May 4, 2021

David Meyers, M.D.
Acting Director
Agency for Healthcare Research and Quality
5600 Fishers Lane
Rockville, MD 20857

Re: Use of Clinical Algorithms That Have the Potential to Introduce Racial/Ethnic Bias into Healthcare Delivery

Dear Dr. Meyers:

The American Osteopathic Association (AOA) and American College of Osteopathic Internists (ACOI), on behalf of the more than 151,000 osteopathic physicians (DOs) and medical students we represent, appreciate the opportunity to support the Agency for Healthcare Research and Quality 's (AHRQ) review of evidence on the use of race/ethnicity in clinical algorithms and the potential for algorithms to contribute to disparities in healthcare. Our organizations recognize that all healthcare stakeholders share responsibility for reducing disparities in healthcare and ensuring that all Americans have access to high-quality, equitable care. DOs are especially connected to this belief and have a unique and important perspective on the delivery of healthcare in our nation, as many osteopathic medical schools are located in proximity to health professional shortage areas and help increase the physician workforce in rural and underserved communities. Additionally, nearly 40 percent of physicians practicing in medically underserved areas are DOs.

While digital, analytic, and clinical decision-making tools are intended to help advance health care, point of care tools and other algorithms also have the potential to introduce biases. Often, these tools take into account racial, ethnic, and socioeconomic factors and can introduce bias unbeknownst to the user. The primary way to ensure that clinical tools are designed and utilized in an unbiased and equitable manner is to engage in extensive review of the evidence available, identify research gaps, and evaluate these tools in a real-world setting. We commend AHRQ for beginning this process, and we appreciate the opportunity to assist by providing the following feedback on the agency's request for information (RFI). We also want to encourage AHRQ to disseminate the results of this review to help improve clinicians awareness when tools have a high likelihood of exacerbating health disparities.

Questions 1 and 2: What clinical algorithms are used in clinical practice, hospitals, health systems, payment systems, or other instances? What is the estimated impact of these algorithms in size and characteristics of population affected, quality of care, clinical outcomes, quality of life, and health disparities? Do these algorithms include race/ethnicity as a variable

and, if so, how was race and ethnicity defined (including from whose perspective and whether there is a designation for mixed-race or multiracial individuals)?

It would be difficult to comprehensively enumerate all algorithms or risk assessment tools used in clinical practice across specialties and settings for all health conditions. Listed below are some commonly used algorithms which to our knowledge include race or ethnicity as a variable:

- Kidney GFR estimation: The equation used to calculate estimated glomerular filtration rate typically has an adjustment only for individuals who are African American.
- Pulmonary function tests: Spirometry tests typically utilize a race/ethnicity adjustment.
- American College of Cardiology/ American Heart Association Atherosclerotic cardiovascular disease (ASCVD) Risk Calculator[1]: The ASCVD risk estimator is among the more commonly used estimators for cardiovascular disease, and it takes a patient's race into account in its calculation only for patients who are white and African American[2]. The race field does not apply to patients of other races.
- FRAX risk calculator: There are many variations, designed for different countries, that account for demographic variations in populations.
- Breast Cancer Risk Assessment Tool: The tool allows for an individual to be identified as white, Hispanic, African American, Asian, Native American, or other.
- Colorectal Cancer Risk Assessment Tool: The tool allows for ethnicity to be identified as Hispanic or non-Hispanic, and then asks for race to be identified as white, African American, or Asian American/Pacific Islander.
- Readmission Risk Scoring Tools: These tools will vary based on the developer.

Many of these algorithms have not been sufficiently studied for their impact on racial/ethnic groups who experience disparities other than African Americans, and further study is needed. It is important to note that this is not a comprehensive list, as use of diagnostic tools will often be limited to a specific specialty. For example, the article by Vyas et al. published in the *New England Journal of Medicine*, cited in the RFI, discusses other common algorithms with the potential to introduce bias. It includes tools used in specialties such as nephrology, obstetrics, oncology, and endocrinology that are not listed here.[3] We also want to highlight that the bias in these algorithms have downstream effects for patients of color in addition to care disparities, such as payor coverage decisions.

Question 3: Do the algorithms in question 1 include measures of social determinants of health (SDOH) and, if so, how were these defined? Are these independently or collectively examined for their potential contribution to healthcare disparities and biases in care?

