The Role of AI-Powered Decision-Making Technology in Medicare Coverage Determinations

January 2022

Table of Contents

Executive Summary

Artificial intelligence (AI)-powered decision-making tools have several applications in the health care field, from diagnosing patients to assisting commercial insurance brokers in the Medicare shopping process. This paper focuses on AI-powered tools that are involved in health care utilization management (UM). Such tools automate the medical review and prior authorization processes, direct post-acute care, and affect admission and discharge planning.

In the Center for Medicare Advocacy's experience, AI-powered decision-making tools may prompt providers to make decisions about the authorization or continuation of care that are more restrictive than Medicare coverage guidelines. Although most AI-powered decision-making tools claim to offer only recommendations that are not intended to substitute for clinical or medical judgment or for Medicare law, in the Center's experience, users often implement the tools' recommendations without any critical examination of their impact on patients. In other words, it is often as if the tools make the relevant decisions themselves. The use of these tools in UM processes may lead to premature terminations of coverage and care for Medicare beneficiaries, including denials that violate the coverage standard for skilled care that was clarified by the *Jimmo v. Sebelius* class action settlement.[1] While Medicare requires an individualized assessment of each beneficiary's qualification for coverage in certain care settings, AI- tools offer recommended decisions that are based on previous patient experiences. This neglects the nuance and individuality of the current patient's condition. Oftentimes, plans, providers, and beneficiaries do not fully understand the scope of these tools' development and use. Clinicians and plans rely on these tools with many questions left unanswered. This is partly due to the proprietary nature of UM-focused AI-powered decision-making tools, which prevents the public from understanding and challenging their results.

This paper centers around the following research questions:
1. To what extent do AI-powered decision-making tools that focus on UM supplant, rather than supplement, clinical decision making?
2. How widespread is the use of these tools in Medicare?
3. Are AI coverage criteria more restrictive than Medicare law and guidance?

This paper starts with an overview of AI-powered decision-making tools and their specific use in health care, detailing how the tools are intended to be complementary to the clinical judgment of medical professionals. Next, the paper explains increasing concerns about the use of such technologies in health care, delving into the absence of performance data on the tools' impact on persons with disabilities and their problematic use in Medicaid. The third section describes the use of these tools in Medicare, to the extent that it can currently be discerned, highlighting the Center's main concerns with regard to the *Jimmo* settlement and individualized assessment requirements, and outlining examples of litigation and the tools' use in different care settings. The paper concludes with five recommendations to consider in further research on this topic.

I. Introduction to Artificial Intelligence

A. General Overview of AI Tools and their Use in Health Care

The use of AI and machine learning has grown over the past few years across sectors. AI is the ability of computer systems and machines to simulate problem-solving and decision-making capabilities commonly performed by the human mind.[2] Machine learning is the application of AI to systems so they can automatically learn and improve through experience without explicit programming.

AI is used in a variety of settings, including for determining eligibility for public benefits and employment, automated test proctoring, predictive policing, and social media monitoring -- all areas that the Center for Democracy and Technology is currently examining.[a] In health care, AI's use is growing rapidly with the development of many AI-powered decision-making tools. A 2020 McKinsey Global Survey found that companies that find value in AI continue to invest in AI during the COVID-19 pandemic. Respondents in the health care services, pharmaceuticals, and medical products industries were most likely to report that their companies have increased investment in AI.[3] Both payers and providers ostensibly utilize AI-powered decision-making tools with the intention of lowering costs, optimizing resources, and improving the quality of patient care. Such tools can be used by providers to make diagnoses (e.g., deciding whether a patient should receive treatment based on the ability of an AI-powered decision-making tool to identify signs of sepsis) and by commercial insurance brokers in the Medicare shopping process (e.g., helping beneficiaries choose Medicare plans with the use of an AI-powered decision-making tool).

