Calculated Risks

Ensuring Accuracy and Equity in Hospital Resource Allocation Software

Kat Hefter

WISE
IEEE

Preface and Acknowledgements

About the Author

Kat Hefter is a joint JD/PhD student at the University of Pennsylvania located in Philadelphia, Pennsylvania. She graduated in 2020 from Duke University with a B.S.E. in biomedical engineering, electrical/computer engineering, and neuroscience. On campus, Kat is on the leadership board for the Disabled and Allied Law Students Association, the STEM Club, the Journal of Law and Innovation, and the Civil Rights Project pro bono organization. Kat is also a member of the Tau Beta Pi and Eta Kappa Nu Honor Societies, and the Institute for Electrical and Electronics Engineers (IEEE). Upon graduating from the University of Pennsylvania, Kat hopes to pursue a career at the intersection of medicine, technology, and law.

About the Wise Program

The Washington Internships for Students of Engineering (WISE) program was founded in 1980 through the collaborative efforts of several professional engineering societies to encourage engineering students to contribute to issues at the intersection of science, technology, and public policy. The program helps students to gain exposure to the legislative and regulatory policy-making process through meetings with leaders in federal agencies, congressional staffers, and advocacy groups. In addition, each student is responsible for independently researching, writing, and presenting a paper on a topical engineering-related public policy issue that is important to the sponsoring society. For more information about the WISE program, visit www.wise-intern.org [1].

About IEEE

IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. Its more than 396,000 members in over 160 countries work to inspire a global community to innovate for a better tomorrow [2].

Acknowledgements

I would like to thank IEEE for providing me with this opportunity, and Erica Wissolik and Marc Canellas for being great mentors and resources for me this summer. I also would like to thank Mark Ames, this year's Faculty Member in Residence (FMR) for his mentorship and support. Thank you to AIChE for supporting this program this year, and thank you to my fellow interns for your wisdom, ideas, and friendship.

Front Cover Image Credits: Adapted from [3].

Acronyms

AAPCHO – Association of Asian Pacific Community Health Organizations
ACEP – American College of Emergency Physicians
ACG – Adjusted Clinical Group
AHRQ – Agency for Healthcare Research and Quality
AI – Artificial intelligence
CFPB – Consumer Financial Protection Bureau
DHS – U.S. Department of Homeland Security
ED – Emergency Department
EEOC – Equal Opportunity Employment Association
EHR – Electronic Health Record
ENA – Emergency Nurses Association
ER – Emergency Room
ESI – Emergency Severity Index
FCC – Federal Communications Commission
FDA –U.S. Food and Drug Administration
HHS – U.S. Department of Health and Human Services
HUD – U.S. Department of Housing and Urban Development
ICU – Intensive Care Unit
IEEE – Institute of Electrical and Electronics Engineers
IOM – Institute of Medicine
ML – Machine learning
MRAS – Medical resource allocation systems
NACHC – National Association of Community Health Centers
NIST – National Institute of Standards and Technology
NCHS – National Center for Health Statistics
OMB – Office of Management and Budget
OPCA – Oregon Primary Care Association
SDOH – Social determinants of health
US – United States
VA – U.S. Department of Veterans Affairs

Contents

Executive Summary

COVID-19 demonstrated to many Americans just how tenuous our medical system is, but the medical system's struggles to adapt to the pandemic are only one symptom of a crisis of medical resource allocation that has been building for decades. Due to heavy demand on the emergency departments (EDs), and often insufficient supplies of space and medication, ED overcrowding requires physicians to determine who gets care, and when [4]. More broadly, doctors must find ways to determine who receives scarce resources like ventilators or organs, or identify patients who would benefit most from extra attention and preventative care.

Unfortunately, these allocation techniques cannot escape medicine's long history of inequity. Study after study has shown that a person's judgment is affected by the order in which they receive information, their desire to resolve situations as quickly as possible, and their preconceived opinions about a given situation – health care practitioners are no exception [5]. As a result, assessments about how to allocate care are influenced by the people making the decisions, resulting in health care inequities. To combat this, hospitals are increasingly turning to machine learning to create tools that can learn patterns from vast amounts of data and use them to generate standardized, impartial predictions of future outcomes [6].

However, left untested, machine learning devices often incorporate and even solidify the same kinds of systemic biases and errors they were meant to prevent. Biases have been found in machine learning tools across a variety of disciplines, including hospital resource allocation [7]. Furthermore, because these tools cannot be patented due to general restrictions on patenting algorithms, companies employ trade secret protections to safeguard the intellectual property (IP) rights of data and methods. The FDA regulates most medical devices, including those that use machine learning, but hospital resource allocation algorithms of all kinds do not fall into their jurisdiction. As a result, there is very little to prevent hospitals from adopting algorithmic systems in an attempt to better serve their patients, only to have those systems reinforce the same problems they were meant to solve.

To improve health equity while spurring innovation in these algorithms, the federal government should implement the following recommendations:

1) **Grant the FDA Jurisdiction Over Medical Resource Allocation Systems.** The Administration should issue an executive order stating that the definition of a medical device includes medical resource allocation systems; this will allow the FDA to bring their considerable expertise and work on AI health applications to bear on this particular application.
2) **Push for Standardized Systems**. NIST should ask medical associations like the Emergency Nurses Association should identify and endorse software that would be a strong candidate for national adoption. This would encourage innovation and standardization in this area.
3) **Create an Artificial Intelligence Bureau**. AI is nearly ubiquitous, and its challenges are too. Members of the AI Caucuses of the House and Senate should put forth, and Congress should pass, a bill creating an independent agency to be the hub of all AI regulation, improving the standardization of AI policy across disciplines and giving the agency the power to enforce standards for fairness, accountability, transparency, and ethics. This may be adapted from the already-existing AI Initiative.

1. Introduction

In 2020, as 'Flatten the Curve' [8] became a rallying cry, the world watched in horror as surges in COVID-19 cases overwhelmed ICU units in hospitals everywhere. During the height of the pandemic in America, one in 12 US hospitals reported that over 95 percent of their total ICU beds were full – compared to 67 percent in 2010 as a baseline [9]. The struggle for scarce resources sparked a national debate about methods of allocation and triage, including an overarching conversation about equity [10]–[12].

But the COVID-19 pandemic, tragic as it has been and continues to be, is only one facet of the challenge of allocating scarce healthcare resources to those who need it most. The problem has been recognized in several other contexts, most obviously in the overcrowding of our emergency medical system.

1.1 Emergency Department Overcrowding: A Case Study

This problem has been documented since 1987, when the first statewide conference on hospital overcrowding was held in New York. Little has changed since then, due both to the challenge and expense of finding a solution, and a lack of regulatory incentive to do so [13]. However, physicians and patients alike agree that overcrowding is an enormous problem that should be taken more seriously than it is [14] – a view that the government also endorses.

In 2007, the Institute of Medicine (IOM) published a report, "Hospital-Based Emergency Care: At the Breaking Point" [4]. They noted that between 1993 and 2003, while the population of the US increased by 12 percent, hospital admissions rose by 13 percent, and the number of ED visits rose by more than 26 percent. Simultaneously, due in part to cost-cutting measures, the US lost 703 hospitals, 198,000 hospital beds, and 425 EDs. As a result, 60 percent of hospitals were over capacity by 2001.

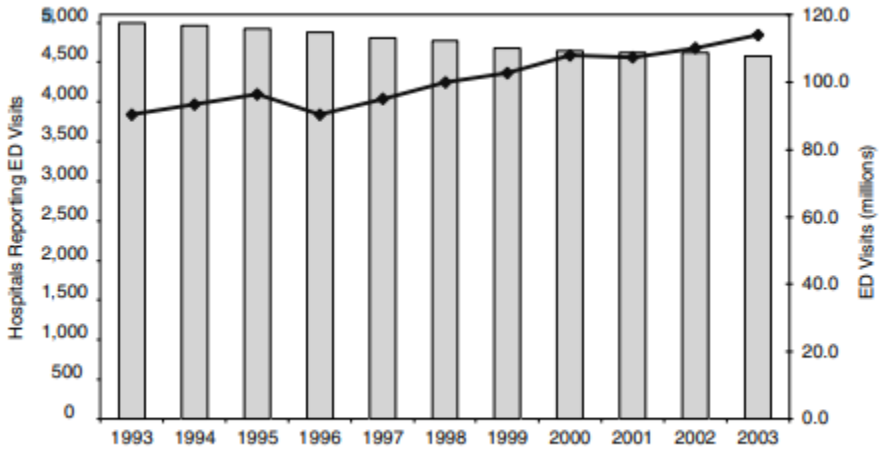


Figure 1: Hospital emergency departments versus numbers of visits. Source: Adapted from [4].

As demands on the emergency system increase, and hospitals continue to close, this trend has continued. By 2015, there were roughly 45 ED visits per 100 people [15]. In 2018, the NCHS reported 130 million ER visits, which comes to 40.4 visits per every 100 persons – slightly lower than the peak in 2015, but higher than the 39.3 visits per 100 persons in 2003 [16]. Overall, in the past 20 years, ED use has increased by 35 percent, while the number of EDs has decreased by 11 percent [17].

1.2 The Emergency Department's Changing Role

A primary reason for the increased use of emergency services is that EDs have gradually become what the IOM termed "a key component of the health care safety net" [4]. They are a crucial source of care for the more than 28.9 million uninsured people (roughly 11 percent of all non-elderly people) in the United States who have limited or nonexistent access to other providers [18].

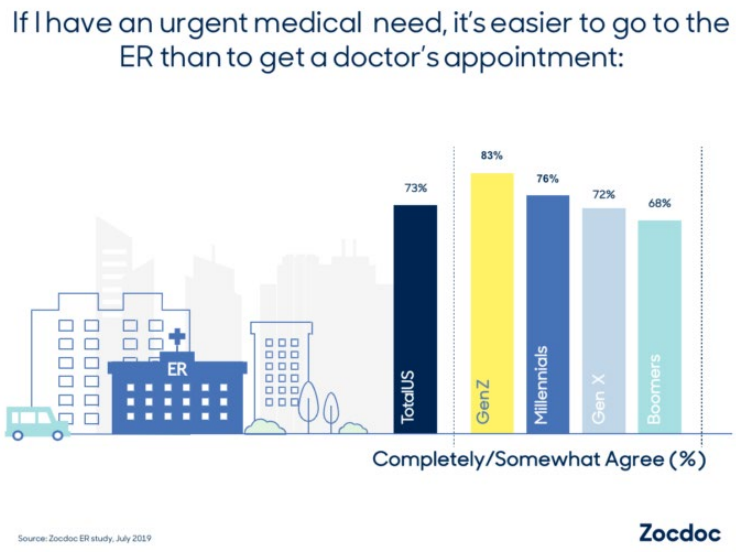


Figure 2: Zocdoc 2017 study result demonstrating reliance on ER. Source: Adapted from [20].

As wait times for physicians grow increasingly longer (one 2017 study found that the average time to schedule a new appointment with a doctor was 24 days [19]), EDs have also become the default choice for Americans who need medical care but can't wait until the next available doctor's appointment. Even though four out of five Americans have an established relationship with a primary care physician, most people are likely to visit the emergency room[1] (ER) if they cannot quickly schedule a doctor's appointment. More than 70 percent of Americans think that it is easier to go to the ER for an urgent care than to try to schedule an appointment with their doctor. People may also use the ER for non-urgent medical care – 15 percent of Americans see the ER as their primary doctor, and one out of every three Americans has used the ER for non-urgent medical care at least once [20].

