



OUR PATIENTS. OUR UNION. OUR VOICE.

WASHINGTON DC
8455 Colesville Road
Suite 1100
Silver Spring MD 20910
phone: 800-287-5021
fax: 240-235-2019

OAKLAND
155 Grand Avenue
Suite 100
Oakland CA 94612
phone: 800-504-7859
fax: 510-663-1625

Via Electronic Webform at ahrq.gov

May 11, 2021

Xavier Becerra, Secretary
U.S. Department of Health and Human Services
Attn: 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, Request for Information (RFI), 86 Fed. Reg. 12,948 (Mar. 5, 2021), Document No. 2021-04509

Dear Secretary Becerra,

On behalf of more than 170,000 registered nurses (RNs) and as the largest union and professional association of registered nurses in the United States, National Nurses United (NNU) submits these comments and materials in response to the Request for Information (RFI) on “Use of Clinical Algorithms That Have the Potential To Introduce Racial/Ethnic Bias Into Healthcare Delivery,” 86 Fed. Reg. 12,948 (Mar. 5, 2021), by the Agency for Healthcare Research and Quality (AHRQ).

NNU applauds AHRQ’s decision to inquire into how clinical algorithms can introduce racial and ethnic bias into clinical practice and degrade the care Black, Indigenous, and other people of color receive. Equally important, is ensuring that the implementation of clinical algorithms does not interfere with clinicians’ professional judgment. NNU supports policies that ensure registered nurses and other clinicians may exercise their professional judgment to determine the best course of action for their patients and override clinical algorithms when that is the best course of action for their patients. As registered nurses, NNU’s members work as bedside health care professionals in nearly every state in the nation and our members have witnessed the ways in which clinical algorithms can harm patients.

Clinical algorithms are typically based on practice guidelines developed by a panel of experts or, through machine learning, based on data from a particular set of patients. Practice guideline development by experts requires significant exercise of judgment which can introduce error and bias. Clinical algorithms based on machine learning also face risks of error and bias as well as an unacceptable lack of transparency. Both approaches to creating clinical algorithms face problems with the evidence base available including racial discrimination in clinical trials and access to health care as well as other deficiencies.

NNU knows that clinical algorithms can interfere with safe, therapeutic health care that meets the needs of each individual patient. Even under optimal conditions, clinical algorithms are based on population-level data and are not appropriate for every patient. In addition, the way clinical algorithms are implemented, regardless of how they are created, often inappropriately constrains the use of health care professionals' judgment, which can worsen the impact of a biased algorithm.

Clinical algorithms should serve as guidelines, but employers often require rigid adherence with the goal of controlling costs and eliminating variation in treatment. First, a lower cost treatment may not be the best option.¹ Second, health care corporations' focus on eliminating variation aims to facilitate automation and the use of less costly labor, as individualized care requires professional judgment by skilled clinicians. Yet patients are diverse, with unique needs and values. Thus, NNU believes that decisions on patient care should be made by skilled clinicians using their professional judgment in a manner that is consistent with an individual patient's preferences and in the patient's best interests. Moreover, health care professionals must have the autonomy to override clinical algorithms, within their scope of practice, and not be constrained by mandates established in a corporate boardroom. Finally, NNU supports mandates for transparency as well as user and patient education on the risks associated with clinical algorithms.

NNU urges AHRQ to address several essential points that apply to all algorithms, not just those that have already been proven to be biased, as it conducts its inquiry into how clinical algorithms can introduce racial and ethnic bias into health care delivery. First, there are fundamental limits on the ability of algorithms to meet the needs of individual patients, especially when those patients are part of racial or ethnic groups that are less well represented in the data. Second, AHRQ must consider the importance of ensuring that health care professionals have full autonomy to override algorithmic recommendations, within their scope of practice and consistent with their patients' needs and preferences. Third, transparency and education about all aspects of the creation, use, and impact of these algorithms is necessary to detect and combat bias. Fourth, transparency must extend to clinicians, patients, and the public, including thresholds for recommendations to provide and deny preventive, diagnostic, or treatment measures by race, ethnicity, gender, and other relevant data. Finally, NNU urges AHRQ to support robust regulation of clinical algorithms as medical devices and to ensure that the supporting evidence and reasoning on which clinical recommendations are based are sufficient, sound, transparent, and intelligible and that recommendation thresholds for providing or denying preventive, diagnostic, or treatment measures are transparent to clinicians, patients, and the public.

¹ For example, doctors at a California hospital contend that the system to which the hospital belongs has standardized clinical guidelines through a shared electronic health record system that are "often driven by cost considerations" and that the guidelines "often [conflict] with their own judgment of best medical practices." Wolfson, BJ. Orange County Hospital Seeks Divorce From Large Catholic Health System. *Kaiser Health News*. <https://khn.org/news/article/orange-county-hospital-seeks-divorce-from-large-catholic-health-system/>. Published April 13, 2012. Accessed May 1, 2021.

These comments have five main sections. Section I discusses problems related to algorithms based on clinical practice guidelines developed by experts. Section II discusses problems related to algorithms developed through machine learning. As part of the response to this RFI, NNU also has included preliminary results of our survey of RNs on their experiences with racial bias in clinical algorithms, which are detailed in Sections III and IV of these comments. Section V offers concluding recommendations.

I. Algorithms based on clinical practice guidelines developed by experts

Answering RFI question 1 on the estimated impact of these algorithms on quality of care, clinical outcomes, quality of life, and health disparities and RFI question 6 on the mechanisms by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color.

A. **Algorithms based on clinical practice guidelines rely on population level data that may be inappropriate for individual patients or subgroups, introducing opportunities for structural racial bias to influence health care delivery.**

Clinical algorithms rely on population level data that does not reflect each individual patient's needs and can introduce bias into health care, if clinicians are pressured to apply clinical algorithm recommendations to all patients despite individual care needs. Practice guidelines are typically based on studies and data regarding a certain percentage of a patient population as a whole and, thus, may not be appropriate for a particular patient.² This is the nature of statistics: a treatment that works in a high percentage of a population is considered high-quality care. The trick, of course, is in determining where an individual patient falls relative to the population as a whole.

Patients have a range of individual factors that influence how well a particular treatment plan works for them, including their own preferences. Some proponents of using algorithms in health care delivery claim that more complex algorithms can successfully predict the best treatment plan for an individual based on a range of factors. However, the degree to which an algorithm works depends on how closely a patient matches the population reflected in the underlying data and on how well the algorithm accounts for all the relevant factors in their life, whether the factors are objective empirical issues or subjective values and preferences. Many algorithms do not even claim to account for patient preference, giving patients only the option to be considered compliant or not.³ As discussed in the next section and in Section IIA, the data underpinning these algorithms is influenced by structural racism in health care research, access, and quality, so reliance on population-level data is less likely to lead to quality care for individuals who are racial and ethnic minorities.

² Tritter, JQ, Lutfey K, McKinlay, J. What are tests for? The implications of stuttering steps along the US patient pathway. *Social Science & Medicine* 2014;107: 37–43.