---

[1]ASCVD Risk Estimator Plus. American College of Cardiology. Available at: https://tools.acc.org/ascvd-risk-estimator-plus/#!/calculate/estimate/

[2] Lloyd-Jones DM, Braun LT, Ndumele CE, Smith SC Jr, Sperling LS, Virani SS, Blumenthal RS. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American Heart Association and American College of Cardiology. Circulation. 2019;139:e1162–e1177. DOI: 10.1161/CIR.0000000000000638. https://www.ahajournals.org/doi/10.1161/CIR.0000000000000638

[3] Vyas et al. Hidden in Plain Sight- Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine* 383;9. June 17, 2020. Available at: https://www.nejm.org/doi/pdf/10.1056/NEJMms2004740

To our knowledge, most of the algorithms or risk assessment tools that we highlight under *Question 1* do not account for social determinants of health, and accounting for social determinants can strengthen use of these assessment tools. We discuss this in more detail, specific to algorithms we identified, in response to *Question 6.*

Question 6: Which clinical algorithms have evidence that they contribute to healthcare disparities, including decreasing access to care, quality of care or worsening health outcomes for Black, Indigenous, and other people of color? What are the priority populations or conditions for assessing whether algorithms increase racial/ethnic disparities? What are the mechanisms by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color?

Most health disparities are the result of structural inequalities that exist in our society, although the way in which some clinical tools account for racial and ethnic differences in disease risk may make it appear that differences are solely related to genetic or biologic factors. Across a number of tools, studies have shown risk to be strongly correlated with other socio-economic factors, such as poverty, lack of or inadequate insurance coverage, out-of-pocket expenses, and access to care.   Therefore, current calibrations that only take race/ethnicity into account are insufficient. Evidence demonstrating this for several different clinical tools from *Question 1* can be found below. We share the information below with the intention of helping to stimulate further research, analysis, and discussion on what variables in algorithm development result in the best care for patients of all backgrounds.

Kidney Glomerular Filtration Rate Estimation (eGFR): There is clear evidence that current approaches to calculating estimated GFR are biased and result in unequal care. However, work needs to be done to identify what factors contribute to an accurate calculation of eGFR to ensure proper identification of kidney disease and delivery of care. Identifying alternatives is important because race is a social construct, and as an article by Eneanya et al. explains, "kidney function estimating equations that include race as a variable cause problem for transparency and unduly restrict access to care in some cases, yet offer only modest benefits to precision." An article published in Nature notes that removal of the correction "would lead to a change in diagnosis for 3.5% of Black adults from 'disease free' to having early-stage kidney disease (extended to the US population, this would be one million Black adults). Removing it [the correction] would also shift the status of 29% of Black patients from having early-stage to advanced disease."[4] However, evidence on the impact of immediate elimination of the race/ethnicity variable in eGFR calculation without a substitute is mixed, and an article published in JAMA Internal Medicine in 2020 found that eliminating the race variable resulted in consistent underestimation of mGFR for African American patients, which "may have unintended consequences in African American individuals, such as inappropriate early transplant or dialysis initiation, overdiagnosis of CKD, overestimation of the association of the risk of adverse outcomes with reduced GFR, inadequate dosing of drugs excreted by glomerular filtration (e.g.,

---

[4] Madhusoodanan. "Is a racially-biased algorithm delaying health care for one million Black people?" Nature. Available at: https://www.nature.com/articles/d41586-020-03419-6#ref-CR4

some antibiotics and cancer chemotherapy), and limited access to tests."[5] However, the article also argues that better methods need to be developed to estimate GFR. Alternatives to the calculation adjustment based on race could be adjustments for biomarkers in addition to creatine levels, or accounting for social factors.