Physician-referrals to the ED also make up a large portion of use cases. In 2017, a study found that nearly one quarter of visits to the ED were due to referrals by outpatient providers. This referral could be due to a number of reasons – more convenient times, more available equipment, or even a physician's unwillingness to directly admit patients to the hospital [21]. Often, this may lead to boarding – holding a patient that needs to be admitted into the hospital in the ER until a bed becomes available [4].

No matter the cause, overcrowding has become a serious concern. In 2017, 50 percent of EDs experienced overcrowding, and 90 percent of EDs reported overcrowding as a recurrent problem [13].

[1] ER is used here to represent the physical location; ED will be used consistently throughout the paper to refer to the field of medicine in the abstract.

1.3 Risks of Overcrowding

Overcrowded EDs can have dangerous, even deadly consequences, such as delays of care for people who need it – a 2014 CDC study reported that patients at the ER waited roughly 30 minutes before they were even seen, and over two hours on average to receive care [22]. The later a patient must wait to be seen, the greater the risks of permanent negative consequences to their health. Delayed care has been shown to increase the total length of a patient's stay in the hospital, using up beds and increasing overcrowding. It also may result in a 25 percent relative increase in mortality [13].

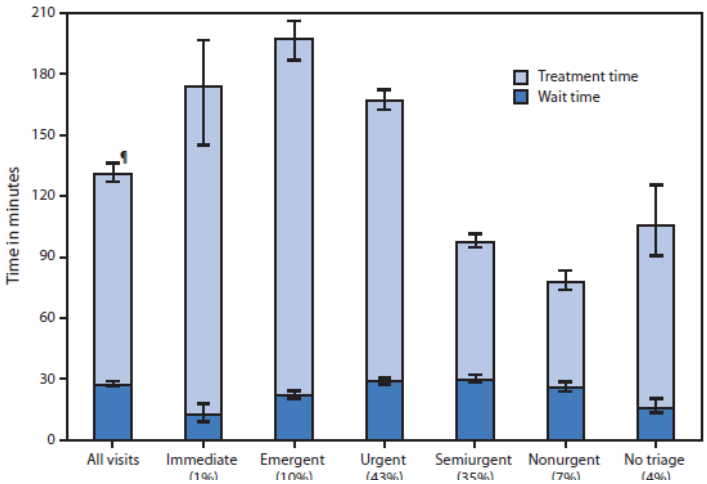


Figure 3: Time to Be Seen in an Emergency Room. Source: Adapted from [22]

When EDs are overcrowded, many patients may walk out rather than wait – no matter how urgently they need the care. Many of these patients will later require hospitalization. Overcrowding also places an enormous strain on healthcare providers, increasing the likelihood of errors and decreasing the quality of medical care [13].

1.4 Beyond EDs: The Need for Resource Allocation Strategies in Healthcare

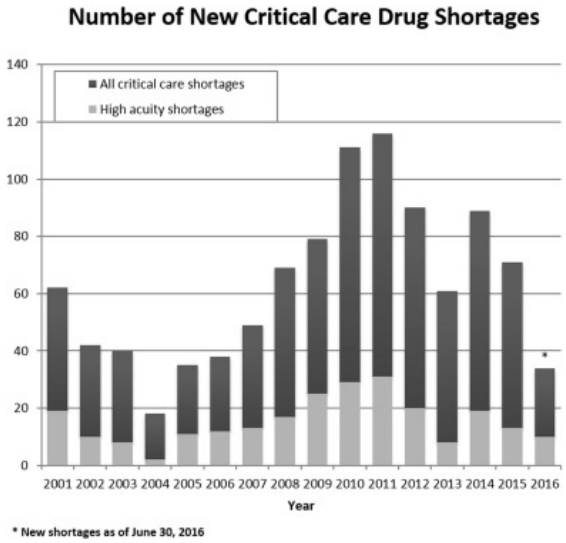


Figure 4: Number of Critical Care Shortages Over Time. Source: Adapted from [24].

The ED is an environment of scarcity. Boarding and overcrowding have limited the number of available beds to the point that in 2015, nearly one third of all EDs resorted to ambulance diversion – sending ambulances to another ED with more space [23]. Even before COVID-19, EDs faced frequent severe drug shortages. Painkillers, antibiotics, and cardiovascular drugs were most affected, with shortages lasting anywhere from three months to two years [24]. In 2018, the American College of Emergency Physicians (ACEP) reported that 90 percent of emergency physicians experienced shortages or absences of critical medicine in their department; over one third of physicians reported that their patients had been harmed by these shortages [25].

As a result, EDs often resort to triaging methods – methods which "prioritize incoming patients and … identify those who cannot wait to be seen." Generally, these methods are performed by triage nurses, who briefly assess each patient and assign them a triage acuity level that serves as a proxy for how long they can safely wait before receiving care. Until recently, no single standardized method for triage existed [26]. Recently, several companies have been working to create e-triage systems that use machine learning to gain insights

from patients across the country. However, these methods are often unvalidated, opaque, and unregulated, leaving many to worry that these algorithms, and indeed, triage in general, may exacerbate healthcare disparities that already exist.

But EDs are only one of several use cases for medical resource allocation systems (MRAS). Indeed, hospitals across the country are moving away from a "one-size-fits-all" practice of offering the same level of resources to every patient, towards a risk stratification strategy that assigns a risk category to each patient that determines eligibility for resources. The National Association of Community Health Centers (NACHC) estimates that 20 percent of the population accounts for 80 percent of the total health care spending in the United States, and of these, five percent account for nearly half of U.S. health expenditures. By assigning risk levels to each patient, hospitals can target their interventions differently for separate groups of patients. Patients in these upper levels of need receive proactive care management services to help prevent unnecessary emergency or acute care, while patients with easily managed care will not require unnecessary services. This kind of resource allocation is meant to be significantly more clinically effective and less prohibitively expensive than non-stratified systems [27].

MRAS, and their potential for decreasing healthcare costs overall by more effectively targeting interventions, are critical given the ever-rising healthcare costs in America. Healthcare costs grew 4.6 percent in 2019, reaching $3.8 trillion or $11,152 per person. Health spending also accounted for 17.7 percent of the nation's Gross Domestic Product [28].

COVID-19 demonstrated the need for effective, data-driven MRAS, and our long history of ER overcrowding shows no sign of stopping. Furthermore, as healthcare costs continue to rise, hospitals struggle to determine how to use their funding most effectively to serve their communities. Improper triage can be expensive, dangerous, and even deadly. So long as the ED continues to play a crucial role as a safety net for the roughly 30 million people without healthcare and anyone with an urgent medical need or lack of access to a non-emergency doctor, and so long as hospitals are forced to resort to risk stratification measures to care for their patients, lives hang in the balance.

2. Background

Although MRAS in hospitals generally are a moderately new development, triage in the emergency department has existed for centuries.

2.1 ED Triage Systems in America

The United States was the first country to create emergency triage as a subspecialty for nurses; Britain and Australia followed within the next few decades. However, there are no national guidelines for triage, and nurses learn through experience [30].

Most ED triage systems are designed by experts in clinical emergency medicine who create decision trees or algorithms based on research evidence, and use those risk assessments to define triage levels. Scales generally have either three, four, or five levels of urgency. Systems generally include vital signs – level of consciousness, respiratory rate, heart rate, blood pressure, oxygen saturation, and body temperature – and clinical descriptors. These variables are weighted and summed together to determine the triage level [30].

The United States historically relied upon several different triage systems, but healthcare professionals have been advocating for a single standard ever since the turn of the 21st century. In 2017, ACEP and the Emergency Nurses Association (ENA) published a statement supporting the adoption of the Emergency Severity Index (ESI), a five-level triage scale. In the following years, hospitals have generally moved towards the ESI and away from other standards [26]. In 2019, the ENA officially purchased the ESI triage system, and the system is currently being used in approximately 80 percent of US hospitals [31].

The ESI has five acuity levels that represent how urgently patients need to be seen by the physician or healthcare provider: immediate (immediately), emergent (one to 14 minutes), urgent (15 to 60 minutes), semi-urgent (one to two hours), and non-urgent (two to 24 hours). The program was developed to consider not only in which order patients should be seen, but what resources are necessary to care for the patient [26]. ESI has been validated extensively through a number of hospital studies [26], and appears to perform approximately as well as other methods when assessing patients who needed emergency interventions [32].

When evaluating triaging systems, researchers generally use two criteria: sensitivity and specificity. Sensitivity, also known as the true positive rate, refers to the algorithm's ability to accurately detect a certain condition. Specificity, also known as the true negative rate, refers to the algorithm's ability to correctly determine when a certain condition is not present [33]. For example, an algorithm that diagnosed every patient with cancer regardless of their actual health would have a very high sensitivity (it detected every instance of cancer) but a low specificity (it had a very high number of false positives). A meta-analysis found that the ESI had a sensitivity of around 85 percent and a specificity of around 90 percent when detecting high-urgency patients (i.e., it failed to detect 15 percent of high-urgency patients, and had roughly a 10 percent false positive rate); this was on par with the triage systems commonly used in other countries [34]. Researchers may also use the false positive rate.

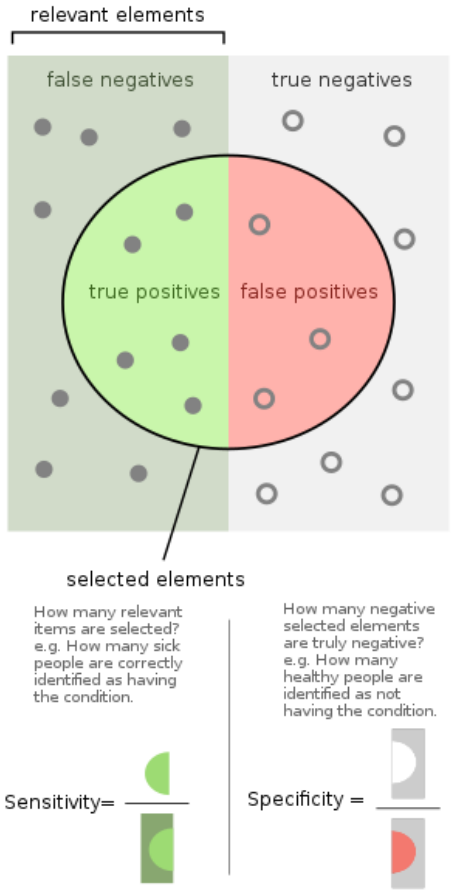


Figure 5 Specificity and Sensitivity. Source: Adapted from [33].

When using the ESI system, the triage nurse must first determine if the patient requires emergency life-saving treatment like assisted ventilation, intubation, defibrillation, or control of major bleeding. If so, the patient requires immediate attention, and is categorized as Risk Level One. Next, the nurse must determine if the patient is in a high-risk situation, evidenced by either confusion, lethargy, disorientation, or severe pain. The ESI handbook directs the nurse to determine if the patient is high risk "based on a brief patient interview, gross observations, and finally the 'sixth sense' that comes from experience." The patient's pain or distress is determined using clinical observation and / or a self-reported pain rating. High

risk patients are assigned to Risk Level Two. The remaining risk levels are determined based on how many different resources (chemical tests, electrocardiograms, intravenous medications, specialty consultations, or procedures) will likely be required to treat the patient. Abnormal vital signs may also elevate a patient into a higher risk category than they would be otherwise. The official algorithm is replicated on the following page [26].

Although the ESI has clear guidelines, the algorithm relies heavily on clinical interpretation of a patient's condition, leaving the triage process open to a number of cognitive biases, including overconfidence, the anchoring effect, information and availability bias, and risk tolerance, and environmental challenges like interruptions and time constraints. As a result, nurse triage accuracy, even using the ESI, is reported at around 60 percent [17].