This paper, however, focuses on AI tools that affect health care coverage, specifically Medicare coverage. Such tools automate the medical review and prior authorization processes, direct post-acute care, and affect admission and discharge planning for hospitals, nursing facilities, and home health agencies. In other words, this paper examines AI-powered decision-making tools that focus on utilization management (UM), defined as "a set of techniques used by or on behalf of purchasers of health care benefits to manage health care costs by influencing patient care decision-making through case-by-case assessments of the appropriateness of care prior to its provision."[4]

Prior authorization is the process by which a provider, on behalf of a patient, requests approval from the patient's health plan or other payer before proceeding with a treatment or service. If approved, prior authorization makes coverage and payment by the health plan or entity certain.[5] In insurers' view, they use prior authorization to confirm that the care ordered by providers is consistent with clinical guidelines and covered by the patient's plan. Patient advocates, however, view prior authorization as a way for insurers to review care and decisions made by the patient's treating provider, oftentimes reducing payer costs. In Medicare, prior authorization is not usually used in traditional Medicare, but it is common in Medicare Advantage (MA) and Part D.[6] A Kaiser Family Foundation report states that MA plans often require prior authorization for relatively expensive services, such as inpatient hospital stays, skilled nursing facility stays, and Part B drugs, and that prior authorization is generally not required for preventive services.[7]

---

[a] See the Center for Democracy and Technology's work here.

Medical reviews are defined by the Centers for Medicare & Medicaid Services (CMS) as the clinical review and collection of medical records and other information to confirm that payment is only made for services that meet coverage, coding, billing, and medical necessity requirements.[8] These reviews are similar to prior authorizations but are usually conducted *after* a service or treatment has been given. Medical reviews are thus used frequently in the traditional Medicare program, where there is limited use of prior authorization and Medicare contractors generally review claims after services have been rendered. Among its medical review activities, CMS has implemented a Medicare Fee-for-Service Recovery Audit Program to identify and correct improper payments.[9] This work is done by Recovery Audit Contractors (RACs), who review claims on a post-payment basis. RACs detect and correct improper payments so that CMS, fiscal intermediaries, and Medicare Administrative Contractors (MACs) can try to prevent improper payments in the future.

The proprietary nature of UM-focused AI tools leads to concerns about transparency for medical professionals, patients, and product testing populations. Developers often hold exclusive rights to the algorithms and processes used to create AI products. UM-focused AI tools are also subject to only limited regulation. Developers of diagnostic proprietary AI-powered decision-making tools often do not seek approval from the Food and Drug Administration (FDA), as the 2016 21st Century Cures Act "was interpreted as taking most medical advisory tools out of the FDA's jurisdiction."[10] However, in guidelines released in the fall of 2019, the FDA stated that it will focus its oversight powers on "AI tools whose rationale cannot be independently evaluated by clinicians," or "tools that don't allow practitioners to see how they arrived at their conclusions."

Like clinicians, patients are also often uninformed about the use of the tool in their care and how the tool works, and more specifically, patients do not have access to the criteria on which coverage decisions are made. **Current discourse about the regulation of AI-powered decision-making tools focuses on tools used by providers to make diagnoses, rather than on tools that affect or determine Medicare coverage. While diagnostic tools fall into a "regulatory gray zone," regulation of the AI tools that affect or determine Medicare coverage is even more unclear. Why should some types of AI-powered decision-making tools (i.e., diagnostic) be subject to regulation while others (i.e., UM) are not?**

B. *The Purportedly Complementary Nature of Diagnostic AI Tools Compared to Utilization Management (UM)-Focused Tools*

Literature on diagnostic AI-tools defends their use by arguing that they are only complementary to a clinician's decision-making process and do not substitute for it. In a study testing a particular algorithm's ability to detect a decline in kidney function, Tomašev et al. state that their model to predict inpatient episodes of acute kidney injury is designed to complement existing routine care.[11] Karyn Baum, internist at M Health Fairview, argues that AI tools help clinicians make better decisions--similar to screening tools used to diagnose sepsis or to CT scans or lab values-- but that these tools do not take the place of clinician-made decisions.[12] Oftentimes, clinicians are hesitant to bring up the use of AI tools to patients as they claim the topic of AI will take time away from "actionable steps that patients can take to improve their health and quality of life" and as they worry that the mention of AI use can result in long discussions about the algorithm's

development, a process on which clinicians do not have all the information.[13] Yet, when the subject is brought up, clinicians generally emphasize that they make final care decisions, not the tools. The supposed complementary nature of these tools even affects the regulation of them. Because developers argue that their tools are a "decision aid," meaning the tools do not replace clinical judgment but rather help clinicians make decisions, developers usually do not seek FDA approval for their products.[14]

However, for AI-tools that focus on UM, available literature suggests that prior authorization decisions are made solely by the tools themselves. The Center has found that many Medicare Advantage plans use AI-tools such as InterQual AutoReview and Milliman Clinical Group's Cite AutoAuth.[15] Through InterQual AutoReview, medical reviews are automated and through InterQual Connect, prior authorizations are automated. This automation results in quick to nearly instantaneous approval or denial of requests. MCG's Cite AutoAuth performs similarly to InterQual AutoReview by allowing providers to submit prior authorization requests and receive automated determinations.[16] The instantaneous provision of a determination gives little-to-no time for any human to be involved in or give feedback for the request.

