July 9, 2021

Dr. Francis Collins
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Dear Dr. Collins:

Nemours Children's Health (Nemours) commends the National Institute of Health's (NIH) efforts to reduce disparities through the use of Artificial Intelligence/Machine Learning (AI/ML). Nemours is one of the nation's largest multistate pediatric health systems, including two free-standing children's hospitals and a network of nearly 80 primary and specialty care practices across five states. Nemours seeks to transform the health of children by adopting a holistic health model that utilizes innovative, safe, and high quality care, while also caring for the health of the whole child beyond medicine. Nemours also powers the world's most-visited website for information on the health of children and teens, KidsHealth.org. In our effort *to create the healthiest generations of children,* Nemours has taken substantive steps to reduce health disparities and advance health equity across our health system for the close to half-a-million unique children Nemours treats annually. Nemours appreciates the opportunity to share our thoughts and recommendations, especially related to the importance of centering efforts on children, a stage of life where it remains possible to shift cyclical and generational disparities.

Nemours Relevant Experience and Expertise

Electronic Health Records

Nemours is among the most advanced pediatric health care systems in its early adoption of an Electronic Health Record (EHR) system, in deriving a research-ready extract from the EHR system that adheres to the Observational Medical Outcomes Partnership (OMOP) standard data model, and in sharing data as part of multiple networks. Use of the OMOP model allows our research EHR data to contribute to national networks such as PEDSnet, PCORnet and the newly developed N3C enclave for COVID-related data. In addition, our health system participates in a number of national disease registries and is a strong leader and advocate of Health Information Exchange for clinical care.

Diversity, Health Equity, and Inclusion

The Nemours Office of Health Equity and Inclusion (OHEI), Research & Education Section (RES), is charged with ensuring that evidence used in advancing care of racial/ethnic minorities, underserved populations, socio-economically challenged, as well as populations defined by health disparities indicators, meets the requirements of rigorous scientific methodology. OHEI also supports an undergraduate internship (Health Equity Summer Scholars Program), which provides an opportunity for undergraduates to acquire research knowledge and skills in areas pertaining to pediatric health disparities science.

Social Determinants of Health Screening

In an ongoing effort to identify and address disparities within our patient population, Nemours developed a strategy to determine how best to learn about our patients' Social Determinants of Health (SDoH) and social needs. Nemours created a pediatric-appropriate SDoH screening tool compiled from several evidence-based social needs questionnaires for adult populations. Between December 2020 and June 2021, Nemours screened over 15,000 primary care patients, with over one third of those requesting assistance with a need. As Nemours advances its SDoH strategy, we will use this and other data to drive the identification of priority areas of focus in order to develop appropriate interventions to address the non-medical needs of our pediatric patient population and their families.

Background

Nemours strongly encourages NIH to center its efforts to reduce disparities through an approach that begins with children and their families. As you know, evidence shows that factors like trauma and stress in a mother's life prior to conception, during pregnancy, and throughout the early postnatal care period can impact the life course of her child.[i] In addition, the health of both parents can all have long-lasting impacts on a child's health and well-being.[ii] Early childhood and adolescence are also critical times of development when one's lived experiences can have lifelong effects on health and wellbeing. Children raised in low-income households are more likely to have poorer neurocognitive outcomes, less educational attainment and lower economic productivity in adulthood – all in turn contributing to the potential for intergenerational poverty.[iii]

This substantial evidence-base indicates that those who have faced challenges in their lives are more likely to not only experience negative health outcomes, but also raise children who will start their lives already at a disadvantage. Nemours urges NIH to focus significant effort on positively impacting the life course of children, starting pre-conception, through childbirth, and continuing into the early years and beyond. Only then can interventions substantively shift cyclical and intergenerational disparities.

Considerations for using AI/ML

AI and ML have immense potential to aid in the assembly and interpretation of multi-generational and multi-site data that is necessary to understand and address health disparities and inequities. However, there is no existing resource that contains the necessary data to make this a reality; hence, the first and perhaps most challenging application of AI/ML techniques will be to stitch together the disparate components needed, and to do so in a way that preserves the privacy of individuals. Privacy-preserving record linkage (PPRL) is an established technology that is already in use in a variety of settings.[iv,v]

However, as the amount of information linked to deidentified individuals increases, so too does the risk of reidentification. Even though they are not identifiers, facts about individuals such as their Rural-Urban Commuting Area (RUCA) code, Child Opportunity Index, Area Deprivation Index, and environmental factors such as air pollutants or distance from known toxic waste facilities, coupled with similar information about their linked parents and other relatives, may make it possible to identify some individuals with high levels of confidence.[vi] **To avoid reidentification, access to combinations of features or facts regarding individuals must be governed carefully and supported by research that clearly identifies reidentification risk given specific combinations of facts.** The opportunity for misuse of such information is substantial, and mistrust—notably among underserved and at-risk populations—threatens to derail efforts to assemble such data unless great care is taken in designing and communicating guard rails to prevent misuse.