³ See Krumholz HM. The New Cholesterol and Blood Pressure Guidelines Perspective on the Path Forward. *Journal of the American Medical Association*, 2014;311(14):1403-1405. doi:10.1001/jama.2014.2634.

The problem of applying population-level data is exacerbated when a hospital or health system uses an average as a benchmark for clinical performance. Clinical algorithms provide recommendations for treatment decisions, such as how long a patient should stay in the intensive care unit (ICU).⁴ When doctors know their performance is measured by the average number of hours their patients stay in the ICU, they may feel pressured to send on any patient who stays longer than that average. That pressure is the point of instituting clinical algorithm benchmarks, which are sold promising increased efficiency and reduced costs.⁵ However, that pressure will disadvantage the patients who need the longer stay. Moreover, if a doctor's patient load skews less healthy than the data mix underlying the algorithm, the doctor will have to choose between risking their patients' lives by moving them out of the ICU early or damaging their career by missing the benchmark. Forcing doctors to make that choice may lead to racial bias because disparities in health care access mean Black, Indigenous, and other people of color are more likely to have underlying conditions that complicate treatment. In addition, residential segregation means that racial groups are likely to be concentrated in certain health care facilities.⁶ Therefore, unless explicitly analyzed and validated, Black, Indigenous, and other people of color are more likely to be treated by a doctor whose patient load skews less healthy than the benchmark population. Moreover, even if explicitly analyzed and validated, an algorithm's recommendation may be inappropriate for a particular patient regardless of their race or ethnicity.

The mismatch between population-level data embodied in clinical algorithms and individual patients has an impact on quality of care, RFI question 1, and is one mechanism by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color, RFI question 6. NNU urges AHRQ to investigate the inherent limitations of algorithms to determine appropriate individual patient care and the disparate impact on racial and ethnic lines of those limitations.

B. The medical evidence base underlying clinical algorithms reflect racial bias in health care research, access, and quality.

Clinical algorithms based on available medical research are often designed without regard to relevant differences among patients. Thus, they often perform poorly for particular subgroups of patients. The influence of structural racism on access to health care, quality of health care, and inclusion in clinical studies means that Black, Indigenous, and other people of color are often

⁴ See, e.g., How Allina Saved \$13 Million By Optimizing LOS. *HealthCatalyst*. February 10, 2016. https://www.healthcatalyst.com/success_stories/allina-saved-13-million-optimizing-length-of-stay/. Accessed April 26, 2021.

⁵ Navathe AS, Emanuel EJ. Physician Peer Comparisons as a Nonfinancial Strategy to Improve the Value of Care. *JAMA*. 2016;316(17):1759–1760. doi:10.1001/jama.2016.13739; Cosgrove T, Lee TH. Engaging Doctors in the Health Care Revolution. *Harvard Business Review* 2014; Also see, e.g., *Ibid.*; Clairvia Announces Breakthrough Care Value Analytics Service, *businesswire* April 12, 2011. <https://www.businesswire.com/news/home/20110412005618/en/Clairvia-Announces-Breakthrough-Care-Value-Analytics-Service>. Accessed May 10, 2021.

⁶ Gaskin DJ, Dinwiddie GY, Chan KS, McCleary RR. Residential segregation and the availability of primary care physicians. *Health Serv Res*. 2012;47(6):2353-2376.

poorly represented in the medical evidence base. Therefore, recommendations made by clinical algorithms are less likely to be the best choices for their care—whether or not the algorithms explicitly incorporate race or ethnicity as input factors.

Race and ethnicity influence how closely a patient resembles the population underlying algorithmic decisions. They are also correlated with conditions that make some treatment more or less effective, particularly those related to poverty. As Robert Hamm and Zsolt Nagykaladi summarize in the *Journal of Cognitive Engineering and Decision Making*:

A larger issue is that there are specific patients who are not well served by guidelines. As Klein, Woods, Klein, and Perry (2016) and Falzer (2018) argue, although an [evidence-based clinical practice guideline (EB-CPG)] will, by design, produce better results for typical patients, those commonly included in the supporting studies, a guideline's recommendation may be inappropriate, infeasible, or harmful for other patients. Exceptions may be needed based on patient medical situation (comorbidities, age), patient preferences and goals (Mold, 2017; Mold, Hamm, & Scheid, 2003), patient competence (inability to adhere to, or understand and stick to, a treatment plan), patient social situation (addiction, homelessness, unreliability about clinic appointments), or patient economic situation (inability to pay for visits, prescriptions, or a healthy diet). With EB-CPG regimens, situations can arise where physicians are penalized for doing what they know will be best for a patient. Unless there is allowance for exceptions, a reward regimen can punish those physicians who use more effort and judgment to make better choices for their patients. This can exacerbate physician income differences contingent on the wealth of the patients served: When the measured conformance to a guideline depends partly on patient behavior, the regimen rewards those physicians who are already better rewarded because they serve patients with orderly middle or upper-class lives and a higher capacity to become or stay healthy.⁷

The Institute of Medicine (IOM) made a similar observation—that clinical practice guidelines are often inappropriate for certain populations—when it issued best practices for guideline development in 2011. After a thorough review of the processes used to develop clinical practice guidelines, IOM concluded that “evidence supporting clinical decision making and [clinical practice guideline] development relevant to subpopulations, such as patients with comorbidities, the socially and economically disadvantaged, and those with rare conditions, is usually absent.”⁸ Structural racism in the U.S. means that Black, Indigenous, and other people of color are more likely to be poor and have less access to health care throughout their lives than

⁷ Hamm RM., Nagykaladi ZJ., Physician Judgment and Clinical Practice Guidelines, *Journal of Cognitive Engineering and Decision Making*, 2018;12(3): 209-214 doi: 10.1177/1555343418782850.

⁸ Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. Eds. Graham, Robin et al. 2011. p.2.

white people.⁹ As we are seeing in the COVID-19 pandemic, this means these groups are more likely to have comorbidities that limit their participation in clinical trials¹⁰ or an economic situation that can affect their ability to maintain the kind of consistent treatment plan assessed in clinical trials. Research into clinical trial diversity bears this out.

Clinical trials often fail to include enough people of color to ensure that their results reflect patient diversity. In the Food and Drug Administration (FDA) snapshot of participants in global drug trials between 2015 and 2019, 76% of participants were white, 11% were Asian, 7% were Black/African American, 5% were marked as “other”, and 1% were American Indian or Alaskan Native.¹¹ The percentage of white people in the U.S. population is similar, 76.3%.¹² However, the percentage of Black people in the U.S. population, 13.4%, is substantially higher than the 7% in trials. Only 13% of global clinical drug trial participants were Hispanic or Latino, compared with 18.5% of the U.S. population. Thus, algorithmic recommendations based on evidence from these trials will represent white patients better than Black patients and non-Hispanic or -Latino patients better than Hispanic or Latino patients, both in absolute terms and relative to their representation in the U.S. population.