II. Growing Concern About the Use of AI

A. Limited Data on AI Tools' Performance and Social Determinants of Health

In addition to ambiguous regulation of UM-focused AI- tools, there is limited data on their performance, specifically data on how well their algorithms work as measured by improved health outcomes. There is also limited data on the patient demographics used in the testing stages.[17] STAT's 2021 report analyzing the role of AI in the health care field discusses how diagnostic tools do not perform consistently across diverse populations, particularly when disaggregated by race, sex, and geography.[b] Developers choose patient data on which to test their tools, and the demographics of these patient data sets are usually homogenous. Thus, tools perform well on the homogenous population they were tested on and perform poorly on demographics not represented in the test population.[18] Moreover, just as the algorithms of AI tools are considered trade secrets, the demographics of the data they are tested on are also treated as proprietary.[19] While the homogeneity of patient testing populations has been analyzed with regard to *diagnostic* AI- tools, **can it be inferred that this homogeneity also affects AI-powered decision-making tools that are used in UM?**

---

[b] STAT's 2021 report, "Promise and Peril: How artificial intelligence is transforming health care," serves as an in-depth analysis of diagnostic AI-powered decision-making tools. One of the articles in this report described how only seven out of 161 AI-powered decision-making tools cleared by the FDA in recent years include any information about the racial composition of their testing datasets (105). In a study testing the influence of gender imbalance in the medical imaging datasets which are used to train proprietary AI-powered decision-making tools, Larrazabal et al. found that their algorithm to diagnose 14 common thoracic diseases performed worse when using male patients for training and female patients for testing, and vice versa (Larrazabal et al., 2020: 12592). Further, overrepresentation of a sex in the training data stymied system performance. Another study, which involved looking through scientific literature on diagnostic AI tools, found only 56 studies which identify the geographic origin of the testing data used to train an image-based diagnostic tool in the United States (Brodwin and Ross, 2021, 111). These articles and studies show that using imbalanced testing data can reproduce already existing bias. Moving forward, both diagnostic and UM-focused AI-powered decision-making tools should consider this research and ensure improved performance on historically underrepresented patient/geographic populations.

B. *Impact on Persons with Disabilities*

AI-powered decision-making tools also appear to negatively impact people with certain chronic conditions. People with disabilities, for example, are often mistreated and discriminated against in health care. According to the World Health Organization, health services for disabled people are unfailingly of poorer quality and are under resourced compared to services provided to individuals without disabilities.[20] Further, the disability community is particularly harmed by AI-powered decision-making tools.[21] In conversation with Henry Claypool, the policy director at Brandeis University's Community Living Policy Center and founding Principal Deputy Administrator of the Administration for Community Living at the Department of Health and Human Services, he said:

> The experience of disability is not often well-represented in the training data set used by algorithms. Adding to the challenges of creating algorithms that treat disabled people fairly is the diversity of experiences that receive protections under civil rights law—building an automated tool used to make [a] determination about one's health care is a very sensitive matter that warrants great examination by health policy professionals.[22]

Tools that use quality-adjusted life years (QALYs) to measure clinical cost-effectiveness undervalue the lives of persons with disabilities in the following way: QALYs are calculated by asking persons *without* disabilities to place a value on the experience of having a disability, an experience non-disabled people cannot fully understand.[c] QALYs do not consider the complexity of experiencing life with a disability and thus reduce the value of treatments that fail to render patients "healthy" or "functioning." Vital treatments that may extend and/or improve the quality of life for persons with disabilities are thereby undervalued by AI tools that rely on QALYs. The inability of disabled people to achieve a particular standard of health – a standard that excludes the viewpoint of people with disabilities – can be viewed by AI-tool algorithms as low restoration potential that does not warrant coverage. If a particular treatment is costly due to more extensive need, for example, a tool may deny persons with disabilities necessary care as a cost-savings measure.