Another ripe area of application for AI and, in particular, Natural Language Processing (NLP) is in the interpretation of EHR data from free-form notes fields. Most existing deidentified or Health Insurance Portability and Accountability Act (HIPAA) limited datasets for research purposes—particularly datasets that aggregate multi-institutional data—use a common data model that contains only discrete elements of the EHRs from which they are derived. Free form notes within EHRs potentially contain a wealth of information regarding the status of patients that is not reflected in the discrete diagnostic billing codes, procedure codes and other discrete data elements. Emerging AI/ML applications based on sequence-to-sequence mapping and Transformer technology show great promise for their ability to answer questions of clinical significance and provide concise summaries of provider notes.[vii] **Research is needed to identify best practices in applying NLP to provide discrete and accurate answers to specific questions about patients based on EHR notes without risk of exposing protected personal health information.**

A third research emphasis area for AI/ML should be studies designed to mitigate bias in our numerical models. Bias can arise from many sources in AI/ML including the analysis design phase (e.g., selection of features and choice of ML algorithms), the training phase due to imbalance in the amount of training data for machine learning such that specific populations of patients are underrepresented or misrepresented and in evaluating ML models after training to examine possible biases in model accuracy for protected subgroups of patients.[viii] **As a multi-state pediatric health care organization, Nemours urges NIH to commit to expanding the availability of pediatric data that is needed to inform AI/ML studies leading to improved pediatric health**.

Beyond merely lacking adequate quantities of pediatric EHR data in aggregated or federated research datasets, use of multigenerational datasets will require careful stratification of patients by age and necessitate modifications to our common data models to ensure use of measurement units (e.g., time, anthropometric units) that do not implicitly bias results by lacking the necessary granularity for pediatric patients.

Nemours already has established expertise in AI/ML techniques, managing research-ready deidentified EHR data, and linking geocoded SDoH data to support research on health care disparities. Nemours is also leveraging nearly two decades of EHR data on over three million pediatric patients to discover patterns of care that lead to improved outcomes. AI/ML can evaluate population level data to find factors that contribute to disease, such as biomarkers or sociomarkers, as well as protective factors. They can also develop predictive tools, risk stratification models and geographic targeting approaches, which are critical as the nation continues to move into value-based and population level health care. **Nemours believes it is critically important for this NIH initiative to build on and scale efforts across the United States to truly impact population health.**
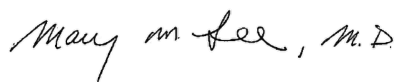
Another area of great promise for AI/ML is its application to precision medicine. Our discussion has focused largely on the incorporation of EHR and SDoH data as input to analyses that will ultimately allow us to create ever more precise patient phenotypes. This greater precision in characterizing patient phenotypes should open the door to AI/ML studies able to untangle the complex relationships between "omics" and health. The massive quantity or data needed to explore this genotype-phenotype complexity demands computer resources with extremely large memory capacity and massively parallel computational capability. Nemours supports a Biomedical Research Informatics Center (BRIC) that provides AI/ML and related computational resources used for a variety of efforts, including basic and clinical research, predictive analytics and quality improvement. Our experiences in developing and maintaining the storage and computational power for BRIC underscore the significant material as well as human resources needed for AI/ML applications in health care. **Nemours strongly encourages NIH to allocate instrumentation grant support for AI/ML-focused special-purpose computer and storage hardware, and training programs**.

Finally, Nemours believes that providing support for academic programs focused on equity, social determinants, disparities and related areas, combined with scholarship and internship opportunities within health care organizations is a way to promote both greater interest and greater diversity in the workforce that will use and develop even more advanced AI/ML tools in the future.

Conclusion

Nemours is committed to developing tools like AI/ML that show promise to improve patient care and long-term health outcomes in our effort *to create the healthiest generations of children*. Nemours has taken steps to address disparities and inequities within our patient population and is committed to continuing this work through new and innovative methods. Nemours thanks NIH for its leadership in advancing this RFI and appreciates your consideration of our response. Nemours looks forward to collaborating with the NIH on this important effort. Please reach out with any questions.

Sincerely,

Mary M. Lee, MD
Executive Vice President, Chief Scientific Officer,
Nemours Children's Health System
Physician in Chief, Delaware Valley Operations
Professor of Pediatrics, Sidney Kimmel Medical College
Jefferson University

Citations

[i] Horsch, A., & Stuijfzand, S. (2019, 2019/05/27). Intergenerational transfer of perinatal trauma-related consequences. *Journal of Reproductive and Infant Psychology, 37(3), 221-223.* https://doi.org/10.1080/02646838.2019.1629190

[ii] Stephenson, J., Heslehurst, N., Hall, J., Schoenaker, D., Hutchinson, J., Cade, J. E., Poston, L., Barrett, G., Crozier, S. R., Barker, M., Kumaran, K., Yajnik, C. S., Baird, J., & Mishra, G. D. (2018). Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet (London, England), 391(10132), 1830–1841.* https://doi.org/10.1016/S0140-6736(18)30311-8

[iii] Jensen SKG, Berens AE, Nelson CA 3rd. Effects of poverty on interacting biological systems underlying child development. *Lancet Child Adolesc Health. 2017 Nov;1(3):225-239.* doi: 10.1016/S2352-4642(17)30024-X. Epub 2017 Jul 27. PMID: 30169171.

[iv] Vatsalan, et al., A taxonomy of privacy-preserving record linkage techniques, Information Systems (2013), http://dx.doi.org/10.1016/j.is.2012.11.005

[v] Datavant. (2021). *Connecting the World's Health Data.* https://datavant.com/

[vi] Kind AJH, Buckingham W. Making Neighborhood Disadvantage Metrics Accessible: The Neighborhood Atlas. *New England Journal of Medicine*, 2018. 378: 2456-2458. DOI: 10.1056/NEJMp1802313. PMCID: PMC6051533.

[vii] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *ArXiv, abs/1706.03762.*

[viii] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of internal medicine, 169*(12), 866–872. https://doi.org/10.7326/M18-1990