Algorithm developers sometimes attempt to account for these disparities by encoding race or ethnicity into their algorithms. However, including race and ethnicity explicitly in a clinical algorithm can also end up perpetuating disparate health care access. Health outcome differences that appear to be attributed to race, which is a social construct and not a biological reality, may be more correctly tied to correlated factors that the algorithm does not measure, such as socioeconomic status or muscle mass.¹³ In these cases, the use of race as a proxy for these factors will make the algorithm a poor fit for individuals of the targeted race who do not share the same correlated factor as the original data set. Racism’s effects are pervasive but vary by context. Correlations with race and other factors may be very different between a study population and the ultimate patient population and will certainly not apply to every individual. The broad application

⁹ See, e.g., Serchen, J, et al. Understanding and Addressing Disparities and Discrimination Affecting the Health and Health Care of Persons and Populations at Highest Risk, Position Paper, American College of Physicians; 2021. https://www.acponline.org/acp_policy/policies/understanding_discrimination_affecting_health_and_health_care_per_sons_populations_highest_risk_2021.pdf. See, generally, Pallok, K, De Maio, F & Ansell, DA. Structural Racism — A 60-Year-Old Black Woman with Breast Cancer. *N Engl J Med* 2019; 380:1489-1493 DOI: 10.1056/NEJMp1811499; Ray, Kristen N. Disparities in Time Spent Seeking Medical Care in the United States. *JAMA Intern Med.* 2015;175(12):1983-1986. doi:10.1001/jamainternmed.2015.4468.

¹⁰ Duma N, Kothadia SM, Azam TU, et al. Characterization of Comorbidities Limiting the Recruitment of Patients in Early Phase Clinical Trials. *Oncologist.* 2019;24(1):96-102. doi:10.1634/theoncologist.2017-0687. Accessed May 10, 2021.

¹¹ U.S. Food and Drug Administration, 2015-2019 Drug Trials Snapshots Summary Report. <https://www.fda.gov/media/143592/download>. Published November 2020. Accessed May 3, 2021. p.9.

¹² United States Census, Quick Facts, Population estimates. <https://www.census.gov/quickfacts/fact/table/US#>. Published July 1, 2019. Accessed May 3, 2021.

¹³ See Letter from U.S. Senators Elizabeth Warren, Ron Wyden, and Cory A. Booker and Rep. Barbara Lee to AHRQ Director Gopal Khanna. Sept. 22, 2020. <https://www.warren.senate.gov/imo/media/doc/9.22.2020%20Letter%20to%20AHRQ%20re%20Use%20of%20Rac%20in%20Clinical%20Algorithms.pdf>. Accessed May 10, 2021.

of a clinical algorithm without the treating health care provider's ability to deviate based on their professional judgment and on the needs and preferences of individual patients can introduce errors into the care of these patients.

The racial bias in the data underlying clinical algorithms has an impact on quality of care, clinical outcomes, quality of life, and health disparities, RFI question 1, and is one mechanism by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color, RFI question 6. NNU urges AHRQ to investigate the ways that Black, Indigenous, and other people of color are systemically excluded or underrepresented in the medical evidence base that underlies clinical algorithms and how this underrepresentation leads to less accurate recommendations for these patients.

C. Clinical algorithms reflect the judgment and biases of their creators.

While clinical algorithms may purport to be an objective analysis of the scientific evidence, in fact their development involves significant use of judgment by their creators and creates the opportunity for creator bias—from conflicts of interest, limited perspective on the lives of racial minorities, or implicit racial bias—to be introduced into the algorithm. To create clinical practice guidelines, panels of experts convene to develop guidelines for a particular medical condition or care situation.¹⁴ Panels go through a process, which varies from institution to institution, designed, in theory, to determine the best care practices based on the best available evidence. The panel must perform a meta-analysis of a body of evidence that entails establishing the criteria for including a research study followed by synthesizing studies that may be based on conflicting assumptions, disparate patient populations, and heterogenous results.

IOM's best practices for guideline development called out a lack of "sufficient attention" to "the role of judgment in the derivation of recommendations," among other issues.¹⁵ It also listed factors that "commonly undermine the quality and trustworthiness of [clinical practice guidelines]" including:

variable quality of individual scientific studies; limitations in systematic reviews (SRs) upon which [clinical practice guidelines] are based; lack of transparency of development groups' methodologies (particularly with respect to evidence quality and strength of recommendation appraisals); failure to convene multi-stakeholder, multi-disciplinary guideline development groups, and corresponding non-reconciliation of conflicting guidelines; unmanaged conflicts of interest (COI); and overall failure to use rigorous methodologies in [clinical practice guidelines] development.¹⁶

¹⁴ Davino-Ramaya C et al., Transparency Matters: Kaiser Permanente's National Guideline Program Methodological Processes. *The Permanente Journal*. 2012;16(1): 55-62.

¹⁵ Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Institute of Medicine. 2011, fn 8, at p. 3.

¹⁶ *Ibid.*

Organizations who develop these guidelines ask their developers to use their judgment to balance different factors, including some that do not directly relate to ensuring the best patient care. For example, Kaiser Permanente uses a guidelines-development approach that includes “formulating recommendations based on quality of evidence, balance of benefits and harms, patient values and preferences, and resource and cost implications.”¹⁷ Thus, costs are considered alongside patient outcomes, and “patient values and preferences” are incorporated on a population, rather than an individual, level. The American College of Physicians uses a similar methodology.¹⁸

Another crucial issue is that there is not always a clear best choice in prevention, diagnostic, and treatment measures. The “scientific evidence about what to recommend is often lacking, misleading, or misinterpreted.”¹⁹ Experts may disagree about whether a particular practice guideline is the best approach to a health issue. For example, doctors have voiced concerns about a sepsis protocol mandated by the state of New York because it included a step that “may not be beneficial.”²⁰ In another example from 2014, clinical experts disagreed about whether major new cholesterol and blood pressure practice guidelines were appropriate. A group of 5 dissenters out of the 17 authors of the high blood pressure guidelines published an article against the new guidelines.²¹ In discussing these disagreements on high blood pressure guidelines, a *Journal of the American Medical Association (JAMA)* editorial stated that “guidelines should inform but not dictate, guide but not enforce, and support but not restrict.”²² Practice guidelines often must “harmonize” across differences in research studies. These judgment calls should be made at the bedside not in a corporate board room.

Finally, practice guidelines may not be trustworthy. The *JAMA* editorial cited above also stated:

Another risk is that the regulatory process may be used to advance commercial interests that may not be in the public interest. Clinical practice guidelines often make recommendations involving proprietary medical devices and pharmaceuticals. Device and pharmaceutical companies could lobby state governments to include these products in future regulations.²³

¹⁷ Davino-Ramaya 2012, fn 14, p. 60.

¹⁸ Qaseem A et al., The Development of Clinical Guidelines and Guidance Statements by the Clinical Guidelines Committee of the American College of Physicians: Update of Methods. *Ann Intern Med* 2019. doi:10.7326/M18-3290.

¹⁹ Woolf SH et al. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;318:527–30.

²⁰ Harris R. Are State Rules for Treating Sepsis Really Saving Lives? *National Public Radio*. May 30, 2017. <http://www.npr.org/sections/health-shots/2017/05/30/530224517/are-state-rules-for-treating-sepsis-really-saving-lives>. Accessed May 3, 2021.