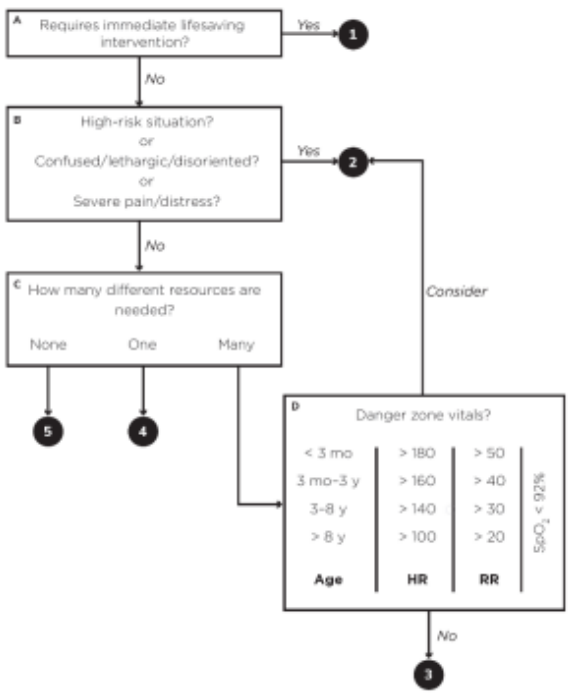
2.2 Risk Stratification

The US does not have a standard risk stratification method. Some systems merely provide an unweighted count of the number of comorbidities a person has. Others, like the Chronic Disease Score and RxRisk systems, use pharmacy data as a proxy for chronic disease and aim to predict future health care costs. The Charlson Index evaluates a patient's predicted care requirements based on their age but informed by their comorbid predictions, while the Adjusted Clinical Groups System (ACG) creates groups using medical records and insurance claims, aiming to predict costs and outcomes. The Cumulative Index Illness Rating Scale compiles severity scales for fourteen domains across the human body and aims to correlate it with health outcomes and care utilization, and the Duke Severity of Illness (DUSOI) measures illness severity using symptoms, complications, prognosis without treatment, and treatment potential. Each scale has its own strengths – the Charlson index and the ACG system are the strongest predictors of mortality, while the ACG system and pharmacy-based systems are the best at predicting healthcare utilization [35].

Overall, the current research shows that the ACG model seems to outperform all other models when predicting hospitalization, ED visits, 30-day admissions, and healthcare expenditures.[36] The algorithm for the ACG system is not available online [37].

However, these methods have recently come under fire, since research has shown that medical records and claims – the data the ACG uses – can only explain a fraction of patient health care outcomes. As a result, the AAPCHO, NACHC, OPCA, and other stakeholders have been advocating for the creation of a new risk stratification algorithm, called PRAPARE (Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences) that would consider not only clinical and claims data, but demographic, mental health/substance abuse, and social determinants of health and could be standardized across the entire United States [38].

Figure 2-2. ESI Triage Algorithm, v4



A. **Immediate lifesaving intervention required**: Airway, emergency medications, or other hemodynamic interventions (intravenous access, supplemental oxygen, monitor, electrocardiogram and labs DO NOT count); and any of the following clinical conditions: intubated, apneic, pulseless, severe respiratory distress, pulse oximetry (SpO_2) < 90%, acute mental status changes, or unresponsive.

Unresponsiveness is defined as a patient who is either
1. Nonverbal and not following commands (acutely) or
2. Requires noxious stimulus (P or U on AVPU scale).

B. **High risk situation:** A patient you would put in your last open bed.

Severe pain/distress is determined by clinical observation and/or patient rating of greater than or equal to 7 on 0-10 pain scale.

C. **Resources**: Count the number of different types of resources, not the individual tests or radiographs (e.g.,: complete blood count, electrolytes, and coagulants equals one resource, while completed blood count plus chest radiograph equals two resources).

Resources	Not Resources
• Labs (blood, urine) • Electrocardiogram, radiographs • Computed tomography, magnetic resonance imaging, ultrasound, angiography	• History and physical exam (including pelvic) • Point-of-care testing
• Intravenous fluids (hydration)	• Saline or heparin lock
• Intravenous, intramuscular, or nebulized medications	• Oral medications • Tetanus immunization • Prescription refills
• Specialty consultation	• Phone call to primary care physician
• Simple procedure = 1 (laceration repair, urinary catheter) • Complex procedure = 2 (procedural sedation)	• Simple wound care (dressings, recheck)) • Crutches, splints, slings

D. **Danger Zone Vital Signs**

Consider uptriage to ESI level 2 if any vital sign criterion is exceeded.

Pediatric Fever Considerations

1–28 days of age: assign at least ESI level 2 if T > 38° C (100.4°F)

1–3 months of age: consider assigning ESI level 2 if T > 38°C (100.4°F)

3 months–3 years of age: Consider assigning ESI 3 if:
1. T > 39°C (102.2°F), or
2. Incomplete immunizations, or
3. No obvious source of fever

Figure 6: ESI Triage Algorithm. Source: [26].

2.3 Machine Learning and Data-Driven Decision-Making

Machine learning is one of the main driving forces behind the development of risk stratification and triage software. Machine learning systems, described most generally, take in vast quantities of data and use statistical methods to generate relationships between sets of variables. One vast field of machine learning is supervised learning, where a computer is given both a set of inputs (for example, a dog's hair color, size, weight, fur texture, and snout length), and outputs (the breed of dog) and tasked with mapping the inputs to the outputs. Sometimes, these algorithms are parametric, which means they can be represented as a function or a weighted sum of the inputs. Regression models, lines of best fit, and logistical regression models are examples of this. Algorithms may also be non-parametric, which means they do not create a single universal equation that can be applied, but instead use a series of decisions to categorize each new occurrence.

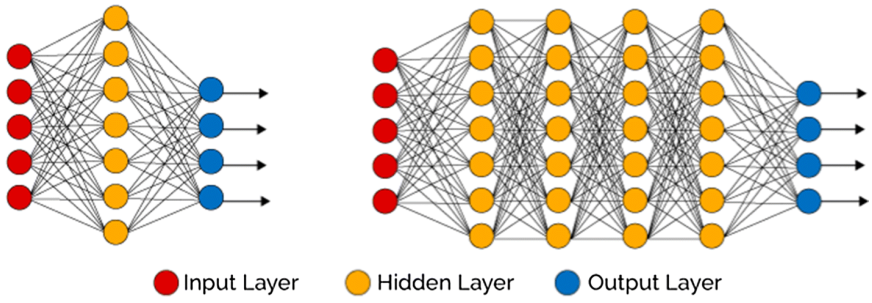


Figure 7: Machine Learning versus Deep Machine Learning. Source: [39]

Deep learning is also a commonly implemented technique. Deep learning is essentially a parametric equation, in that it uses a complicated function, rather than a decision tree, to map inputs to outputs. However, the equation resulting from deep learning is infinitely more complex than that from shallow learning, due to the numbers of internal inferential layers in between. Each layer allows the algorithm to extract secondary features from the data, and hopefully detect new patterns that would not otherwise have been visible [39].

As discussed earlier, one of the main concerns of hospital triaging systems is their subjectivity – the algorithm implicitly requires nurses to make their own judgments, often without the uninterrupted time needed to make the decision accurately. Furthermore, human decision-makers of any kind are subject to many unintentional cognitive biases. Among these are the anchoring effect, in which people pay the most attention to the first information they receive; the diagnostic bias, which leads people to attempt to resolve an issue as early as possible to move on to other things; the confirmation bias, which makes people more receptive to information that supports their hypotheses, and risk tolerance [40]. Computer algorithms can often make decisions in real-time, and are not subject to these cognitive biases because they do not 'think' in the same way that humans do. Furthermore, machine learning systems can be standardized and well-validated, and thus are subject to less of the uncertainties that come with human decision-makers.

As a result, machine learning approaches have become increasingly common in patient risk assessments. Algorithms can be used to predict mortality, hospitalization, readmission rates, lengths of stay, ED crowding, wait time, admission rates, and a number of other outcomes [17], [41]. The ENS is currently

urging its members to adopt KATE, an algorithm developed by Mednition that uses natural language processing and other statistical tools to determine triage acuity, which seems to perform at least 25 percent better than nurses alone [17], [31]. Similarly, machine learning approaches for risk stratification and other forms of resource allocation are growing in prevalence and popularity [6].

Overall, engineers and clinicians alike are excited about the potential of machine learning to revolutionize risk allocation and provide better care to patients. Over the upcoming years, machine learning will become an ever-more-crucial component of emergency medicine and population management [42]–[44].

3. Key Concerns

The excitement and potential of machine learning in medicine must be balanced in view of the enormous risks that come with adopting such a technology. Triage and risk stratification has already been extensively challenged for not considering the right factors, and for embedding potential biases into decision-making processes. Furthermore, case studies in other fields demonstrate that machine learning may be prone to incorporating implicit biases or errors and ratifying them.

The kinds of machine learning technologies used for complicated decision-making are also plagued with another problem – a lack of transparency. Several factors contribute to this opacity, including trade secrets, health privacy, and the technical challenges of deep learning.

Perhaps the most issue is that there is currently no oversight for these kinds of algorithms – although medical devices are subject to regulation, risk stratification is not technically a medical device. With all of these concerns in play, public pushback against artificial intelligence in decision-making may turn hospitals away from adopting these technologies, thus losing the benefits to both fiscal stewardship and patient outcomes that will come as a result of their usage.

3.1 Risk of Health Inequities

One of the major concerns of any resource allocation algorithm is that it will exacerbate preexisting inequities in healthcare. Indeed, healthcare disparities have been a long-standing problem throughout our nation's history. Healthy People, a government program tasked with monitoring and setting objectives to improve the country's health, defines a health disparity as "a particular type of health difference that is closely linked with social, economic, and/or environmental disadvantage. Health disparities adversely affect groups of people who have systematically experienced greater obstacles to health based on their racial or ethnic group; religion; socioeconomic status; gender; age; mental health; cognitive, sensory, or physical disability; sexual orientation or gender identity; geographic location; or other characteristics historically linked to discrimination or exclusion" [45].

Indeed, one of Healthy People's overarching goals has constantly been to eliminate health disparities [46]. However, these disparities are clearly evident in healthcare generally, and in MRAS more specifically. A 2003 IOM report that studied differences in kinds and quality of healthcare received by racial and ethnic minorities in the United States found that "racial and ethnic disparities in healthcare exist, and because they are associated with worse outcomes in many cases, are unacceptable" [5].

These disparities still exist. People of color fare significantly worse than their white counterparts on a majority of health status measures [47]. Although much of this may be due to systemic features from areas ranging from economic stability to education, to even community, safety, and social context, these disparities exist at the clinical level as well. In an emergency room setting, people of color are less likely to receive analgesia, even when the severity of their pain as assessed by a physician was the same [5]. Some studies have shown that misconceptions about biological differences between people of different races may also contribute to this, as physicians may have a tendency to perceive the pain of a person of color as less severe than of a white person [48]. This is particularly concerning given that analgesics are one of the most common drug shortages.