C. *Use in Medicaid*

The use of AI-powered decision-making tools in Medicaid raises concern, according to both the National Health Law Program (NHeLP)[d] and the Center for Democracy and Technology (CDT). In May of 2021, NHeLP's responded to the Agency for Healthcare Research and Quality's request for information on the potential of clinical algorithms to introduce racial and ethnic bias into health care delivery. This report discusses how AI- tools exacerbate the issue of "black box" health care decision-making. This is the idea that AI will predict the behavioral traits of an individual or make decisions without providing an explanation.[23] This type of decision-making is contrasted against "explainable AI," which refers to AI that is used to complement human

---

[c] For more information on QALYs, see Disability, National Council on Disability. *Quality Adjusted Life Years and the Devaluation of Life with Disability.* (November 6, 2019). https://ncd.gov/sites/default/files/NCD_Quality_Adjusted_Life_Report_508.pdf.

[d] NHeLP's work on Medicaid in their May 2021 letter to the Agency for Healthcare Research and Quality on the use of clinical algorithms that may introduce racial and ethnic bias into health care delivery helped influence the Center's work on the role of AI in Medicare.

activity and not replace it.[24] NHeLP writes that information on existing AI-powered decision-making tools is unavailable, as these tools are shielded by intellectual property protections; even the underlying study, training, and validation data are not publicly available.[25] Lack of transparency starts at the outset of these tools' journeys and is still present at the end when diagnoses or coverage decisions are made. A person denied care as a result of a recommendation made by a tool may simply be told the decision without any other information about how their specific medical circumstance triggered the denial.[26] This can prevent beneficiaries from appealing the denial. This issue, which NHeLP reports is negatively impacting Medicaid beneficiaries, is also affecting Medicare beneficiaries, as discussed below. Additionally, in the development process, program design decisions are often driven by the prioritization of cost-savings, clinician-usability, or performance.[27] Developers' desired outcomes and priorities affects everything about these tools. For example, while AI-powered decision-making tools focused on Medicaid eligibility generally prioritize efficiency and performance measures, advocates report that these tools are unable to identify errors and outliers because the reporting functions required to do so and the people affected by such errors "were not built into the system." NHeLP expresses concern over the top-down perspective of bias: NHeLP finds that much of the literature on AI-powered decision-making tools discusses bias at a broad level (e.g., bias on the basis of population/identity groups such as race, ethnicity, sexuality, gender, etc.) but fails to acknowledge that individuals, with their own characteristics and medical needs, are being denied medically necessary health.[28] Both instances of bias, at higher levels and from a subaltern perspective, are important.

CDT's October 2020 report discusses cases of state governments adopting AI-tools to assess eligibility for home-and community-based services (HCBS) under Medicaid.[29] The report centers an example of a budget allocation tool adopted by Idaho. Idaho adopted a new program to assess recipients' approved budgets for HCBS under Medicaid in 2011.[30] Recipients traveled to a medical assessment center where an Independent Assessment Provider (IAP) filled out a proprietary form scoring the recipient's need for assistance in feeding, toileting, dressing, and other functions. Once the IAP manually entered this data into a digital "Budget Tool," the Tool automatically calculated an "Assigned Budget Amount" for the reported needs based on data in the Tool's proprietary database. This "Assigned Budget Amount" could only be increased if the recipient required increased funds for their "health and safety," as determined by program managers. The term "health and safety" is undefined, which led to major funding cuts for recipients' individualized budgets, and subsequently, to lengthy and complex appeals. A court found that this "Budget Tool" was developed using a small and under-representative dataset. Because IAPs recorded large data sets from the medical assessment and manually entered them into the digital "Budget Tool," the judge concluded that there was a "high likelihood of human error." Idaho did not perform the annual recalibration required to ensure appropriate assessment of budget allocation. Further, Idaho's state agency did not show recipients the proprietary assessment form or allow recipients to access their assessment results. Without an audit process, there was no mechanism to confirm that budgets assigned by the Tool accurately met recipients' needs.