²¹ Wright JT Jr, Fine LJ, Lackland DT, Ogedegbe G, Dennison Himmelfarb CR. Evidence supporting a systolic blood pressure goal of less than 150 mm Hg in patients aged 60 years or older: the minority view. *Annals of Internal Medicine*. doi:10.7326/M13-2981. Published online January 14, 2014. Cited in Krumholz 2014, fn 3.

²² Krumholz 2014, fn 3.

²³ Hershey TB, Kahn, JM. State Sepsis Mandates — A New Era for Regulation of Hospital Quality. *New England Journal of Medicine*. May 21, 2017. doi: 10.1056/NEJMp1611928. <http://www.nejm.org/doi/pdf/10.1056/NEJMp1611928>. See also: Lenzer, J. Why We Can't Trust Clinical Guidelines. *BMJ* 2013;346:f3830 doi: <https://doi.org/10.1136/bmj.f3830>.

This risk has been born out in practice. In the early 1990s, the Cochrane Collaboration put out guidelines making high-dose steroids the standard of care for acute spinal cord injury. The sole reviewer of the guidelines was a consultant to steroid manufacturers. Despite skepticism from neurosurgeons in polls, the guidance was widely followed until it was reversed in March 2013 by new Congress of Neurological Surgeons guidelines that found no good evidence for use of the steroids and substantial evidence of harmful side effects including death.²⁴ This is not an exceptional case. One survey found 71% of clinical policy committee chairs and 90.5% of co-chairs had financial conflicts.²⁵ While financial conflicts create a different kind of bias than racial or ethnic discrimination, these examples show that judgment calls made by authors of guidelines can twist patient outcomes in serious ways.

The problem of unreliable guidelines is magnified when treatment and diagnosis guidelines make different recommendations based on race and ethnicity. Claims about racial and ethnic differences in medical needs are often based on poor quality evidence. Evidence shows that developers of clinical practice guidelines do not give racial differences the necessary level of scrutiny. One review of the use of race correction in clinical algorithms found that when algorithm developers offer rationales for why race correction is included, their origins can be traced “to outdated, suspect racial science or to biased data.”²⁶ Other developers offer no explanation for why racial or ethnic differences may exist. When designing the algorithm, they choose to translate correlations between race and outcome into different clinical treatment without understanding why those correlations exist or what they really represent.²⁷ As shown above, both available data and expert judgment can be affected by the structural racism in society and the health care system.

It is essential that the use of race or ethnicity in clinical algorithms is scrutinized, including whether race or ethnicity are serving as proxies for other factors that should be identified explicitly. Studies must also look at racial differences in recommendations made by algorithms that do not explicitly incorporate race, and in their ultimate patient outcomes. However, it will not be possible to eliminate the use of judgment or the need for individual assessment in care decisions. These judgments should be made at the bedside between the patient and their health care provider, not by a committee based on population-level data. Moreover, the underlying research and the process for developing an algorithm must be available to the health care professionals providing patient care.

Encoding biases into algorithms and taking health care decisions away from patients and providers has an impact on quality of care, clinical outcomes, quality of life, and health disparities, RFI question 1, and is one mechanism by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color, RFI question 6. NNU urges AHRQ to investigate

²⁴ See Lenzer 2013, fn 23.

²⁵ *Ibid.*

²⁶ Vyas DA, Eisenstein LG, Jones DS. Hidden in plain Sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med.* 2020 Aug 27;383(9):874-82. doi: 10.1056/NEJMms2004740. PMID: 32853499. <https://www.nejm.org/doi/full/10.1056/NEJMms2004740>.

²⁷ *Ibid.*

the role judgment and bias plays in the development of these algorithms, how to minimize the effects of that bias, and how to ensure that final judgment calls are made at the bedside by fully informed clinicians and patients.

II. Algorithms developed through machine learning

Answering RFI question 1 on the estimated impact of these algorithms on quality of care, RFI question 6 on the mechanisms by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color, and RFI question 7 on clinician and patient awareness of bias.

While many clinical algorithms reflect decisions made by human experts, such as clinical practice guidelines, an increasing number of clinical algorithms are driven by machine learning. These algorithms pose several risks. Clinical algorithms developed through machine learning can make serious errors, amplify patterns of bias in the underlying data, and endanger population subgroups. They are also often opaque to users. This can be because they are “black boxes,” where even their creators do not know how they work, because of protections for proprietary trade secrets, or because they require high level technical knowledge to understand. This opacity undermines patient and professional trust and magnifies the risk to patients from errors. Thus, it is essential that these algorithms are made transparent and intelligible to their users and to ensure that health care professionals can override them as needed to provide their patients therapeutic and effective care that is consistent with their needs and preferences.

A. Machine learning algorithms may be trained on datasets that do not apply to racial and ethnic subpopulations, leading to bias.

Machine learning algorithms, often referred to as artificial intelligence (AI), are programmed to find patterns in large quantities of data. They are trained on one set of data and then used to classify new information based on patterns they detected in the training data. They are frequently used to make diagnoses or to make predictions based on information in electronic health records and assign risk levels. Problematically, data from electronic health records often reflect differential access to health care, quality of care, and other forms of structural racism and biases. Similarly, Black, Indigenous, and other people of color face discriminatory assessment and treatment by health care providers because of implicit bias or false beliefs about race which are entered into their electronic health record and used to train clinical algorithms.

For example, in one recent study, researchers found that models trained on electronic health record data to predict suicide attempts among patients who had outpatient mental health visits performed substantially worse for Black patients, American Indian/Alaskan Native patients, and patients without ethnicity recorded than it did for white, Asian, or Hispanic patients.²⁸ There were

²⁸ Coley RY, Johnson E, Simon GE, Cruz M, Shortreed SM. Racial/Ethnic Disparities in the Performance of Prediction Models for Death by Suicide After Mental Health Visits. *JAMA Psychiatry*. Published online April 28, 2021. doi:10.1001/jamapsychiatry.2021.0493.

several reasons for this failure. One was a more limited pool of data on Black and American Indian/Alaskan Native patients and patients without ethnicity recorded. The authors speculated that systemic barriers to affordable, culturally competent mental health care led to lower utilization and therefore sparser records on racial or ethnic minority populations.²⁹ Practitioner bias and institutionalized discrimination throughout the treatment process means that Black, Indigenous, and other people of color are less likely to be screened or receive high-quality treatment for depression and more likely to experience discrimination in health care settings.³⁰ The researchers also pointed to evidence that discrimination and low quality care deters health care use: studies showing Black and Asian people were less likely than whites to initiate and continue mental health treatment,³¹ and that non-Hispanic Black respondents who had experienced discrimination in health care settings were less likely to opt for talk therapy over medication.³² The authors also suggested that suicide deaths may be misclassified more often among some groups than others. Setting specific thresholds for intervention for each measured race and ethnic group improved performance somewhat for some groups poorly served by the global thresholds but not for others. In any case, setting race and ethnicity-based thresholds to account for poor predictivity meant that Black and American Indian/Alaskan Native patients would be subject to “unnecessary and possibly intrusive interventions” at a higher rate for the same number of interventions in eventual suicides compared to white patients.³³

Numerous studies have documented implicit bias in health care providers which may then be recorded in electronic health records used to train algorithms. For example, clinicians are less likely to prescribe pain medications to Black patients than white patients. In one study, fewer opioids were prescribed for Black patients with migraines and back pain, which rely on self-reported pain, than similar white patients, but there was no difference for patients with long bone fractures, which are visible on an X-ray.³⁴ In another study, Black children with appendicitis in emergency departments were less likely than white children to receive any pain medication for moderate pain and less likely to receive opioids for severe pain.³⁵ These studies demonstrate that Black patients suffer because their doctors do not trust them with opioid prescriptions or believe them when they say they are in pain. This may be due, in part, to the fact that substantial numbers of white medical students and residents falsely believe that Black people feel less pain than white

²⁹ *Ibid.* at p. E6.