Pulmonary Function Tests: Pulmonary function tests using spirometers involve a race-based adjustment, assuming lower lung capacity for individuals of ethnic minorities. It assumes 10-15% lower lung capacity for African American patients and 4-6% lower capacity for Asian patients when compared to white patients. However, lung function is strongly associated with socioeconomic status as individuals with lower income are more likely to live in areas with high air pollution, have poorer air quality in housing, smoke, and have occupational exposure to pollutants. Individuals with these risk factors are more likely to have poor ling function, and ultimately pulmonary disease.[6] As a result, adjustments that assume biologic differences are most likely inaccurate and pose the likelihood of missed diagnoses and failure to connect patients to treatment they may need. These adjustments are especially concerning when taken in the context of COVID-19. An article published in *The Lancet* notes that "during the COVID-19 pandemic, these race adjustments could potentially cause clinicians to miss important diagnoses. For example, restrictive ventilatory dysfunction is emerging as a problem in COVID-19 patients, evident for at least 2 weeks after hospital discharge. This defect is indicated by a spirometry measure of forced vital capacity below the lower limits of normal for the appropriate reference population. Clinicians might miss this diagnosis if lower lung capacity measures are considered normal for minority populations."[7]

American College of Cardiology/ American Heart Association Atherosclerotic Cardiovascular Disease (ASCVD) Pooled Cohort Equation: Further investigation is required on disparate effects of this calculation's use in different populations. There is a significant knowledge gap in regard to "accurate ASCVD risk estimation specific to persons of Asian/Pacific Islander and Hispanic ethnicities, who are currently combined with the white population in the Pooled Cohort Risk Equation." A study recently published in the Journal of the American College of Cardiology found that the current ASCVD Pooled Cohort Risk Equation should be recalibrated, especially considering the Pooled Cohort Risk Equation overestimated actual ASCVD risk across all 4 ethnic groups studied.[8] Another analysis found concerns with the studies used to calibrate the equation for the African American population, especially in light of the fact that African Americans have 2 times the risk of heart failure and 1.5 to 2 times the risk of coronary heart disease when compared to the non-Hispanic white population.[9] Additionally, there is evidence that ASCVD risk calculation is poorly calibrated when accounting for social determinants of

[5] Levey, Andrew S et al. "Estimation of Glomerular Filtration Rate With vs Without Including Patient Race." *JAMA internal medicine* vol. 180,5 (2020): 793-795. doi:10.1001/jamainternmed.2020.0045 Available at:

[6] Hegewald MJ, Crapo RO. Socioeconomic status and lung function. Chest. 2007 Nov;132(5):1608-14. doi: 10.1378/chest.07-1405. PMID: 17998360.

[7] Anderson et al. "Could routine race-adjustment of spirometers exacerbate racial disparities in COVID-19 recovery?" The Lancet Volume 9, ISSUE 2, P124-125, February 01, 2021 Available at: https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30571-3/fulltext

[8] Rana et al. "Accuracy of the Atherosclerotic Cardiovascular Risk Equation in a Large Contemporary, Multiethnic Population" J Am Coll Cardiol. 2016 May, 67 (18) 2118–2130

[9] Goff DC, Lloyd-Jones DM. The Pooled Cohort Risk Equations—Black Risk Matters. JAMA Cardiol. 2016;1(1):12–14. doi:10.1001/jamacardio.2015.0323

health, such as income and education level. A study found that the ASCVD may be underestimating risk in socioeconomically disadvantaged populations, defined by indicators that include income, education, and housing.[10] This indicates that ASCVD risk should better take these factors into account, and current calibration may need updating.

FRAX Risk Calculator: The NEJM article by Vyas at al. questions the utility of the race/ethnicity variable in the US FRAX calculator used to determine fracture risk. The article cites that the calculator determines a lower fracture risk for women who are Black (by a factor of 0.43), Asian (0.50) or Hispanic (0.53), often resulting in delayed care for osteoporosis treatment. There is further evidence that alternative approaches to calculating fracture risk without race/ethnicity information can be just as accurate.[11] However, recent articles published in *Osteoporosis International* in response to the NEJM article argue that FRAX still serves as a useful guide.[12,13] Based on conflicting evidence, and the seriousness of the concern that women of color could receive delayed care as a result of this tool, we would encourage further investigation into the impact of the FRAX calculator on patients of color, and whether the tool requires adjustment.