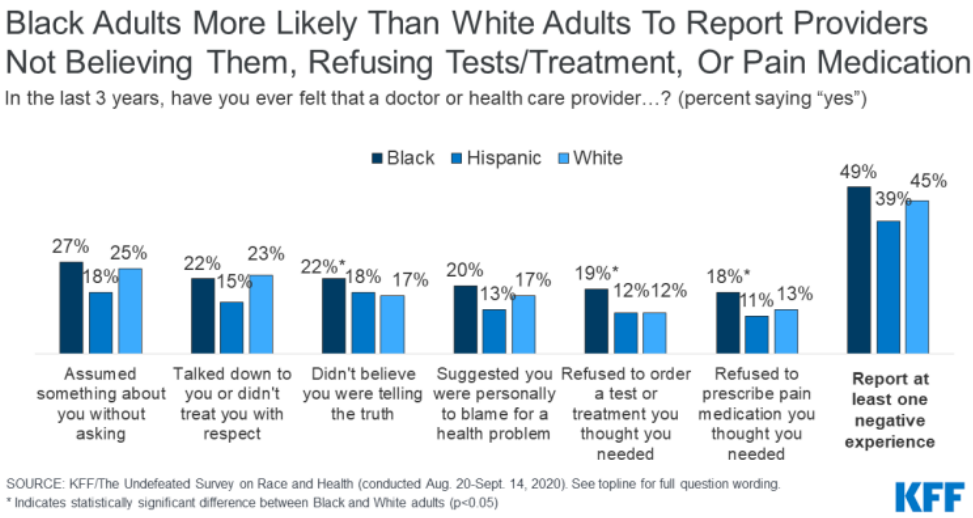


Figure 8: Report on Patient Experiences with Providers. Source: Adapted from [49].

Indeed, race seems to play a large role in how patients and healthcare professionals view each other. One in five Black adults reported experiencing discrimination in healthcare because of their race or ethnicity, and Black adults are significantly more likely to report an experience where a physician didn't believe them, refused to order a test or treatment they thought they needed, or refused to prescribe pain medication [49]. On average, Black patients face significantly longer ED wait times than white patients, although this disparity seems to diminish as illness severity increases [50].

Gender also tends to affect the patient / provider relationship. Some studies have suggested that women are less likely to receive ICU treatment, regardless of how severe their illness is [51]. Women are significantly more likely to report negative experiences with a health care provider, particularly because they assumed something without asking, talked down to them, or didn't believe them [49].

Physicians may even look to other factors beyond acuity score when determining which patient to see. One study found that patients who were referred to the ED were more likely to be hospitalized than self-referrals, even after controlling for the seriousness of their medical condition [21]. Another study found that as long as patients were within the targets for wait times, physicians were more likely to prioritize the patients that had been waiting longer, regardless of acuity [52].

Proponents of machine learning hope that a standardized algorithm, free from the implicit biases of human decision-makers could help to overcome both the systemic disparities and the bias implicit to a

patient-physician relationship. However, critics of machine learning argue that past cases demonstrate that artificial intelligence is more likely to ratify and solidify these biases, than to overcome them.

3.2 Machine Learning May Carry Biases of Its Own

All decisions are prone to error, whether they are based on instinct, mathematical rules, or something in between. Intuitively, many people believe that a decision based on a computer is somehow more fair or impartial because the machine itself cannot make cognitive errors. However, machine learning systems are certainly not infallible, and cases across disciplines have made evident the very real problems inherent in – and sometimes obfuscated by – machine learning algorithms.

Although these errors might take many forms, one of the most common is disparate prediction, or an algorithm that performs more accurately in certain situations. For example, facial recognition algorithms perform far better on white men than any other demographic, achieving near-perfect accuracy for white men, but misclassifying one in every three darker-skinned women [53]. When the ACLU tested Amazon's facial recognition software against members of Congress, the software falsely identified 28 members as people who had been arrested; 11 of these, or 39 percent, were people of color, even though only 20 percent of Congress were people of color at the time [54]. These errors likely came about because the algorithm was given far more examples of white faces than Black faces – an issue that will likely arise in clinical cases as well.

Another common issue with algorithms is that they may fail to generalize to a new context because they are basing their decision on an underlying variable not present in other situations. This has been frequently shown in medical contexts – software meant to detect skin lesions may actually detect rulers in the picture [55], or pen markings [56]. In multiple x-ray cases, computer vision algorithms based their diagnostic decisions on information about the type of x-ray or scanner equipment used, rather than the portion of the image that showed the patient's bones [57], [58]. Since every hospital is different, and algorithms are often trained with data from only a few hospitals, there is a non-negligible chance that MRAS are also subject to these kinds of concerns.

Algorithm performance may also degrade if it is not kept up-to-date – for example, a model trained on data from 2017 will struggle to predict data from 2027 if the underlying data and patterns within the data have changed over time [59].

Perhaps most sensationally, artificial intelligence has a well-known propensity to learn from the biases of the data that it receives.

Artificial intelligence may also work 'correctly' but still lead to bad outcomes. For example, Amazon's AI recruiting tool was scrapped after it was discovered that it penalized resumes with the word 'women's', or female-coded language like "sorority", "softball", or "baking." This error occurred because the data had been trained on the resumes Amazon received for the past 10 years, in a field that has been dominated by men since its inception. Thus, in learning to match resume terms to job offers, the resume had learned to favor men, just as Amazon's HR department had. Although people hoped the algorithm would reduce implicit bias in hiring, it instead learned from that bias and codified it [60].

Existing AI triage and risk stratification systems are clearly not exempt from these errors and biases. One widely-cited article found that for any given risk level, the Black patients were far sicker than white patients, and correcting this disparity increased the number of Black patients receiving additional help from 17.7 percent to 46.5 percent. This error was caused because the algorithm predicted health care costs, but preexisting biases in the medical system meant that Black patients received less care, and thus less money spent on them, than white patients with the same level of need, and the algorithm reflected that bias [7]. Studies conducted by the VA have found disparate impact on the basis of race in two risk assessment algorithms: STORM (Stratification Tool for Opioid Risk Management), and CAN (Care Assessment Needs) – and the research is still ongoing. These researchers believed these errors could be caused by class imbalances (as was the case in facial recognition), potentially biased outcomes (similar to Amazon's recruiting tool), potential confounding variables, omitted variables, and / or unmeasured confounding variables [61].

Algorithms used to predict medical risk may have deadly consequences if they fail to act as they should, but building a fair algorithm is a challenge for all of the reasons mentioned. If the preexisting data is biased, incomplete, unbalanced, out of date, or subject to some confounding force, the algorithm's accuracy on new data may be compromised. But even more fundamental to these issues is how to determine if the algorithm is fair to begin with – a simple question with a complicated answer.

3.3 What is Fairness?

One of the major challenges of fairness in artificially intelligent systems is the problem of defining what exactly fairness means. There are two leading definitions of fairness – false negative parity, and calibration [61].

The first of these, false-negative parity, means that people within a certain group have the same chance of false negatives as people in another group. In a triage system, that would mean that if white people were inaccurately categorized as low risk at a rate of 5 percent, Black people would also be categorized as low risk at a rate of 5 percent. If 10 percent of the Black people categorized as low risk were actually high risk, while only 5 percent of the white people categorized as low risk were actually high risk, the algorithm would be unfair based on the definition of false negative parity.

The other definition of fairness is calibration, which is shown when the risk score given to a group matches the observed outcome. In a triage system, consider the risk of hospitalization. If a patient was given a risk score corresponding to a 10 percent chance of hospitalization, the group of all patients with that same risk score should be hospitalized roughly 10 percent of the time. However, if of that group of patients, 20 percent were hospitalized, then the algorithm would not be calibrated, and thus would not be fair.

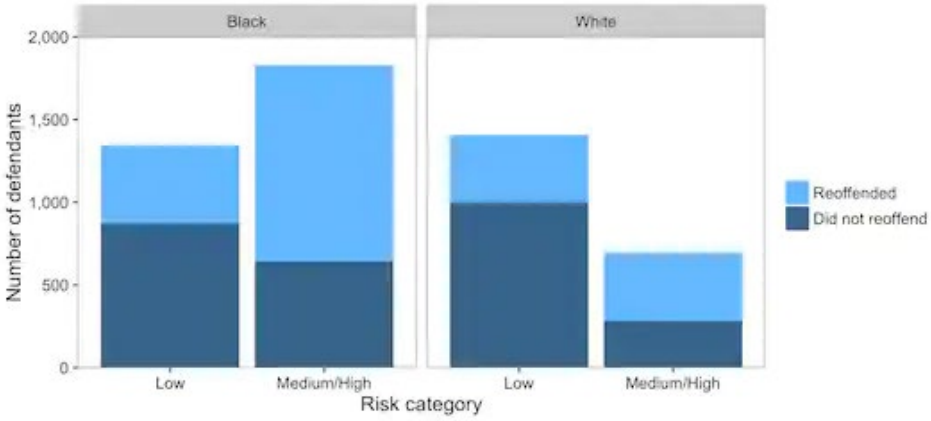


Figure 9: Results of Recidivism Algorithm, displaying challenges of defining fairness. Source: [62].

These two definitions of fairness may, at times, be mutually exclusive. The figure to the right shows an example of how an algorithm can be calibrated without having false-negative parity. The data came from an algorithm used to predict the likelihood of a prisoner reoffending if they were released on parole, classifying each prisoner as low or medium/high risk. Prisoners in the medium/high category had a roughly 60 percent chance of reoffending, no matter their race; this was equal between the two groups. However, because Black people were less likely to be categorized as low risk, a non-reoffending Black person had a higher chance of being considered medium/high risk than a non-reoffending White person, thus there was not false-negative parity [62].

In medical triage, these two definitions of fairness both seem quite important. We want to ensure that people do not lose out on care because they are falsely classified as low risk – thus, false-negative parity is an important definition of fairness. But at the same time, we do not want our system to provide different outcomes for different people, such that high-risk people of one group are only half as likely to need intensive care as high-risk people in another group. Determining which standard(s) of fairness and how to weight them is an ongoing issue, one that requires conversations between subject matter experts and algorithm developers.

3.4 "Black Box" Algorithms and the Challenge of Validation

One of the principal challenges to analyzing algorithms is that their inner workings are often inscrutable. There are a number of reasons for this – first, some algorithms, especially deep learning and non-parametric models, are highly mathematically complex and challenging for even an experienced machine learning expert to understand. An algorithm can easily rely on thousands, if not millions, of weighted values – far too many for any person to make logical sense of. For this reason, many AI algorithms are often referred to as 'black boxes' – inputs go in, and answers come out, without any reason given.

But algorithms are often black boxes for another reason – companies have an incentive to keep their algorithms, and the data they are trained on, a secret. As a general rule, AI technologies cannot be patented, and technologies that rely on health data that must be kept private under HIPAA are even less patentable. Thus, to protect their intellectual property, companies treat their algorithms as trade secrets, posting very little about their structure, the data used to build them, and even their performance. So long as they claim their algorithms as proprietary information, external researchers cannot access the data and validate it.

Even the question of how to validate it becomes a challenging one, since there are so many different metrics at play.

3.5 Lack of Regulatory Oversight

Unfortunately, it's not immediately clear who is responsible for overseeing these algorithms. Although there are several potential alternatives, none of them have clear jurisdiction.

3.5.1 FDA and Software as a Medical Device

One obvious choice might be the FDA. The FDA's Center for Devices and Radiological Health has the authority to regulate all medical devices sold in the United States. Medical devices are classified into three tiers based on their level of potential harm, each with their own regulatory requirements [63].

Class	Percent of Medical Devices	Description	Examples	Type of Regulation
I	47	Minimal potential harm to user, simple.	Elastic bandage Enema kit	General; exemption (95% of all Class 1 devices are unregulated)
II	43	Slightly greater potential harm; more complicated technology.	Powered wheelchair	General or Special
III	10	Sustain or support life, implanted, present possible unreasonable risk of illness or injury.	Implantable pacemaker, breast implant	General and Premarket Approval

Table 1: Medical Device Classifications. Source: Self-generated, based on information from [64].