CDT also finds that other AI-powered decision-making tools make benefits determinations using "added eligibility criteria not required by law." They conclude that challenges to such tools have generally followed four primary legal theories:

1. AI-powered decision-making and proprietary algorithms may violate constitutional or statutory due process rights, including the right to a fair hearing, and to "ascertainable standards."[31]
2. AI-powered decision-making can violate federal or state notice requirements, under which agencies must explain to recipients why their benefits or eligibility status has changed.[32]
3. The *Olmstead* community integration mandate under the American with Disabilities Act can be jeopardized by benefits cuts induced by AI-powered decision-making. Persons with disabilities should be able to stay in their communities if they wish to, rather than enter an institution for care.[33]
4. The proprietary nature of AI tools should not prevent beneficiaries from accessing necessary information to understand and challenge decisions that affect them. Access to an algorithm alone and not to additional, more accessible information about the algorithm may not be enough to meaningfully to challenge the algorithm.[34]

The Center is also seeing AI tools used in Medicare, which raises concerns for patients, providers, and plans. In the Center's experience, Medicare providers regularly deny coverage to patients on the ground that the treatment is "maintenance only" citing patients' low restoration potential, despite *Jimmo v. Sebelius*, discussed below.[35] Research indicates that AI-powered decision-making tools may follow in the same footsteps as these Medicare providers, denying necessary treatment to particularly vulnerable and historically marginalized patients due to improvement concerns, cost-saving measures, or unrepresentative product development testing.

III. Use of AI-Powered Decision-Making Tools in Medicare

The use of AI tools in traditional Medicare is less prevalent and less clear than it is in Medicare Advantage. CMS encourages and, in some cases, requires[e] its contractors that conduct medical reviews of fee-for-service claims, such as Medicare Administrative Contractors and Recovery Audit Contractors, to use "prepayment and postpayment screening tools or natural language coding software" to identify claims with improper billing.[36] CMS does not mandate the use of a specific program for such screenings. Yet the agency has contracted directly with developers of AI tools to provide its audit contractors, Administrative Law Judges, and own departments with access to, for example, InterQual.[37] The Center has also observed the use of AI tools by providers in the traditional Medicare context.[38] For example, many hospitals use AI tools as part of their utilization review process to determine whether patients are classified as "inpatients" or as "outpatients" receiving observation services - a determination that affects Medicare coverage. Hospitals are not required to use any particular screening tool when making patient status determinations, but use many such tools to try to anticipate what Medicare will deem to be the correct patient status.

---

[e] To further clarify, MACs are "encouraged" by CMS to use screening tools and RACs are required to use screening tools to identify claims which are most likely to contain improper billing. In Chapter 3 of the Medicare Program Integrity Manual, section 3.2.1 states that MACs are "encouraged" to use such tools while RACs "shall use" such tools.

A. Overview of the Center's Concerns

Awareness of and efforts to challenge the use of AI-powered decision-making tools in Medicare are far behind those in Medicaid. Building on the work of NHeLP and CDT, it is long past time to focus on the use of such tools in Medicare. Many of the challenges AI tools pose in Medicaid are also evident in Medicare, notably, the proprietary nature of the tools and the resulting inability of beneficiaries to access necessary information for appealing and challenging decisions in which the tools were used. Unlike the state context where it is well-documented that Medicaid agencies use AI tools, the use of such tools in Medicare is more opaque. The Center's research did not discern precisely how widespread the use of these tools is in Medicare (especially in traditional Medicare), but marketing by the tools' manufacturers indicates that their use in Medicare is growing. NaviHealth, a post-acute care management company, states that it has served around nine million Medicare Advantage beneficiaries and is used by nearly 16,000 post-acute providers.[39] MyNEXUS, a benefits management service, claims to deliver clinical support services to nearly 1.7 million Medicare Advantage members across 20 states. [40]

In the Center's experience, AI tools make decisions that seem to be more restrictive than Medicare coverage guidelines. While Medicare is adamant that no claim should be denied based on a screening tool alone, the Center sees cases where beneficiaries do not receive necessary care when an AI tool is used to determine their care options.[41] The use of these tools in UM processes may prematurely terminate care and coverage for Medicare beneficiaries. The Center is often told that such tools are "just a screen," but providers often act as if they are the law. In practical terms, these tools appear to be used like prior authorization devices in that if providers don't follow their results, they believe they will not be paid. The Center has encountered use by providers in Medicare Advantage plans. For example, it is our understanding that Aetna contracts with myNEXUS to provide decisions to Medicare-certified home health agencies on which services may be provided to beneficiaries. Under the home health (HH) benefit, Medicare guidance states that determinations of whether HH services are reasonable and necessary should be based on an individualized assessment of each beneficiary's care needs and that denial of services based on the recommendations of "numerical utilization screens, diagnostic screens, diagnosis or specific treatment norms" is inappropriate.[42]

