [updated 2021 May 4, 2023; cited 2023 October 26]. Available from: https://www.cdc.gov/stroke/facts.htm.
15. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol. 2021;20(10):795-820. PMCID: PMC8443449.
16. Parker TM, Farrell N, Otero-Millan J, Kheradmand A, McClenney A, Newman-Toker DE. Proof of Concept for an "eyePhone" App to Measure Video Head Impulses. Digit Biomark. 2021;5(1):1-8. PMCID: PMC7879263.
17. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, Muti P, Jaeschke R, Guyatt GH. GRADE: assessing the quality of evidence for diagnostic recommendations. ACP J Club. 2008;149(6):2.
18. AHRQ Diagnostic Safety & Quality - Federal Interagency Workgroup on Improving Diagnostic Safety and Quality in Healthcare. Agency for Healthcare Research and Quality, Rockville, MD; 2023 [updated 2023; cited 2023 October 26]. Available from: https://www.ahrq.gov/topics/diagnostic-safety-and-quality.html.
19. Office of Emergency Care Research. OECR; 2023 [updated 2023; cited 2023 October 26]. Available from: https://www.ninds.nih.gov/current-research/trans-agency-activities/office-emergency-care-research.
20. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. BMJ Qual Saf. 2018;27(7):557-66. PMCID: PMC6049698.
21. Hollingworth W, Jarvik JG. Technology assessment in radiology: putting the evidence in evidence-based radiology. Radiology. 2007;244(1):31-8.