³⁰ *Ibid.*

³¹ Hines AL, Cooper LA, Shi L. Racial and ethnic differences in mental healthcare utilization consistent with potentially effective care: the role of patient preferences. *Gen Hosp Psychiatry*. 2017; 46:14-19. doi:10.1016/j.genhosppsy.2017.02.002.

³² Sonik RA, Creedon TB, Progovac AM, et al; Health Equity Consortium. Depression treatment preferences by race/ethnicity and gender and associations between past healthcare discrimination experiences and present preferences in a nationally representative sample. *Soc Sci Med*. 2020;253:112939. doi:10.1016/j.socscimed.2020.112939.

³³ *Ibid.* at p. E5.

³⁴ Tamayo-Sarver JH, Hinze SW, Cydulka RK, Baker DW. Racial and ethnic disparities in emergency department analgesic prescription. *Am J Public Health*. 2003 Dec;93(12):2067-73. doi: 10.2105/ajph.93.12.2067.

³⁵ Goyal MK, Kuppermann N, Cleary SD, Teach SJ, Chamberlain JM. Racial disparities in pain management of children with appendicitis in emergency departments. *JAMA Pediatr*. 2015;169(11):996–1002. doi:10.1001/jamapediatrics.2015.1915.

people.³⁶ That belief carried over to rating Black patients' pain as lower and making less accurate treatment recommendations. Similarly, another study found that physicians showed implicit bias favoring white patients over Black patients that could contribute to racial and ethnic disparities in use of thrombolysis for myocardial infarction.³⁷

These examples, and there are many others, demonstrate that racially biased behavior by clinicians is a serious problem that should be addressed through changes in training and education as well as through increased staff diversity. It is also a problem in the so-called evidence base of medical data. If an algorithm were to be created to regulate opioid prescriptions and based on health records data, it could easily end up encoding anti-Black bias and hiding it behind a veneer of objective technology.³⁸

Finally, algorithms may exhibit racial and ethnic bias whether or not race and ethnicity have been explicitly factored in as inputs. For example, in 2019, researchers Obermeyer et al. found that an algorithm that did not use race as an input still ended up disadvantaging Black patients in its recommendations because it assumed that patients who use less health care are less sick, thereby making health inequities self-reinforcing.³⁹ Obermeyer et al. analyzed the recommendations made by an algorithm used by large health systems and payers to target patients for "high-risk care management" programs and found that it assigned sicker Black patients the same level of risk as less-sick white patients.⁴⁰ This bias reduced the number of Black patients identified for extra care by more than half. This occurred because the algorithm was using health costs as a proxy for health needs. Structural racism has ensured that Black people in the U.S. have lower access to health care, which means they have both higher health needs and lower health costs.⁴¹

Discrimination at many different stages of the patient experience introduces bias into health records data and therefore into any clinical algorithm that depends on it. Reduced access to health treatment is pervasive among Black, Indigenous, and other people of color. These groups face higher rates of many diseases and premature death, including infant mortality, than white people.⁴²

³⁶ Hoffman, Kelly M. et al. Racial bias in pain assessment. *Proceedings of the National Academy of Sciences* Apr 2016, 113 (16) 4296-4301; DOI: 10.1073/pnas.1516047113.

³⁷ Green AR, Carney DR, Pallin DJ, et al. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J Gen Intern Med.* 2007;22(9):1231-1238. doi:10.1007/s11606-007-0258-5.

³⁸ NNU recognizes that analysis of electronic health records, whether by human beings or artificial intelligence programs, could expose implicit biases. However, this does not invalidate the need for clinical algorithms to be transparent and intelligible to their users nor the need to ensure that health care professionals have the autonomy to override them if it is in their patients' interest.

³⁹ See Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019 Oct 25;366(6464):447-53. doi: 10.1126/science.aax2342. PMID: 31649194. <https://science.sciencemag.org/content/366/6464/447>.

⁴⁰ *Ibid.*

⁴¹ See citations in footnote 9, above.

⁴² National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Committee on Community-Based Solutions to Promote Health Equity in the United States; Baciu A, Negussie Y, Geller A, et al., editors. *Communities in Action: Pathways to Health*

This is due to social, not biological, differences. People of color are more likely to lack health insurance than white people, usually due to affordability.⁴³ Locations with majority Black and Hispanic populations are more likely to face primary care physician shortages.⁴⁴ Additionally, underserved areas Hospitals in Black neighborhoods may have fewer specialists than those in white neighborhoods.⁴⁵ These and many other disparities add up to less and lower-quality care, which means less representation in health data. When embedded in electronic health records used to train clinical algorithms, these algorithms reinforce and reproduce both structural racism and implicit clinician bias.

B. Machine learning algorithms may draw inferences from data that are not relevant to health care, potentially introducing racial bias.

One of the most disturbing problems with machine learning is that it may draw spurious inferences based on data that is not relevant to patient care.⁴⁶ Moreover, because many algorithms are opaque, as discussed in the next section, it is not always clear what information an algorithm is using to make its categorizations.

For example, a group of Stanford computer science students, in consultation with radiologists, used a set of over 100,000 chest X-rays tagged with diagnoses to teach a program to read similar X-rays for signs of tuberculosis.⁴⁷ The algorithm appeared to reach the correct diagnosis 75% of the time. Tuberculosis can be a challenging diagnosis for doctors in South Africa, where it is prevalent, so the project hoped to combine machine and human expertise for a better combined success rate.⁴⁸ To make sure it was working correctly, the team designed the program to highlight the parts of the image it was using to make the diagnosis and shared it with others in the field to critique. Fortunately, a medical resident discovered that the X-ray analysis program was basing its diagnosis in part on the information on the edge of the scan image that showed the type of machine used to take the scan. If the machine was of the portable type used in hospitals, instead of the type used in doctors' offices, the diagnosis was more likely to be tuberculosis. The machine had found a pattern on the images, but it was not in the medically relevant data that it was intended to analyze. In a system where both certain diseases and the ability to access care in certain

Equity. Washington (DC): National Academies Press (US); 2017 Jan 11. 2. The State of Health Disparities in the United States. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK425844/>.

⁴³ Artiga S, Orgera K, Damico A. Changes in Health Coverage by Race and Ethnicity since the ACA, 2010-2018. *KFF*. March 5, 2020. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/changes-in-health-coverage-by-race-and-ethnicity-since-the-aca-2010-2018/>. Accessed May 3, 2021.