As these tools are proprietary, challenging the decisions they produce is difficult. AI tools seem to supplant rather than supplement clinicians' own decision-making. Often, the decisions recommended by such tools, notably for providers to terminate care prematurely or for plans to not cover a treatment based on previous patient data, result in cost-savings for plans. CDT finds that there may be an underlying and masked motivation for states to decrease the costs of public benefits programs; as a result, policymakers can obscure budget cuts under the banner of efficiency.[43] Like CDT, the Center shares concern about the prioritization of cost-savings by these tools. Absent the use of these tools, the Center asserts that more care would likely be covered since such tools err on the side of prematurely terminating or outright denying coverage. However, quantifying how much more care would be covered is difficult. The Center finds that the use of AI tools in coverage determinations has undermined proper coverage for skilled care clarified by the *Jimmo* settlement and may not comply with individualized assessment requirements in certain care settings. These topics are analyzed in the following sections.

1) Jimmo v. Sebelius

The 2013 *Jimmo v. Sebelius* settlement clarified that Medicare covers skilled nursing and therapy services when they are needed to maintain a beneficiary's current condition or to prevent or slow further deterioration.[44] Coverage does not depend "on the presence or absence of an individual's potential for improvement, but rather on the beneficiary's need for skilled care."[45] The settlement re-emphasized what was already provided for by regulation: restoration potential is not the deciding factor in determining whether skilled care is required.[46] Skilled nursing or therapy services are coverable when an individualized assessment of the beneficiary's clinical condition indicates that the specialized judgment, knowledge, and skills of a nurse or therapist are necessary to safely and effectively deliver services.[47] The settlement applies in the skilled nursing facility, home health, and outpatient physical therapy settings.

Despite this settlement, the Center finds that many providers continue to inappropriately deny services, and Medicare often inappropriately denies coverage, for skilled care when it is required for non-improvement goals such as maintenance or preventing deterioration. Medicare covers skilled care equally for improvement, maintenance, and slowing decline of a condition, yet in our experience, AI-powered decision-making tools tend to limit coverage on the grounds of improvement, similar to the actions of many providers.

2) Individualized Assessment Requirements

Medicare Part A provides coverage for inpatient hospital services, including long-term care hospitals and rehabilitation hospitals, in addition to coverage for skilled nursing facility (SNF) care, home health care, and hospice care.[48] Individualized assessments are essential in Medicare coverage determinations as they indicate if skilled care is required for patients to maintain or prevent further deterioration of their condition.

Medicare provides limited coverage for SNF services.[49] Part A covers up to 100 days per benefit period if the patient has had a qualifying three-day hospital stay and requires daily skilled care.[50] In addition, the skilled care must be related to the condition for which the patient was hospitalized.[51] Under the SNF benefit, skilled nursing services are covered if an individualized assessment of a patient's clinical condition indicates that the specialized judgment, knowledge, and skills of a registered nurse or licensed practical (vocational) nurse are required.[52] Direct skilled therapy services, including skilled physical therapy, occupational therapy, and speech/language pathology therapy, are covered if an individualized assessment of the patient's clinical condition shows that the specialized judgment, knowledge, and skills of a qualified therapist are necessary to perform the rehabilitation service.[53] The same applies to maintenance therapy in an SNF.[54]

In Medicare's home health (HH) benefit, skilled nursing services are covered if an individualized assessment indicates that the patient's clinical condition requires the specialized judgment, knowledge, and skills of a registered nurse, or a licensed practical (vocational) nurse ("skilled care").[55] Under the HH benefit, Medicare guidance states that determinations of whether HH services are reasonable and necessary should be based on an individualized assessment of each beneficiary's care needs and that denial of services based on the recommendations of "numerical utilization screens, diagnostic screens, diagnosis or specific treatment norms" is inappropriate.[56]

Skilled therapy services to perform a maintenance program are covered if an individualized assessment of the patient's clinical condition shows that the specialized judgment, knowledge, and skills of a qualified therapist or qualified therapist assistant under the supervision of a qualified therapist are necessary to perform a safe and effective maintenance program.[57] Coverage of skilled speech-language pathology and skilled occupational therapy services follow the same rules as for skilled therapy services to perform a maintenance program.[58]

Even when an individualized assessment of a patient's clinical condition indicates that, for example, the specialized judgment of a nurse is required, AI-powered decision-making tools can recommend that clinicians do not pursue further treatment on the grounds that the patient may not improve or that the treatment is too costly.[59] These decisions directly contradict guidance stating that Medicare coverage does not depend on a patient's restoration potential but rather on their need for skilled care to maintain their current condition or prevent/slow further deterioration.