General controls require that the device be registered, and that producers notify users if the device needs to be replaced, recalled, or repaired for any reason. They also require companies to keep records and reports on any adverse events, track their devices, and report any removals or corrections. Lastly, these controls contain provisions for good manufacturing practice. Special controls go a step further, and include performance standards, postmarket surveillance, special labelling requirements and guidelines, and some premarket data requirements [65]. A Class III device triggers premarket approval, which is the FDA's scientific review to ensure that the device is safe and effective. It is similar to drug testing in its stringency [66].

The FDA recognizes that medical devices might include software, even in the absence of any physical component. They rely upon the IMDRF definition of software as a medical device (SaMD) as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device" [67]. Software as a medical device is broken into four categories with varying levels of risk, depending on whether a device is intended to treat or diagnose, drive clinical

management, or inform clinical management, and whether the healthcare condition is critical, serious, or non-serious [68].

The FDA has also taken steps to create an entirely new review framework for machine learning software, given its unique challenges. This framework would implement guidelines for good machine learning practice for SaMD, and has stated a focus towards "usability, equity, trust, and accountability." They are also working to support the development of methods to evaluate and improve machine learning algorithms, including detecting and eliminating bias. They have also stated a commitment to requiring real-world monitoring of AI software to ensure that it works as well on new data as it did during development [69].

Unfortunately, triaging and risk stratification systems are not technically medical devices. Under Section 201(h) of the Food, Drug, and Cosmetic Act, a device is:

> *an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is-*
>
> *(1) recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them,*
> *(2) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or*
> *(3) intended to affect the structure or any function of the body of man or other animals, and*
>
> *which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes [70].*

It might be possible to argue that triage software is "intended for the diagnosis… cure, mitigation, treatment, or prevention of disease" in that it determines the order in which people are seen, but this is not a particularly strong argument, since it is more of a process aid than a treatment aid. The case for risk stratification might be stronger in that it is meant to prevent disease, but it does so obliquely by helping identify which members of a population might benefit most from extra help.

The FDA recently released draft guidance for clinical decision support software, which is software that is not intended to acquire, process or analyze data, but rather intended for the purpose of supporting or providing recommendations. While this might seem on its face to apply to risk stratification software, the FDA defines this second criteria as "intended to assist HCPs in making patient-specific care decisions" [71]. Thus, it would not apply to triaging systems where the healthcare provider would use the software to learn who to prioritize, or risk assessment systems where the provider would need to identify members of the population most in need of extra support.

As a result, although the FDA's framework for machine learning applications would provide an excellent regulatory scheme for MRAS, the current legal definitions of "medical device" and "clinical decision support" mean that these systems are left unregulated on an administrative level.

3.5.2 EMTALA

Some policymakers have argued that the Emergency Medical Treatment and Labor Act (EMTALA) could be used to regulate triage systems.[5] EMTALA requires that hospitals cannot transfer uninsured or Medicaid patients to public hospitals without providing medical screening to ensure they are safe to transfer. It also requires hospitals to screen and treat the emergency medical conditions of patients, without respect to ability to pay, insurance status, national origin, race, creed, or color. If a hospital violates EMTALA, both the CMS and OIG can administer hospital fines, physician fines, or termination Medicare provider agreement. Individual people harmed by the policy may attempt to sue the hospital [72].

However, there are several reasons why EMTALA is insufficient. First, judicial interpretation over time has limited EMTALA's ability to deter disparate treatment by requiring only that hospitals conform to their regular practice, in lieu of a national standard of care. This forces plaintiffs to challenge local hospital policy, without any knowledge of the inner workings of the hospital. State laws are similarly permissive [5]. Second, physicians cannot be held liable for their triaging decisions in a private right of action [73]. It is not immediately clear how this would transfer over to an algorithmic decision-maker, and this brings the added challenge of determining who would be liable. Would the hospital be liable for implementing the system? Would the physician, for relying on it? Would the system's creators be liable? Liability for algorithmic decision-making is an open question in any field.

Third, given the black-box nature of most machine learning software, it would be near-impossible to prove that the algorithm is acting discriminatorily. Bias can only be detected through rigorous analysis with large amounts of data, and since companies may claim proprietary protections on their algorithms, this analysis cannot be conducted.

Lastly, bringing a lawsuit at all is prohibitively expensive, especially for low-income patients.

As a result, although EMTALA technically grants a right to sue, and the potential for regulatory action against discriminatory algorithms, in practice it cannot be sufficient to deal with this problem.

3.5.3 Title VI

Similar to EMTALA, some have argued that Title VI of the Civil Rights Act could be used as a means to regulate biased systems. Title VI bars discrimination on the basis of race, color, and national origin in any program receiving federal financial assistance. Although on its face it applies only to intentional discrimination, funding agencies often have regulations that prohibit effective discrimination as well. Individuals have the right to file a complaint with the funding agency, or to sue in federal court [74]. However, physicians who receive Medicare payments to physicians do not receive "federal financial assistance" under the title, which means that private physicians are not subject to Title VI [5]. Furthermore, Title VI does not create a private right of action to ensure that regulations implementing Title VI are actually enforced [75]. As a result, Title VI also lacks the strength to implement controls.

4. Policy Recommendations

A number of different legislative levers could be used to address this problem – either alone, or in conjunction with other tools. These options are set out and evaluated.

4.1 Grant the FDA Jurisdiction Over Medical Resource Allocation Systems

The FDA is already doing critical work in ensuring that medical algorithms are safe and secure. They have a clear track record of working with stakeholders to respond to AI's unique challenges, and are currently working on providing a framework for algorithmic evaluation. Because they already have subject matter jurisdiction over all other forms of medical devices, they are the obvious choice to regulate medical risk assessment systems like triage and risk stratification systems as well.

4.1.1 Objectives and Impacts

Granting the FDA jurisdiction over these systems would solve several problems. It would ensure that there was some government oversight of MRAS, and the FDA's prior work on AI would enable them to easily envelop these systems without a large amount of additional effort. Indeed, housing authority over MRAS in another agency runs the risk of duplicating all of the FDA's work, which would be time-consuming, unnecessary, and could lead to inconsistencies between how the FDA regulates AI and how another department regulates AI.

The FDA's processes and procedures would also be an ideal fit for MRAS. The FDA often regulates through informal, non-mandatory guidelines. Because they are non-binding, the FDA does not need to go through drawn-out notice-and-comment rulemaking processes, and thus can more rapidly adapt their guidance as necessary. Given the speed at which AI develops, this is an enormous benefit, allowing them to better keep up with the technology. Furthermore, the FDA's informal guidance provides far less of a burden on manufacturers than a mandatory regulation would, while still providing best practices that could help shape system design and testing moving forward.

Although companies and hospitals alike might worry that the possibility of regulation of medical risk assessment systems would make these systems more costly to develop and thus less accessible, this is not necessarily the case. It is true that bringing a medical device to market could range from $31 million to $94 million, depending on the regulatory scheme – an enormous barrier to entry for any company, but particularly for a small startup like many of the ones developing AI applications [66]. However, the FDA has a number of safeguards in place. First of these is the "Small Business Program," which waives or reduces many of the required fees for device development for any business with less than $100 million gross its first year [76].

The FDA also has a Breakthrough Devices program for any device that "provides for more effective treatment or diagnosis of life-threatening or irreversibly debilitating human disease or conditions," and represents a new type of technology, lacks an approved alternative, offers significant improvements over existing technology, or for which increased availability would be in the patient's best interests [77]. Medical risk assessment technology would likely qualify. Triage would certainly qualify because the very purpose of triage is to avoid life-threatening or irreversibly debilitating conditions and, at the very least, its availability would help everyone; the first few technologies in this area could also make a claim because there would be no approved alternative. An argument could also be made for other forms of risk assessment and stratification, by saying that enabling providers to direct care to the most vulnerable members of the population, there is the ability to prevent conditions from worsening or becoming life-

threatening or irreversibly debilitating. The Breakthrough Devices program would allow manufacturers to work closely with the FDA to quickly move through the approval process, and would also grant prioritized review of their submission [77]. This would significantly limit the costs of development by lessening the time for the process, and would give manufacturers access to the FDA's significant expertise in device development. Both the Small Business program and the Breakthrough Devices program would do a great deal to reduce the potential cost and time delay to manufacturers and developers, and would also provide them with substantial support when navigating the FDA's approval process.

Manufacturers might also express concerns about disclosure of their algorithm, but this is not a concern. Under 21 C.F.R. §20.61(c), people who submit records to the FDA may designate those records as exempt from public disclosure if they are trade secrets, commercial, or financial information. The FDA also has the right to redact any additional information that they think could be reasonably considered exempt from public disclosure [78]. The FDA constantly handles highly confidential applications for various drugs and devices, including other systems that rely heavily on artificial intelligence; they are well-versed in safeguarding IP.

Perhaps the most likely drawback to granting the FDA jurisdiction over medical risk assessment systems is that while it ensures that these systems are well-validated, it does not encourage their adoption. Indeed, the increased regulation may increase the price and even make them less likely to adopt it. Indeed, on its own, it may not be a perfect solution – but we must not let good be the enemy of great. It is better to first design a system that can be well validated, and then work towards implementation, rather than advocating for the acceptance of a poorly designed and tested algorithm. Indeed, ensuring that these systems receive the same oversight and guidance as all others in the healthcare sphere is a critical step.

Furthermore, there are clear economic benefits to FDA approval. Without FDA oversight, hospitals would be unwilling to make the expensive gamble of adopting an untested system; approval gives hospitals the surety and confidence they would need to be willing to invest in MRAS. Approval would signal to potential customers that the device can be trusted, and could actually increase hospital willingness to adopt them. Approval could also potentially lead to future changes – for example, allowing government funding to offset the cost of implementation for FDA-certified triage or risk stratification systems. Furthermore, one of the primary aims of risk stratification systems is to reduce hospital operating costs by preventing future harm, and not providing unneeded care – as a result, implementation of these systems would be an investment that would repay itself. (It is worth noting here that hospitals earn roughly $700 more for each elective admission than for each patient admitted through the ER. Therefore, if a hospital can reduce emergency admissions by either by triaging patients and getting to those who need care quickly, or by using preventative care management on patients identified using MRAS, this will have an enormous financial gain to hospitals by allowing them to perform more elective admissions [79]). FDA approval could also provide some legal security against any lawsuits alleging biased provision of care – if the FDA certified the soundness of the software making the decision, this could lessen the liability of the hospitals, which would be clearly beneficial, especially as questions of health equity become more relevant.

4.1.2. Implementation

One benefit of this policy solution is that there are several possible paths to implementation.

The most surefire route would be to amend the law defining a medical device, Section 201(h)(B) of the Federal Food, Drug, and Cosmetic Act [80], to read something similar to: "intended for use in the diagnosis

of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, **or in the determination of triage priority or other allocation of medical resources**, in man or other animals, or…" (bolded phrase added). This would clearly, unambiguously, ensure that the FDA's jurisdiction extended to MRAS. Furthermore, it will show that Congress has a demonstrated interest in FDA oversight of MRAS, which will spur the agency to take action rather than to use their discretion and choose not to regulate.

A change of this kind would require legislative action, but due to the simplicity of the change, this could easily be added as an amendment to another bill. At the time of writing, the 117th Congress, two potential candidates would be S.346/H.R.959, the Black Maternal Health Momnibus Act of 2021 (Momnibus) [81], [82], and S.893/H.R.937, the Tech to Save Moms Act [83], [84]. The Momnibus would likely be a better choice, simply because it has more bipartisan co-sponsors, is larger and more inclusive, and takes an equity-based approach to health in general, rather than to a specific medical situation. Potential champions exist in both houses of Congress – many of the co-sponsors of either bill would be likely to favor an amendment to improve the equity of MRAS, which certainly affect the people each bill is meant to aid.