⁴⁴ Gaskin DJ, Dinwiddie GY, Chan KS, McCleary RR. Residential segregation and the availability of primary care physicians. *Health Serv Res*. 2012;47(6):2353-2376.

⁴⁵ Pallok 2019, fn 9.

⁴⁶ Paullada A, et al. 2020. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *arXiv preprint arXiv:2012.05345*. Accessed May 10, 2021.

⁴⁷ See Harris, R. How can doctors be sure a self-taught computer is making the right diagnosis? *NPR*. April 1, 2019. <https://www.npr.org/sections/health-shots/2019/04/01/708085617/how-can-doctors-be-sure-a-self-taught-computer-is-making-the-right-diagnosis>. Accessed April 29, 2021.

⁴⁸ *Ibid*.

venues is highly correlated with poverty and race, this type of unexpected behavior has the potential to introduce bias.

Machine learning algorithms consistently make more obvious errors, leading experts in the field to insist on the importance of a human override option. If an algorithm continues learning from new data while operating, its results can become truly strange. Machine learning algorithms can be easily fooled—Google image software identified a picture of a cat as guacamole after MIT students changed a few pixels.⁴⁹ To explain the necessity of a human override, one public health expert compared black box algorithms to a plane where autopilot said the plane was going up and pilots saw it was going down but had no override option.⁵⁰

Inappropriate inferences from data have a negative impact on quality of care, clinical outcomes, quality of life, and health disparities, RFI question 1, and are one mechanism by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color, RFI question 6.

C. Machine learning algorithms can be dangerously opaque making it difficult to identify racial bias in their development or underlying data.

There are several reasons that machine learning algorithms are frequently opaque, but none of them are inevitable. If developers of these algorithms and the health care providers who implement them commit to transparency principles, many systems can be made transparent. If they cannot be made comprehensible to users and patients, they generally should not be used for medical applications.

Some machine learning algorithms function as “black boxes.” Their creators can show that they categorize input data correctly some percentage of the time but cannot explain the reasoning behind the categorizations. The complexity of their reasoning and the number of data points they use to reach conclusions may be difficult for a human to comprehend. They may use a complex form of mathematical representation which is not intelligible for humans.⁵¹ An algorithm that is designed to be comprehensible might be less useful, if the goal is to interpret data at a higher level of complexity than a human expert can.⁵² Moreover, the decision logic may change over time as the program learns.

The black box is an easier design choice to program, but it is often not the only or the best option. Some programmers create black boxes that give hints about their reasoning, such as the X-ray image highlighting that allowed researchers to spot flaws in the tuberculosis diagnosis

⁴⁹ See Powell, A. AI revolution in medicine, *Harvard Gazette*. November 11, 2020. <https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/>.

⁵⁰ *Ibid.*

⁵¹ See Burrell J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*. January-June 2016:1-12.doi:10.1177/2053951715622512. p. 5-9.

⁵² See *Ibid.* at p. 5.

algorithm. However, these clues do not give a full understanding of what an algorithm is doing and may provide a false sense of security to users.⁵³ The need for this opacity may be overstated. One computer scientist, Cynthia Rudin, said “I’ve worked on many predictive modeling problems... and I’ve never seen a high-stakes decision where you couldn’t come up with an equally accurate model with something that’s transparent, something that’s interpretable.” She explains that medical decisions with life-or-death consequences merit the extra effort to build a program based on clinical knowledge that allows humans to see how it reaches its conclusions.⁵⁴

Creators of algorithms can also avoid transparency to protect their trade secrets. Avoiding sharing information with competitors may be an incentive to design a black-box model or to keep the information about an algorithm restricted inside the corporation that owns it.⁵⁵ One observer of financial algorithms argued that corporations also choose to keep algorithms opaque to hide “sidestepped regulations, the manipulation of consumers, and/or patterns of discrimination.”⁵⁶ Developers of certain algorithms may also keep them secret to prevent users from changing their behavior to achieve a particular result from an algorithm.⁵⁷ For instance, a search engine may keep its inner workings secret to prevent websites from gaming the system.

Sometimes algorithms have information about how they work publicly available, but that information can only be understood by people with a high level of technical knowledge. Most medical professionals, patients, and journalists do not have the appropriate education or the time to read computer code to understand the algorithms governing care.⁵⁸

If a system is opaque, then researchers, users, and patients are limited in their ability to detect and counteract biases. As shown above, there is a serious risk that these algorithms will have biases. Biased data or flawed reasoning can be hidden behind a veneer of objective technology if a machine learning algorithm cannot or does not show its work in a way users and patients can understand. This can have devastating results for individual patients and increase overall health care inequities.

Transparency can and should be required, however. It is essential that corporations make their code available to the public, along with appropriate documentation, and that they explain how the algorithm functions so that clinicians can make informed decisions and explain them to their patients. Some types of algorithms will not be appropriate for medical use, despite some promising features, because they cannot be made intelligible to their users. Experiments to check for discriminatory outputs are essential but do not obviate the need for transparency. These measures are essential to safe and trustworthy implementation of this technology.

⁵³ See *Ibid.* at p. 9; Harris 2019, fn 47.

⁵⁴ Harris 2019, fn 47. See also Rudin, C. Stop explaining black box learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*. 2019;1:206-215. <https://arxiv.org/abs/1811.10154>.

⁵⁵ See Harris 2019, fn 47.

⁵⁶ See Burrell 2016, fn 51, at p. 4, summarizing arguments from Pasquale, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. 2015.

⁵⁷ Burrell 2016, fn 51, at p. 4.

⁵⁸ *Ibid.*

Users and patients are often unaware of the inclusion of variables that can introduce bias in clinical algorithms, RFI question 7, due to the opacity of machine learning algorithms. NNU urges AHRQ to investigate how machine learning algorithms produce errors and perpetuate bias as well as support both their full transparency and the autonomy of clinicians to fully exercise their professional judgment. NNU supports regulation of clinical algorithms to ensure results are unbiased and reasoning is transparent.

III. Preliminary results from an NNU survey on clinical algorithms suggest that RNs are often prohibited from overriding algorithms, preventing them from correcting mistakes made by biased algorithms.

Answering RFI question 1 on the estimated impact of these algorithms on quality of care and RFI question 6 on the mechanisms by which use of algorithms contribute to poor care for Black, Indigenous, and other people of color.

Clinical algorithms and technology could serve as tools that assist health care professionals in making decisions about patient care, in consultation with the patient and in the patient's best interests. In practice, however, nurses and other health care professionals are often compelled by their employers, or the health system in which they practice, to accept the recommendations of these clinical algorithms without the autonomy to override them regardless of whether the recommendation is appropriate for and in the best interests of individual patients.

Clinical guidance is designed to be used in concert with direct care professionals' expertise to find the best treatment for an individual patient. In practice, however, employer rules discourage or prohibit the use of autonomous professional judgment and require adherence to decisions made by clinical algorithms. This limits the ability of health care professionals to counteract biased or inappropriate algorithmic recommendations. Addressing bias in algorithms is necessary but not sufficient to ensure safe, high-quality therapeutic care that meets the needs of an individual patient. Every patient is different, even those who share many demographic and medical history commonalities, and algorithms cannot account for all the relevant differences among patients. To ensure appropriate care is provided, health care professionals must have the right to override algorithms to account for individual needs and preferences.