B. Medicare Screening Tools and Litigation

While there has not been as much litigation challenging these tools in Medicare as in Medicaid, there have been some relevant decisions made about screening tools in Medicare. The 1986 decision in *Fox v. Bowen* found that it is contrary to Medicare regulations to use informal presumptions or "rules of thumb" applied across the board without considering the medical condition or therapeutic requirements of the individual patient.[60] The 1989 decision in *Vorster v. Bowen* found that a "utilization screen" regarding frequency of use could be used as a "guide" throughout the claim review process rather than as the direct basis of denial of claims. Further, beneficiaries must be notified of the basis for denials and given the opportunity to argue for medical necessity.[61] In 1989, the parties in *Hooper v. Sullivan* agreed that denials of admissions, services, and/or Medicare coverage based on "numerical utilization screens, diagnostic screens, diagnosis, specific treatment norms, the 'three hour rule,' or other 'rules of thumb,'" in the context of rehabilitation hospitals are inappropriate.[62] Additionally, the court ordered the Medicare agency to issue an agreed-upon bulletin explaining that the "three hour rule"[f] is only intended to be used as a screening criterion to identify those cases which require further review by the Medicare contractor.[63] These guidelines should never be used as the sole basis to deny coverage. While these three cases, all of which the Center were involved in, pre-date the use of AI-powered decision-making tools in health care, they are still "good law" and their principles hold.

C. Examples in Specific Care Settings

The Center has encountered the use of such tools in the SNF and HH settings. The following example shows that coverage decisions issued using naviHealth software might not comply with the standard clarified by *Jimmo v. Sebelius* in the SNF setting. The Center learned about the daughter of a Medicare beneficiary who spent five days in a hospital for cellulitis and was then discharged to a SNF. At the required SNF plan-of-care meeting (7 days after discharge from the hospital), the daughter was given a "naviHealth Predict Outcome Report" projecting her

---

[f] The "three-hour rule" defined medical necessity for rehabilitation hospital patients to include requiring and receiving at least three hours per day of physical and/or occupational therapy. *Hooper*, 1989 WL 107497, at *1.

mother's success at rehabilitation in the SNF "against 'a patient database of over 3 million,' and determining her end-date for rehab. The SNF informed us that 'it was out of their hands' and Mom would be discharged because she was not making 'significant progress.'"[g] The daughter cited the *Jimmo* settlement, but the SNF was adamant that naviHealth's report determined when her mother would be discharged and that the SNF had to comply.

We have encountered the greatest use of AI tools by Medicare-certified home health agencies (HHAs). A HHA shared with the Center a chart from myNEXUS, a post-acute benefits management company that uses AI to "highly automate the home visit authorization process, speed time to care, increase provider effectiveness and improve member satisfaction." [64] The chart, which was to be used with patients in a certain Medicare Advantage plan the HHA contracted with, listed various diagnoses and the corresponding amount of care to be provided for each diagnosis (e.g., patients who had a stroke receive three nurse visits).[65] The chart appeared highly prescribed in terms of what was approved for each diagnosis; it did not allow for any home health aides for any condition, even though Medicare covers reasonable and necessary aide services for beneficiaries who also have a skilled care need. The HHA reported that over a span of two years it constantly requested additional services to be approved by myNEXUS, but all requests were denied. The myNEXUS algorithm, which is generally used by BCBS and Humana, two major players in the Medicare Advantage industry, appears to only authorizes a certain number of services.[66] If a HHA wants more services to be authorized, they must seek approval from myNEXUS, which can be difficult to obtain, as demonstrated above. The Center has also encountered corporate home health chains that have developed their own in-house algorithm/tool. Professional staff at these agencies have shared their concerns about having to strictly adhere to "the model" even if, in their professional opinion, more services are necessary. They report being instructed by their agencies to tell beneficiaries that the allotted benefits are "all Medicare will cover."