While this option would be powerful and unequivocal, it is perhaps the most challenging given a deadlocked Congress. On the other end of the spectrum, the easiest option would be for the FDA itself to release a guidance saying that the agency is interpreting MRAS as medical devices under Section 201(h)(B) because they are "intended for use in the mitigation… of disease". Since there appears to be no legal precedent to the contrary, there is nothing barring the FDA from doing so. Certainly, opponents to this change could argue against the FDA's interpretation, but under the legal doctrines of agency statutory interpretation, courts will use "Chevron deference" – deferring to the agency's interpretation because the interpretation is not unreasonable, and because Congress has not directly spoken on the issue [85]. The agency may do this either as a rulemaking or an informal guidance; given the FDA's normal procedures, an informal guidance is more likely. This, however, means that such an interpretation is far from permanent, and can easily be overturned by new leadership.

The ideal implementation strategy likely falls between these two alternatives, in the form of an executive action. Such an action would fit well within President Biden's stated priorities of racial equity and health care [86]. An executive order saying that Section 201(h)(B)'s language should be construed to include MRAS, and that the FDA should use their authority to investigate these systems, would have much the same effect as any other implementation, but it would be more authoritative than the FDA, and less likely to be overturned by a change in agency leadership. It also would be likely more insulated from legal action than the FDA's interpretation of their own statute. There is still the risk that it may be overturned by another executive action from a later presidential administration, but undoing this action is unlikely to be an important priority for a new president and likely it will still stand. Indeed, improved oversight of MRAS is a bipartisan issue – when implemented correctly, it will cut hospital operating costs and decrease reliance on the emergency system; furthermore, it will promote ideals of health equity and transparency.

Thus, to ensure that there is oversight of MRAS by an agency well-equipped to handle the challenges of healthcare AI, President Biden should issue an executive order stating that the definition of medical devices should be construed to include MRAS.

4.2 Identify and Endorse Standardized MRAS

To spur development and adoption of MRAS, professional organizations like ACEP, ENA, and NACHC should identify and endorse effective technologies. While the FDA's approval may indirectly induce hospitals to adopt MRAS, endorsements from the professional organizations that would use these systems would be far more persuasive. NIST should request that professional organizations undertake this process.

Indeed, professional organization endorsements have already been responsible for the increasing acceptance of the ESI triage system. In 2003, ACEP and ENA published a joint position statement advocating for the standardization of ED triage systems. In 2005, they specifically recommended the ESI or another scale, the Canadian Triage and Acuity Scale. In 2017, they recommended the ESI specifically as the single best scale [26]. The ENA acquired the rights to ESI in 2019, and it is now the official triage system used. As a result of their endorsement, 80 percent of hospitals in the country use the ESI [31].

Professional organization endorsements would have several benefits. They could lead to one single standardized system, where patients are assured their needs would be evaluated in the same way no matter where they went. One single standardized system would result in the availability of copious amounts of data from which programmers could learn and refine their systems. Standardization would also allow for easier updates and testing within a single system, eliminating the complexity of multiple systems. The process of developing a standardized system would likely allow external researchers to gain easier access to the system for the purpose of vetting it and recommending it for the association's endorsement.

Furthermore, as mentioned before, doctor and nurse buy-in is a powerful force to ensuring that hospitals adopt impartial, standardized MRAS. If hospital administrators understand that employees would significantly benefit from using the software, this is likely to be far more persuasive than simply hearing that the software is FDA-approved. Of course, the conjunction of both FDA and practitioner buy-in is likely to be the most persuasive, and may also improve hospital perception of these technologies as a 'best practice', rather than a regulatory burden.

Practitioner endorsement could also spur competition and innovation – as the practitioners announce that they are looking for a specific software, developers may push for their systems to be adopted, investing in research and testing to ensure that their system is the best on the market. They may even work with the FDA through the Breakthrough Technology program. This would not be particularly challenging – for example, ACEP already hosts a yearly competition for medical device and IT companies [87]. They might easily put out a press release saying that they are particularly interested in triage systems; other organizations could do something similar.

Endorsement could also allay many concerns about potential costs of implementation. Organizations often release their software for free – for example, NACHC and others are providing free use of their PRAPARE risk stratification tool[38], and the ESI handbook is free to everyone on ENA's website [31]. Organizations may choose to purchase a certain software, making it free to use for their members. They may also choose to establish programs for free training courses, or provide stipends for hospitals who choose to adopt it.

There are several methods of implementation, depending on each organization and their specific processes. They might, like ENA, simply acquire the rights to a certain technology, and advertise another on their website [31]. They might, like ACEP, host contests and give prizes for the best technology, in order

to spur industry innovation [87]. They might, like NACHC and others, use their own internal expertise to create their own algorithms, and release it for free [38]. It might take years of research, or a contracted partnership with a certain company. Indeed, much of this should be left to each organization to decide for themselves. But a standardized system, approved and validated, would improve equity and quality of care for all patients, and provider buy-in is critical to making that happen.

Although this policy decision could be done on its own, its impacts would be strengthened by the adoption of the policy described above. The combination of the two would ensure that the technology is vetted. Asking each group to make their own endorsements, for each use case, leaves open the possibility that groups will simply decline to choose; FDA oversight ensures that there are no regulatory gaps. However, individual practitioner groups calling for development of new technologies and promising potential markets is more enticing than the FDA announcing a new regulatory scheme. Indeed, the combination of both policies would combine the strengths of each, while mitigating their respective weaknesses.

4.3 Establish a Federal Artificial Intelligence Bureau

The FDA is certainly not the only organization currently working on ways to regulate artificial intelligence. Bills proposed just in the 116th and 117th sessions of Congress would task the OMB, the National Academies, HHS, the Secretary of Labor, NIST, NSF, FTC, DHS, CFPB, HUD, VA, the Department of Education, FCC, EEOC, and the Civil Rights Division of the Department of Justice, among organizations, with developing AI standards in a number of different contexts [83], [88]–[91]. Indeed, AI usage spans nearly every industry, and the jurisdiction of every industry. Rather than asking agencies to create standards for every use case, leading to a patchwork of regulation, Congress should pass a bill establishing a new independent agency that would have jurisdiction over AI systems in all industries, providing a single common locus for all machine learning regulation and enforcement.

Such a solution is certainly not novel – in fact, Congress followed a similar policy in 2010 when they passed the Dodd-Frank Act of 2010, which among other things established the Consumer Financial Protection Bureau (CFPB), the US's youngest independent agency [92]. Indeed, the CFPB's mission statement sounds quite familiar: "The CFPB was created to provide a single point of accountability for enforcing federal consumer financial laws and protecting consumers in the financial marketplace. Before, that responsibility was divided among several agencies. Today, it's our primary focus" [93]. Like consumer protection, AI is divided among several agencies, and it's not immediately clear who has regulatory oversight over them all. Indeed, CFPB was instituted in a similar situation, with a similarly deadlocked Congress, so the political situation does not necessarily rule out the creation of another agency. An Artificial Intelligence Bureau would share several of the CFPB's benefits: it would create a concentrated hub of expertise and rulemaking for AI, ensuring that regulations were standard and making it easier for companies to comply. It would ensure that all use cases of AI had some kind of oversight, even if it would have been difficult to locate a particular agency with jurisdiction – as is the current issue with MRAS. A single hub would ensure that regulators were not constantly 'reinventing the wheel' with AI standards and solutions, and would likely be faster to adapt to change than several agencies with limited inter-agency communication would be. Finally, a single bureau would give consumers who felt that they had been harmed by AI a clear means of redress, and provide information about AI development in a centralized place, which would benefit industries.

It is also possible to defray concerns about cost by following the CFPB's example. The CFPB, like some other independent agencies, derives its funding from the earnings of the Federal Reserve, and does not

rely on congressional appropriations [92]. An Artificial Intelligence Bureau could do the same thing, and would thus not cost taxpayer dollars to implement; its continued funding would not be a concern.

Although some might argue that creating an entire new agency is inefficient, the opposite is true. Passing a law to establish a Bureau would require a single bill; the alternative would be to legislate every time a new AI issue arose. Rather than popping AI pimples, Congress could instead simply add a layer of topical cleanser – an AI Bureau – to cut the problem off at its source.

Creating a new bureau is also the more responsible course of action. Congress does not have expertise in the field of AI, but a new agency could draw upon the institutional knowledge of organizations like the FDA, NIST, the VA, and industry. In fact, the purpose of agencies is to serve as a repository of subject matter expertise and to help craft useful and scientifically valid policies. Furthermore, currently, Congress must dedicate its attention between thousands of different topics; a dedicated agency could give the topic the attention and care it deserves. Given the ubiquity of AI, it is only logical that Congress establish a Bureau to deal with these issues in a timely and expert manner.

Such a Bureau would need authority over all uses of AI. This would mean carving out jurisdiction from other agencies, or creating a provision for joint rulemaking – i.e., the FDA would create guidance for medical devices, but the AI Bureau would sign off on any rules that involved AI, and would keep a copy on their own website as well. The Bureau would also need an enforcement arm, and a way for individuals to complaint or leave comments. The Bureau's focus areas could include substantive fairness (ensuring that algorithms do not discriminate against people in an unacceptable way, including but not limited to protected classes), procedural fairness (ensuring that algorithms are auditable and it's clear how people's data is being used), and privacy. It could also handle more niche areas like deepfakes, or policy issues concerning the legality of tactics like "nudging" (using artificial intelligence to influence peoples' decisions.)

Admittedly, the National AI Initiative Act of 2020 has already established a coordinated program across the government to accelerate AI research [94]. However, it does not go far enough. Its purpose is merely to coordinate research activities within the government, and not commercially. It does not seem to have any authority of its own, and its founding statute explicitly says it does not hold any unique jurisdiction of its own. Furthermore, it has no enforcement power. As a result, it cannot safeguard the rights of people affected by AI software, and is insufficient – it must be replaced, or augmented, by an agency with the features described above.

This bill would be a perfect project for members of the AI Caucuses of the House and Senate, both of whom hope to support competitive innovation in AI while still protecting the rights of Americans [95], [96]. The bipartisan makeup of the caucus demonstrates that both sides believe this to be an important issue.

5. Conclusion

In a time where medical supplies, drugs, and hospital beds are in short supply, and hospital visits continue to increase, automated algorithmic software may prove the key to standardizing medical resource allocation systems, targeting inequity and ensuring that people who need the care the most get access to it. However, AI systems must be carefully designed and tested to ensure that they do not perpetuate the very errors they were meant to mitigate. Unfortunately, companies have every incentive to hide their

proprietary data and algorithms, making it incredibly challenging for external parties to validate the software. Furthermore, hospital triage and risk assessment systems are not currently covered by regulation, including the FDA's moderately well-developed guidance on artificial intelligence in medical device software.

To counter this problem, there are several steps that should be taken. First, President Biden should issue an executive order stating that the FDA's definition of a medical device should be construed to include medical resource allocation software, granting the FDA jurisdiction to regulate and approve these technologies. Medical associations that would use these technologies should also call for the development of AI systems, and search for a single system that they can endorse for adoption. Finally, the leaders of the AI Caucus should introduce and work with other Members of Congress to pass a bill that would create a new independent agency to handle all issues of artificial intelligence, creating a designated hub with the expertise and authority to handle this issue and all other questions of artificial intelligence as they continue to arise.