As discussed above in Section IIA, Obermeyer et al. discovered that an algorithm that was supposed to identify patients in need of extra care failed to identify large numbers of Black patients because it relied on health costs as a proxy for health needs. Optum, the company responsible for the biased algorithm, responded to the research with a statement calling it "misleading" because "[t]he cost model is just one of many data elements intended to be used to select patients for clinical engagement programs, including, most importantly, the doctor's expertise."⁵⁹ Academics and

⁵⁹ See Ledford, Heidi. Millions of black people affected by racial bias in health-care algorithms. *Nature*. Published online October 26, 2019. <https://www.nature.com/articles/d41586-019-03228-6>.

developers of diagnostic algorithms driven by machine learning frequently tout their effectiveness when combined with users' expertise but stress the importance of an override option.⁶⁰ An FDA report on health information technology stated that clinical decision support software "is not intended to replace clinicians' judgment, but rather to assist clinicians in making timely, informed, higher quality decisions."⁶¹ Clinical algorithm developers may explicitly state that they are not a substitute for clinicians' professional judgment and that clinical decisions must consider the characteristics of individual patients, but in practice this is not always the case.⁶²

NNU's members are bedside registered nurses who often face pressure from management and threats of discipline if they deviate from clinical algorithms. Clinical algorithms often are used to reduce the time RNs spend with patients and limit the amount of care they offer them. In mid-April 2021, NNU began surveying RNs and other health care workers, both members and non-members, on their experiences with clinical algorithms in their workplaces, including the use of race and ethnicity as inputs, their awareness of possible bias, and their ability to override recommendations. The survey was initially administered in continuing education classes but is now widely available online. This RFI response will focus on the preliminary results from 170 registered nurses. These preliminary results offer evidence of a negative impact of these algorithms on quality of care, RFI question 1, and of one of the mechanisms by which the use of algorithms contributes to poor care for Black, Indigenous, and other people of color, RFI question 6.

Out of 142 nurses who responded to a question regarding overriding algorithms, 36 respondents (25%) said they were not allowed to "override clinical practice guidelines, clinical pathways, or electronic or computer-based tools that [they] believe are not in the best interest of the patient", 32 respondents (23%) could only do so with the approval of a doctor or supervisor, and 45 respondents (32%) did not even know if they were allowed to override the recommendations. Only 13 respondents (9%) were allowed to override algorithms based on their own judgment. 16 respondents (11%) said the question was "not applicable."

The fact that a majority of registered nurses who responded cannot override or do not know if they can override algorithms is concerning because 57 (40%) said they had been "been prompted by a clinical practice guideline, clinical pathway, or electronic or computer-based tool to make choices about patient care, patient care staffing, or other clinical issues that [they] believed were not in the best interest of the patient based on [their] clinical judgment and scope of practice" Out of those nurses, 18 (32%) could not override the algorithms at all, while 8 (14%) said they did not know. Twenty-two (39%) needed approval from a doctor or supervisor to override. Only 9 (16%) of RN respondents could override based on their professional judgment without approval.

⁶⁰ E.g., Powell 2020, fn 49; Powell A. The algorithm will see you now. *Harvard Gazette*. February 28, 2019. <https://news.harvard.edu/gazette/story/2019/02/in-health-care-ai-offers-promise-and-hype/>.

⁶¹ Food and Drug Administration Safety Innovation Act (FDASIA) Health IT report, Proposed strategy and recommendations for a risk-based framework. April 2014. <https://www.fda.gov/media/87886/download>. p. 26.

⁶² See Krumholz 2014, fn 3.

The need to combat inappropriate recommendations creates deficiencies in care. Worryingly, one RN responded:

[Her employer] calls them BPA's (best practice alerts) that pop up in the middle of your charting. Sometimes they are useful reminders, but often it leads to nurses following the BPAs instead of exercising their clinical judgment to individualize each patient's needs. I have been told by an educator "you are not supposed to think or question it. Just do what the computer tells you to."

Another RN commented:

Healthcare should not be one standard protocol for all but that is what it's turned into. We are not being asked to use our skills, knowledge, or critical thinking abilities. We are now just asked to follow protocols, policies and procedures. It's disheartening and disappointing because healthcare is becoming more about financial gains & not actual personalized health for each individual as a whole.

Doctors also feel substantial pressure to comply with treatment guidelines prescribed by clinical algorithms. Polls of doctors show that they frequently will follow guidelines that become the "standard of care" even if they do not agree with the scientific conclusions that underlie them, due to fears that they will be penalized for variations from that standard.⁶³ External payers such as insurance companies and government agencies also reward or penalize individual doctors and medical groups based on compliance with guidelines.⁶⁴ Payers apply these same rewards and penalties to hospitals and other types of health care facilities. Additionally, the health systems implementing clinical algorithms are often responding to promises by manufacturers that they will be able to reduce staffing costs or to financial incentives to implement electronic health records systems, rather than strong data on patient outcomes.

Many discussions of clinical algorithms mention excessive health care costs as a driving factor.⁶⁵ Achieving the "efficiency" gains promised by clinical algorithms requires forcing health care professionals to comply with their recommendations, despite possible weaknesses in the tools and the need for individualized patient care. When use of clinical algorithms is heavily incentivized, patients who fail to fit those guidelines and expectations suffer. In 2009, an assessment of a pay-for-performance plan in California found that doctors dropped noncompliant patients and refused to treat people with complicated illnesses to prevent bad outcomes from being added to their statistical records.⁶⁶

⁶³ See Lenzer 2013, fn 23.

⁶⁴ See Hamm, Nagykaldi 2018, fn 7.

⁶⁵ See, e.g., Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Institute of Medicine 2011, fn 8. See also Woolf 1999, fn 19.

⁶⁶ Groopman J, Hartzband P. Why 'quality' care is dangerous. *Wallstreet Journal*. April 8, 2009. <https://www.wsj.com/articles/SB123914878625199185>.

The literature cited by AHRQ in the RFI includes numerous examples of systemically biased algorithms. It is likely that these studies on racial bias in algorithms represent only a small sampling of the biased algorithms in use, due to obstacles to research on proprietary algorithms. As discussed above, bias is also introduced by the ways in which population-level data can be a poor fit for individual patients. When algorithms make biased decisions on patient care based on race or ethnicity, registered nurses and other health care professionals have the expertise and connection to patients to find the appropriate course of action. These health care professionals can counteract problems in algorithms if they are given the autonomy to do so.

For these reasons, clinical algorithms must serve as guidelines, not rigid protocols. There may be good reasons to deviate from a recommendation in a manner that has been neither authorized nor anticipated. Health care professionals must be free to exercise their professional judgment regarding the care that is appropriate for each patient based on the patient's particular clinical indications, circumstances, needs, and preferences.