Breast Cancer and Colorectal Cancer Risk Assessment Tools: We are unaware of how or whether the algorithm may contribute to bias. However, it is worth noting that significant care disparities exist in screening and diagnosis for both conditions. For colorectal cancer, 47% of non-Hispanic Whites are up to date on screening compared with 38% of Blacks.[14] A study by Peterson et al. found that "referral rates for genetic counseling in women with breast cancer differ by race. Race was independent of employment, Medicaid insurance, and chemotherapy status in influencing referral patterns."[15]

Questions 7 and 11: To what extent are users of algorithms, including clinicians, health systems, and health plans, aware of the inclusion of race/ethnicity or other variables that could introduce bias in these algorithms and the implications for clinical decision making? What evidence is available about the degree to which the use of clinical algorithms contributes to bias in care delivery and resulting disparities in health outcomes? To what extent are patients aware of the inclusion of race/ethnicity or other variables that can result in bias in algorithms that influence their care? Do providers or health systems communicate this information with patients in ways that can be understood?

[10] Colantonio et al. "Performance of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations by Social Deprivation Status". Journal of the American Heart Association. 17 Mar 2017. Available at: https://www.ahajournals.org/doi/10.1161/JAHA.117.005676
[11] Ishii, Shinya et al. "Fracture risk assessment without race/ethnicity information." The Journal of clinical endocrinology and metabolism vol. 97,10 (2012): 3593-602. doi:10.1210/jc.2012-1997
[12] Kanis, J.A., Cooper, C., Dawson-Hughes, B. et al. FRAX and ethnicity. Osteoporos Int 31, 2063–2067 (2020). https://doi.org/10.1007/s00198-020-05631-6
[13] Lewiecki, E., Wright, N. & Singer, A. Racial disparities, FRAX, and the care of patients with osteoporosis. Osteoporos Int 31, 2069–2071 (2020). https://doi.org/10.1007/s00198-020-05655-y
[14] Fiscella, Kevin et al. "Eliminating disparities in cancer screening and follow-up of abnormal results: what will it take?." Journal of health care for the poor and underserved vol. 22,1 (2011): 83-100. doi:10.1353/hpu.2011.0023
[15] Peterson et al. "Racial disparities in breast cancer hereditary risk assessment referrals". Journal of Genetic Counseling. March 2020. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/jgc4.1250

To our knowledge, physicians and other clinicians are often unaware of potential biases that can be produced by clinical tools such as those we previously mentioned. This is largely because physicians may not be informed on how the race/ethnicity variable in the tool impacts the final calculation. In order to use clinically recommended tools, a physician may not have visibility to the population sample and methods used in designing a clinical algorithm. Physicians rely on many of these listed tools because they are clinically recommended by practice guidelines, and the physician is doing what they believe is in the best interest of the patient.

This lack of awareness is especially true for novel tools such as analytics and machine learning tools that are continuously updated based on population data. The average clinician does not have visibility to all inputs and assumptions that these tools are built upon, and we believe standards should be developed to ensure that what happened in the case of the Optum tool discussed by Obermeyer et al., and referenced in the RFI, does not continue to occur. Gianfrancesco et al. identify several sources of bias in the development of analytics and machine learning tools, and they provide a good framework for how new tools should be evaluated for these sources of bias.[16] Biases in algorithms can originate from various sources, whether it is missing data, flawed sample sizes, misclassification or measurement error, or other flawed assumptions.

If physicians are aware that a tool has a high likelihood of exacerbating health disparities, they would likely question its design and utility. Additionally, if physicians are unaware of how variables factor into an algorithm's output, they would be unable to communicate this with patients.

Conclusion
The AOA and ACOI truly appreciate this opportunity to share input on how certain clinical algorithms have the potential to introduce bias into clinical care. If you have any questions or would like to discuss our comments, please contact either Lisa Miller, MS, AOA Senior Director of Regulatory Affairs and Policy Engagement at lmiller@osteopathic.org or (202)-349-8744; or Tim McNichol, JD, ACOI Deputy Executive Director at tmcnichol@acoi.org or (301) 231-8877. Our organizations stand ready to assist you in subsequent policy development if called upon. On behalf of the entire osteopathic medical profession, thank you for considering our comments.

Sincerely,

| Thomas L. Ely, DO | Kevin M. Klauer, DO, EJD | Michael A. Adornetto, DO, MBA |
| President, AOA | Chief Executive Officer, AOA | President, ACOI |

[16] Gianfrancesco, Milena A et al. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data." JAMA internal medicine vol. 178,11 (2018): 1544-1547. doi:10.1001/jamainternmed.2018.3763