[1] “Washington Internships for Students of Engineering.” https://wise-intern.org/ (accessed Jun. 15, 2021).
[2] “About IEEE.” https://www.ieee.org/about/index.html (accessed Jun. 15, 2021).
[3] Oleksandr_Delyk, *Stethoscope on laptop keyboard. Concept 3D image*. [Online]. Available: https://www.shutterstock.com/image-illustration/stethoscope-on-laptop-keyboard-concept-3d-243153847
[4] Institute of Medicine, *Hospital-Based Emergency Care: At the Breaking Point*. Washington, DC: The National Academies Press, 2007. doi: 10.17226/11621.
[5] “Download: Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care | The National Academies Press.” https://www.nap.edu/download/12875 (accessed Jun. 18, 2021).
[6] A. Kumar, “Application of Machine Learning in Hospital Resource Allocation - A Survey,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. 3, pp. 505–508, Mar. 2021, doi: 10.22214/ijraset.2021.33276.
[7] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, vol. 366, no. 6464, p. 447, Oct. 2019, doi: 10.1126/science.aax2342.
[8] “Covid-19 is now in 50 countries, and things will get worse,” *The Economist*, Feb. 29, 2020. Accessed: Jun. 15, 2021. [Online]. Available: https://www.economist.com/briefing/2020/02/29/covid-19-is-now-in-50-countries-and-things-will-get-worse
[9] M. Conlen, J. Keefe, A. Sun, L. Leatherby, and C. Smart, “How Full Are Hospital I.C.U.s Near You?,” *The New York Times*, Dec. 16, 2020. Accessed: Jun. 15, 2021. [Online]. Available: https://www.nytimes.com/interactive/2020/us/covid-hospitals-near-you.html
[10] T. R. Brown, L. P. Francis, and J. Tabery, “Embedding the Problems Doesn’t Make Them Go Away,” *Am. J. Bioeth.*, vol. 20, no. 7, pp. 109–111, Jul. 2020, doi: 10.1080/15265161.2020.1779864.
[11] D. B. White and B. Lo, “Mitigating Inequities and Saving Lives with ICU Triage during the COVID-19 Pandemic,” *Am. J. Respir. Crit. Care Med.*, vol. 203, no. 3, pp. 287–295, Feb. 2021, doi: 10.1164/rccm.202010-3809CP.
[12] D. W. Wallace *et al.*, “An adapted emergency department triage algorithm for the COVID-19 pandemic,” *J. Am. Coll. Emerg. Physicians Open*, vol. 1, no. 6, pp. 1374–1379, 2020, doi: 10.1002/emp2.12210.
[13] R. Salway, R. Valenzuela, J. Shoenberger, W. Mallon, and A. Viccellio, “EMERGENCY DEPARTMENT (ED) OVERCROWDING: EVIDENCE-BASED ANSWERS TO FREQUENTLY ASKED QUESTIONS,” *Rev. Médica Clínica Las Condes*, vol. 28, no. 2, pp. 213–219, Mar. 2017, doi: 10.1016/j.rmclc.2017.04.008.
[14] K. Krisberg, “Study: Not enough hospitals addressing ER overcrowding,” *Nations Health*, vol. 46, no. 2, pp. E6–E6, Mar. 2016.
[15] “Trends in the Utilization of Emergency Department Services, 2009-2018,” *ASPE*, Mar. 03, 2021. https://aspe.hhs.gov/pdf-report/utilization-emergency-department-services (accessed Jun. 16, 2021).
[16] “FastStats,” Apr. 09, 2021. https://www.cdc.gov/nchs/fastats/emergency-department.htm (accessed Jun. 15, 2021).
[17] O. Ivanov *et al.*, “Improving ED Emergency Severity Index Acuity Assignment Using Machine Learning and Clinical Natural Language Processing,” *J. Emerg. Nurs.*, vol. 47, no. 2, pp. 265-278.e7, Mar. 2021, doi: 10.1016/j.jen.2020.11.001.
[18] J. Tolbert, A. D. P. Nov 06, and 2020, “Key Facts about the Uninsured Population,” *KFF*, Nov. 06, 2020. https://www.kff.org/uninsured/issue-brief/key-facts-about-the-uninsured-population/ (accessed Jun. 16, 2021).

[19] "Physician Appointment Wait Times up 30% from 2014." https://www.merritthawkins.com/news-and-insights/media-room/press/physician-appointment-wait-times-up-from-2014/ (accessed Jun. 16, 2021).
[20] "Study: Nearly 3 in 4 Americans Say It's Easier to Go to the ER Than to Get a Doctor's Appointment," *Zocdoc*. https://www.zocdoc.com/about/news/2019-er-report/ (accessed Jun. 16, 2021).
[21] M. C. Raven and F. Steiner, "A National Study of Outpatient Health Care Providers' Effect on Emergency Department Visit Acuity and Likelihood of Hospitalization," *Ann. Emerg. Med.*, vol. 71, no. 6, pp. 728–736, Jun. 2018, doi: 10.1016/j.annemergmed.2017.10.013.
[22] "QuickStats: Median Emergency Department (ED) Wait and Treatment Times,* by Triage Level† — National Hospital Ambulatory Medical Care Survey, United States, 2010–2011§." https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6319a8.htm (accessed Jun. 16, 2021).
[23] "National Hospital Ambulatory Medical Care Survey: 2016 Emergency Department Summary Tables," p. 38, 2016.
[24] M. Mazer-Amirshahi *et al.*, "U.S. drug shortages for medications used in adult critical care (2001–2016)," *J. Crit. Care*, vol. 41, pp. 283–288, Oct. 2017, doi: 10.1016/j.jcrc.2017.06.005.
[25] "Most Emergency Physicians Report Hospitals Lack Critical Medicines; Not 'Fully Prepared' for Disasters, Mass Casualty Incidents." http://www.emergencyphysicians.org/press-releases/2018/5-22-2018-90-percent-of-Emergency-Physicians-have-Experienced-Drug-Shortages-in-Past-Month (accessed Jun. 16, 2021).
[26] Gilboy, Nicki, Tanabe, Paula, Travers, Debbie, and Rosenau, Alexander, *Emergency Severity Index: A Triage Tool for Emergency Department Care*, 2020th ed. Schaumberg, IL: Emergency Nurses Association, ENA.
[27] "Value Transformation Framework," *NACHC*. https://www.nachc.org/clinical-matters/value-transformation-framework/ (accessed Jun. 17, 2021).
[28] A. B. Martin, M. Hartman, D. Lassman, A. Catlin, and The National Health Expenditure Accounts Team, "National Health Care Spending In 2019: Steady Growth For The Fourth Consecutive Year: Study examines national health care spending for 2019.," *Health Aff. (Millwood)*, vol. 40, no. 1, pp. 14–24, Jan. 2021, doi: 10.1377/hlthaff.2020.02022.
[29] G. W. Mitchell, "A Brief History of Triage," *Disaster Med. Public Health Prep.*, vol. 2, no. S1, pp. S4–S7, Sep. 2008, doi: 10.1097/DMP.0b013e3181844d43.
[30] E. Dippenaar, "Triage systems around the world: a historical evolution," *Int. Paramed. Pract.*, vol. 9, no. 3, pp. 61–66, Sep. 2019, doi: 10.12968/ippr.2019.9.3.61.
[31] "ESI," *enaorg*. https://www.ena.org/education/esi (accessed Jun. 16, 2021).
[32] R. Hong, R. Sexton, B. Sweet, G. Carroll, C. Tambussi, and B. M. Baumann, "Comparison of START triage categories to emergency department triage levels to determine need for urgent care and to predict hospitalization," *Am. J. Disaster Med.*, vol. 10, no. 1, pp. 13–21, 2015, doi: 10.5055/ajdm.2015.0184.
[33] "Sensitivity and specificity," *Wikipedia*. Jul. 01, 2021. Accessed: Jul. 12, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sensitivity_and_specificity&oldid=1031356310
[34] J. M. Zachariasse, V. van der Hagen, N. Seiger, K. Mackway-Jones, M. van Veen, and H. A. Moll, "Performance of triage systems in emergency care: a systematic review and meta-analysis," *BMJ Open*, vol. 9, no. 5, p. e026471, May 2019, doi: 10.1136/bmjopen-2018-026471.
[35] A. L. Huntley, R. Johnson, S. Purdy, J. M. Valderas, and C. Salisbury, "Measures of Multimorbidity and Morbidity Burden for Use in Primary Care and Community Settings: A Systematic Review and Guide," *Ann. Fam. Med.*, vol. 10, no. 2, pp. 134–141, Mar. 2012, doi: 10.1370/afm.1363.
[36] "Risk-Stratification Methods for Identifying Patients for Care Coordination," *AJMC*. https://www.ajmc.com/view/risk-stratification-methods-for-identifying-patients-for-care-coordination (accessed Jun. 17, 2021).

[37] "About the ACG System," *Johns Hopkins ACG® System*. https://www.hopkinsacg.org/advantage/ (accessed Jun. 17, 2021).
[38] A. Team, "Protocol for Responding to and Assessing Patient Assets, Risks, and Experiences (PRAPARE)," *AAPCHO*. https://aapcho.org/focusareas/prapare/ (accessed Jun. 18, 2021).
[39] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Mar. 2019, doi: 10.1186/s40537-019-0192-5.
[40] G. Saposnik, D. Redelmeier, C. C. Ruff, and P. N. Tobler, "Cognitive biases associated with medical decisions: a systematic review," *BMC Med. Inform. Decis. Mak.*, vol. 16, no. 1, p. 138, Nov. 2016, doi: 10.1186/s12911-016-0377-1.
[41] M. Gul and E. Celik, "An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments," *Health Syst.*, vol. 9, no. 4, pp. 263–284, Oct. 2020, doi: 10.1080/20476965.2018.1547348.
[42] J. Stewart, P. Sprivulis, and G. Dwivedi, "Artificial intelligence and machine learning in emergency medicine," *Emerg. Med. Australas.*, vol. 30, no. 6, pp. 870–874, Dec. 2018, doi: 10.1111/1742-6723.13145.
[43] Y. Berlyand *et al.*, "How artificial intelligence could transform emergency department operations," *Am. J. Emerg. Med.*, vol. 36, no. 8, pp. 1515–1517, Aug. 2018, doi: 10.1016/j.ajem.2018.01.017.
[44] "How could artificial intelligence benefit emergency medicine?," *AIMed*, Sep. 03, 2018. https://ai-med.io/ai-med-news/artificial-intelligence-emergency-medicine/ (accessed Jun. 18, 2021).
[45] The Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives for 2020, "Phase I Report: Recommendations for the framework and format of Healthy People 2020." U.S. Department of Health and Human Services. Accessed: Jun. 18, 2021. [Online]. Available: https://www.healthypeople.gov/sites/default/files/PhaseI_0.pdf
[46] "Leading Health Indicators - Healthy People 2030 | health.gov." https://health.gov/healthypeople/objectives-and-data/leading-health-indicators (accessed Jun. 18, 2021).
[47] 2021, "Disparities in Health and Health Care: 5 Key Questions and Answers," *KFF*, May 11, 2021. https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5-key-question-and-answers/ (accessed Jun. 18, 2021).
[48] K. M. Hoffman, S. Trawalter, J. R. Axt, and M. N. Oliver, "Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites," *Proc. Natl. Acad. Sci.*, vol. 113, no. 16, pp. 4296–4301, Apr. 2016.
[49] L. Lopes, C. Muñana, and 2020, "KFF/The Undefeated Survey on Race and Health - Main Findings," *KFF*, Oct. 14, 2020. https://www.kff.org/report-section/kff-the-undefeated-survey-on-race-and-health-main-findings/ (accessed Jun. 18, 2021).
[50] W. P. Qiao, E. S. Powell, M. P. Witte, and M. R. Zelder, "Relationship between racial disparities in ED wait times and illness severity," *Am. J. Emerg. Med.*, vol. 34, no. 1, pp. 10–15, Jan. 2016, doi: 10.1016/j.ajem.2015.08.052.
[51] A. Todorov *et al.*, "Gender differences in the provision of intensive care: a Bayesian approach," *Intensive Care Med.*, vol. 47, no. 5, pp. 577–587, May 2021, doi: 10.1007/s00134-021-06393-3.
[52] Y. Ding, E. Park, M. Nagarajan, and E. Grafstein, "Patient Prioritization in Emergency Department Triage Systems: An Empirical Study of the Canadian Triage and Acuity Scale (CTAS)," *Manuf. Serv. Oper. Manag.*, vol. 21, no. 4, pp. 723–741, Oct. 2019, doi: 10.1287/msom.2018.0719.
[53] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," p. 15.
[54] "Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots," *American Civil Liberties Union*. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 (accessed Jun. 21, 2021).