It is also worth noting that these algorithms have been introduced into health care settings to manage and take advantage of the large quantities of data stored about patients through electronic health records systems and diagnostic imaging computers. As we have seen, this means their performance is limited by the quality of data available. This data is often entered by staff, under pressure from their employers as well as time constraints, tasked with recording a tremendous amount of information for billing and insurance purposes as well as inputs for many different algorithms. That means there will always be some mistakes in data pulled from electronic health records. It also means that the more algorithms health care facilities choose to implement, the more time a health care provider must spend facing a computer instead of their patient. This may be particularly problematic for patients who have limited time with their health care providers at under-resourced facilities and who have difficulty trusting health care providers due to previous experiences with racism.

NNU supports laws and regulations, such as AB 858, a bill introduced in California's state legislature in its 2021-2022 legislative session, that promote patient safety and reduce the risk from biased algorithms by allowing health care professional who provide patient care to exercise independent professional judgment to override health information technology, requiring employers to train workers on the use and limitations of new technology, and requiring consultation with workers before adopting new technology.

As AHRQ conducts its inquiry into bias in clinical algorithms, NNU urges the examination of the ability of health care professionals to use their independent professional judgment to override these algorithms and the role override can play in combatting bias.

IV. Lack of transparency in algorithms and lack of user education on bias impedes use of professional judgment and makes it difficult for health care professionals to detect bias.

Answering RFI questions 7 and 11 on user awareness and education.

In the experience of survey respondents, RNs are rarely aware of the inclusion of race, ethnicity or other variables that could introduce bias in these algorithms and the implications for clinical decision making. Patients are not aware of how these algorithms may bias their care and providers do not communicate this information to patients. This experience is corroborated by extensive evidence on the ways that proprietary algorithms lack the transparency necessary to allow users to evaluate their recommendations.

Opacity in proprietary systems means that health care professionals often do not know how the algorithms reach recommendations for patient care.⁶⁷ As discussed in Section IIC, the results from electronic systems trained using machine learning can be opaque even to their creators without techniques to ensure transparency. Health care professionals enter a wide range of information into electronic health records systems, some of which informs clinical algorithms. They are often asked to enter race or ethnicity or other information that can in some contexts serve as a proxy for race, such as income, location, or health care usage history. They later see treatment recommendations pop up on screen. They will often not know what information they entered caused that recommendation or how the algorithm reached that conclusion.

Registered nurses are often required to use clinical algorithms but receive minimal education about how algorithms are developed or may influence clinical decision-making. This impedes RN's ability to effectively assess the applicability of algorithmic recommendations in their clinical practice and impedes their ability to recognize where population-based and possibly biased data does not apply to an individual patient.

NNU's survey of RNs on clinical algorithms asked questions about the extent to which users and patients are aware of the inclusion of variables that can introduce bias into algorithms, RFI question 7, and the education available on these algorithms, RFI question 11. The preliminary results show that many registered nurses and patients do not know about the risk of bias in algorithms. As of May 10, 74 out of 168 (44%) registered nurses responding responded that they were not aware that clinical algorithms that include race or ethnicity as a factor may introduce bias into patient care. Fifty-five nurses (33%) responded that they were aware and 39 (23%) said they were somewhat aware.

Even among registered nurses who did know about the potential for bias, this awareness did not come from education and training provided by their employers. When asked "Has your workplace provided any education about how clinical algorithms could introduce bias into patient care?", an overwhelming majority of 153 out of 168 registered nurses (91%) responded "No, I have not received any education about this." Only three nurses surveyed said "Yes, I have received extensive education about this."

⁶⁷ See Ledford 2019, fn 59.

Nurses are not the only ones who are not being informed about the risks of bias in algorithms. When asked if, in their experience, patients were “informed about the use of race, ethnicity, or other factors that could result in bias in algorithms that influence their care,” the majority of registered nurses, 98 out of 167 respondents (59%) said “no.” Five nurses replied “rarely,” five replied “sometimes” while only one nurse responded “always.” 58 nurses (35%) did not know.

When health care providers cannot see explanations for individual decisions recommended by a clinical algorithm, dangerous errors based on bias can persist without being spotted, preventing users from advocating for the fair treatment of their patients. This opacity impedes the ability of registered nurses and other health care professionals to effectively and appropriately incorporate the scientific information represented in the algorithms into their own clinical decisions based on their individual patients. Our members live in the communities they serve and see patients every day. They have important information about how their individual patient or community may differ from wider data trends. They cannot use that information to counteract biases in the data if they do not know what data is included or how it is used.

This is not just a provider and patient education problem. Often, even researchers do not have access to data on proprietary algorithms. Protections for proprietary tools combined with the sensitivity of health care data make comprehensive studies of racial and ethnic bias in recommendations and patient outcomes rare.⁶⁸ This means that there is limited evidence available for health care professionals who do know about the possibility of bias to review the performance of the particular algorithms they use in their practice.

In our current market-driven health care system, there is often financial pressure to implement such systems at facilities that serve low-income patients who may be both sicker and less able to pay medical bills. Thus, lower-income communities, that are disproportionately communities of color, and their health care providers may be subject to opaque performance systems to determine what resources they receive. This is also a form of bias in algorithmic implementation.

NNU supports measures to increase the transparency of clinical algorithms and to make information on how these algorithms work accessible to health care providers. Before an algorithm is implemented, rules, content sources, and other relevant information should be readily available to health care professionals and posted on a public website. Crucially, health care professionals, patients, and the public should be educated on the inapplicability of population statistics to individuals and clinicians should have, and be made aware of, the right to override technology in the interests of their patients based on their professional judgment.

AHRQ should include in its analysis the extent to which these algorithms are transparent about their inputs, reasoning, and potential for biased recommendations, and the training and

⁶⁸ Ledford 2019, fn 59.

education provided to the health care professionals who use these tools and the patients whose care is affected by them.

V. Conclusion

As AHRQ conducts its inquiry into how clinical algorithms can introduce racial and ethnic bias into health care delivery, there are several essential points to consider that apply to all algorithms, not just those that have already been proven to be biased. First, there are fundamental limits on the ability of algorithms to meet the needs of individual patients, especially when those patients are part of racial or ethnic groups that are less well represented in the data. Second, AHRQ must consider the importance of ensuring that health care professionals have full autonomy to override algorithmic recommendations, within their scope of practice and consistent with their patients' needs and preferences. Third, transparency and education about all aspects of the creation, use, and impact of these algorithms is necessary to detect and combat bias. Fourth, transparency must extend to clinicians, patients, and the public, including thresholds for recommendations to provide and deny preventive, diagnostic, or treatment measures by race, ethnicity, gender, and other relevant data. Finally, NNU urges AHRQ to support robust regulation of clinical algorithms as medical devices and to ensure that the supporting evidence and reasoning on which clinical recommendations are based are sufficient, sound, transparent, and intelligible and that recommendation thresholds for providing or denying preventive, diagnostic, or treatment measures are transparent to clinicians, patients, and the public.

Thank you for considering National Nurses United's comments and materials in response to the Request for Information (RFI) on "Use of Clinical Algorithms That Have the Potential To Introduce Racial/Ethnic Bias Into Healthcare Delivery," 86 Fed. Reg. 12,948 (Mar. 5, 2021).

Sincerely,



Carmen Comsti
Lead Regulatory Policy Specialist
National Nurses United