[55] A. Narla, B. Kuprel, K. Sarin, R. Novoa, and J. Ko, "Automated Classification of Skin Lesions: From Pixels to Practice," *J. Invest. Dermatol.*, vol. 138, no. 10, pp. 2108–2110, Oct. 2018, doi: 10.1016/j.jid.2018.06.175.
[56] J. K. Winkler *et al.*, "Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition," *JAMA Dermatol.*, vol. 155, no. 10, pp. 1135–1141, Oct. 2019, doi: 10.1001/jamadermatol.2019.1735.
[57] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLOS Med.*, vol. 15, no. 11, p. e1002683, Nov. 2018, doi: 10.1371/journal.pmed.1002683.
[58] M. A. Badgeley *et al.*, "Deep Learning Predicts Hip Fracture using Confounding Patient and Healthcare Variables," *ArXiv181103695 Cs*, Nov. 2018, Accessed: Jun. 21, 2021. [Online]. Available: http://arxiv.org/abs/1811.03695
[59] B. Nestor *et al.*, "Rethinking clinical prediction: Why machine learning must consider year of care and feature aggregation," *ArXiv181112583 Cs Stat*, Nov. 2018, Accessed: Jun. 21, 2021. [Online]. Available: http://arxiv.org/abs/1811.12583
[60] "Amazon scraps secret AI recruiting tool that showed bias against women | Reuters." https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (accessed Jun. 21, 2021).
[61] "Reducing Racial and Ethnic Algorithmic Bias and Ensuring that VA Algorithms are Fair," May 21, 2021. https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=3950 (accessed Jun. 21, 2021).
[62] "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.," *Washington Post*. Accessed: Jun. 21, 2021. [Online]. Available: https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/
[63] C. for D. and R. Health, "Overview of Device Regulation," *FDA*, Sep. 04, 2020. https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/overview-device-regulation (accessed Jun. 22, 2021).
[64] C. for D. and R. Health, "Learn if a Medical Device Has Been Cleared by FDA for Marketing," *FDA*, Nov. 2018, Accessed: Jun. 22, 2021. [Online]. Available: https://www.fda.gov/medical-devices/consumers-medical-devices/learn-if-medical-device-has-been-cleared-fda-marketing
[65] C. for D. and R. Health, "Regulatory Controls," *FDA*, Feb. 08, 2019. https://www.fda.gov/medical-devices/overview-device-regulation/regulatory-controls (accessed Jun. 23, 2021).
[66] C. for D. and R. Health, "Premarket Approval (PMA)," *FDA*, Jul. 09, 2019. https://www.fda.gov/medical-devices/premarket-submissions/premarket-approval-pma (accessed Jun. 23, 2021).
[67] C. for D. and R. Health, "Software as a Medical Device (SaMD)," *FDA*, Sep. 09, 2020. https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd (accessed Jun. 22, 2021).
[68] "'Software as a Medical Device': Possible Framework for Risk Categorization and Corresponding Considerations," p. 30, 2014.
[69] C. for D. and R. Health, "Artificial Intelligence and Machine Learning in Software as a Medical Device," *FDA*, Jan. 2021, Accessed: Jun. 23, 2021. [Online]. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

[70] O. of the Commissioner, “Classification of Products as Drugs and Devices and Additional Product Classification Issues,” *U.S. Food and Drug Administration*, May 12, 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/classification-products-drugs-and-devices-and-additional-product-classification-issues (accessed Jun. 22, 2021).
[71] “Clinical Decision Support Software - Draft Guidance for Industry and Food and Drug Administration Staff,” p. 27.
[72] “EMTALA Fact Sheet.” https://www.acep.org/life-as-a-physician/ethics--legal/emtala/emtala-fact-sheet/ (accessed Jun. 23, 2021).
[73] “Eberhardt v. City of Los Angeles | Cases | California | Westlaw Edge.” https://1.next.westlaw.com/Document/Iccf583d2919f11d993e6d35cc61aab4a/View/FullText.html?navigationPath=Search%2Fv1%2Fresults%2Fnavigation%2Fi0ad604ad0000017a39875da1edffeff3%3Fppcid%3D79e4ca1ebc1e436298c9d905d1bb16f7%26Nav%3DCASE%26fragmentIdentifier%3DIccf583d2919f11d993e6d35cc61aab4a%26parentRank%3D0%26startIndex%3D1%26contextData%3D%2528sc.Search%2529%26transitionType%3DSearchItem&listSource=Search&listPageSource=ac04c42869b2c5ffb7472dc5efe7acf5&list=CASE&rank=1&sessionScopeId=2f5f9d797d6480789747440b5aaef896bfb0e35efcd62f943c9defbae63c8d21&ppcid=79e4ca1ebc1e436298c9d905d1bb16f7&originationContext=Search%20Result&transitionType=SearchItem&contextData=%28sc.Search%29 (accessed Jun. 23, 2021).
[74] “Title VI Of The Civil Rights Act Of 1964 42 U.S.C. § 2000d Et Seq.,” Aug. 06, 2015. https://www.justice.gov/crt/fcs/titlevi-overview (accessed Jun. 23, 2021).
[75] “Alexander v. Sandoval | Cases | Supreme Ct. | Westlaw Edge.” https://1.next.westlaw.com/Document/I318fc9199c2511d9bdd1cfdd544ca3a4/View/FullText.html?originationContext=typeAhead&transitionType=Default&contextData=(sc.Default) (accessed Jun. 23, 2021).
[76] C. for D. and R. Health, “Reduced Medical Device User Fees: Small Business Determination (SBD) Program,” *FDA*, Jul. 2019, Accessed: Jul. 02, 2021. [Online]. Available: https://www.fda.gov/medical-devices/premarket-submissions/reduced-medical-device-user-fees-small-business-determination-sbd-program
[77] C. for D. and R. Health, “Breakthrough Devices Program,” *FDA*, Jan. 2021, Accessed: Jul. 02, 2021. [Online]. Available: https://www.fda.gov/medical-devices/how-study-and-market-your-device/breakthrough-devices-program
[78] “Electronic Code of Federal Regulations (eCFR),” *Electronic Code of Federal Regulations (eCFR)*. https://www.ecfr.gov/ (accessed Jul. 08, 2021).
[79] “Hospitals Prioritizing People Getting Elective Procedures Over Emergency Patients Because They’re More Financially Attractive,” *Kaiser Health News*, Feb. 25, 2019. https://khn.org/morning-breakout/hospitals-prioritizing-people-getting-elective-procedures-over-emergency-patients-because-theyre-more-financially-attractive/ (accessed Jul. 08, 2021).
[80] “21 U.S. Code § 321 - Definitions; generally,” *LII / Legal Information Institute*. https://www.law.cornell.edu/uscode/text/21/321 (accessed Jun. 22, 2021).
[81] C. A. Booker, “S.346 - 117th Congress (2021-2022): Black Maternal Health Momnibus Act of 2021,” Feb. 22, 2021. https://www.congress.gov/bill/117th-congress/senate-bill/346 (accessed Jul. 08, 2021).
[82] L. Underwood, “H.R.959 - 117th Congress (2021-2022): Black Maternal Health Momnibus Act of 2021,” Apr. 23, 2021. https://www.congress.gov/bill/117th-congress/house-bill/959 (accessed Jul. 08, 2021).
[83] R. Menendez, “Text - S.893 - 117th Congress (2021-2022): Tech to Save Moms Act,” Mar. 23, 2021. https://www.congress.gov/bill/117th-congress/senate-bill/893/text (accessed Jun. 25, 2021).

[84] E. B. Johnson, “H.R.937 - 117th Congress (2021-2022): Tech To Save Moms Act,” Feb. 09, 2021. https://www.congress.gov/bill/117th-congress/house-bill/937 (accessed Jul. 08, 2021).
[85] “CHEVRON, U.S.A., INC., Petitioner, v. NATURAL RESOURCES DEFENSE COUNCIL, INC., et al. AMERICAN IRON AND STEEL INSTITUTE, et al., Petitioners, v. NATURAL RESOURCES DEFENSE COUNCIL, INC., et al. William D. RUCKELSHAUS, Administrator, Environmental Protection Agency, Petitioner, v. NATURAL RESOURCES DEFENSE COUNCIL, INC., et al.,” *LII / Legal Information Institute*. https://www.law.cornell.edu/supremecourt/text/467/837 (accessed Jul. 08, 2021).
[86] “Priorities,” *The White House*. https://www.whitehouse.gov/priorities/ (accessed Jul. 02, 2021).
[87] “incubatED: Cultivating Solutions that Transform Emergency Medicine.” https://www.acep.org/administration/incubated/ (accessed Jul. 02, 2021).
[88] G. C. Peters, “Text - S.1353 - 117th Congress (2021-2022): Advancing American AI Act,” May 12, 2021. https://www.congress.gov/bill/117th-congress/senate-bill/1353/text (accessed Jun. 25, 2021).
[89] S. Horsford, “Text - H.R.1458 - 117th Congress (2021-2022): Unemployment Insurance Technology Modernization Act of 2021,” Mar. 01, 2021. https://www.congress.gov/bill/117th-congress/house-bill/1458/text (accessed Jun. 25, 2021).
[90] M. Takano, “Text - H.R.2438 - 117th Congress (2021-2022): Justice in Forensic Algorithms Act of 2021,” Apr. 08, 2021. https://www.congress.gov/bill/117th-congress/house-bill/2438/text (accessed Jun. 25, 2021).
[91] C. A. Coons, “Text - S.5052 - 116th Congress (2019-2020): Algorithmic Fairness Act of 2020,” Dec. 17, 2020. https://www.congress.gov/bill/116th-congress/senate-bill/5052/text (accessed Jun. 25, 2021).
[92] B. Frank, “Text - H.R.4173 - 111th Congress (2009-2010): Dodd-Frank Wall Street Reform and Consumer Protection Act,” Jul. 21, 2010. https://www.congress.gov/bill/111th-congress/house-bill/4173/text (accessed Jul. 02, 2021).
[93] “The Bureau,” *Consumer Financial Protection Bureau*. https://www.consumerfinance.gov/about-us/the-bureau/ (accessed Jul. 02, 2021).
[94] “The National Artificial Intelligence Initiative (NAII),” *National Artificial Intelligence Initiative*. https://www.ai.gov/ (accessed Jul. 12, 2021).
[95] “Artificial Intelligence Caucus,” *Artificial Intelligence Caucus*. https://artificialintelligencecaucus-olson.house.gov/ (accessed Jul. 02, 2021).
[96] “Portman, Heinrich Launch Bipartisan Artificial Intelligence Caucus,” *Senator Rob Portman*, Mar. 13, 2019. https://www.portman.senate.gov/newsroom/press-releases/portman-heinrich-launch-bipartisan-artificial-intelligence-caucus (accessed Jul. 02, 2021).