IV. Recommendations/Further Inquiry

Some experts in the field of AI and health care, such as Satish Gattadahalli, director of digital health and health informatics at Grant Thornton Public Sector, recommend that AI technologies should continue to complement and augment clinician care.[67] The Center suggests a different approach, however. While AI proliferates in various settings and applications, the focus of AI moving forward should be to overcome some of its inherent problems, such as correcting biases that result in poor performance among diverse and historically underrepresented populations. The Center is concerned that tools focusing on UM err on the side of cost-savings rather than on streamlining UM or improving the quality of patient care. The prioritization of cost-savings can be one of many factors contributing to AI tools' inefficacies. Rather than adopting a passive approach of ensuring that AI tools continue to complement clinicians, the Center advocates for adopting a proactive approach of fixing large-scale issues as such tools continue to be used in clinical settings. The Center offers the following policy recommendations to mitigate the previously discussed issues posed by AI-powered decision-making tools:

1. CMS should conduct better oversight and enforcement of its own policies, as the current widespread use of AI may violate stated policy and appears to be growing unchecked.

---

[g] The Center received this communication in July of 2019.

As outlined in the "Individualized Assessment Requirement" section above, the use of AI-powered decision-making tools to determine whether certain services are covered is leading to inappropriate denials and terminations of coverage in the Medicare program. The increased use of such tools appears to violate Medicare's own rules requiring individualized assessments and prohibiting rules of thumb. CMS should engage in more oversight of the use of these tools, including analysis and transparency measures outlined below.

2. Policymakers should ensure transparency concerning the use of AI decision-making tools at all stages.

The entire "life-cycle," or each stage of AI-powered decision-making tools, from the identification of an issue to be addressed by the tool and the many stages of development and implementation, to the continuing stages of implementation and quality control, should be transparent and undergo regulation.[68] Each of these stages require decisions that are opportunities for bias. While discourse on bias in AI-powered decision-making tools is growing, NHeLP writes that a review of current biases in such tools is incomplete until there is transparency at each stage that these tools undergo. Policy approaches focused on mitigating bias in AI-powered decision-making tools through improved transparency should center the individuals who both supply training data and are affected by the tools' decisions.[69] At the outset, AI developers must be transparent about the justification for developing an AI-powered decision-making tool, the processes used to develop the tool, the selection of patient data for training, the regulatory process the tool undergoes before implementation, and the quality control measures put in place after implementation. In other words, the proprietary nature of these tools should be questioned, and policymakers and other relevant actors should push to ensure that the label "proprietary" is not used as a defense to prevent the disclosure of actionable information.

NHeLP highlights that the justification for why building AI-powered decision-making tools requires transparency: "Often, [automated decision-making tools] ADS are created under the banner of increased efficiency and objectivity, but the underlying rationale behind the initial decision may instead prioritize cost-savings at the expense of patients' needs."[70] "The use of ADS may even be more the result of successful efforts by sales people or lobbyists about the purported benefits of a given ADS."[71]

Currently, the proprietary nature of UM-focused AI-powered decision-making tools prevents providers and patients from challenging the guidelines and coverage criteria of such tools. On the contrary, National Coverage Determinations (NCDs) and Local Coverage Determinations (LCDs), decisions by Medicare Administrative Contractors (MACs) whether to cover a certain service in accordance with Medicare's reasonable and necessary criteria, are published by MACs for use by the public and medical community.[72] As such, aggrieved parties, such as a Medicare beneficiary, can initiate a review of a NCD or LCD.[73] The public nature of NCDs and LCDs allows the public to challenge guidelines that affect them. The public should be able to challenge the coverage criteria and decisions of AI-powered decision-making tools in this same manner.

3. Due process requires full disclosure.

   Those affected by the use of AI-powered decision-making tools should be aware of how a tool's decision was made through the provision of complete, accessible information about the algorithm. The *Goldberg v. Kelly* Supreme Court decision serves as the foundation for the principle that beneficiaries must be provided information that allows them to understand the reasons for actions that have affected their benefits.[74] The Court held that when "the means to obtain essential food, clothing, housing, and medical care" are at stake, beneficiaries are entitled to "timely and adequate notice detailing the reasons for termination" and a hearing where the "benefits recipient can present evidence and confront adverse witnesses."[75] All approaches to AI-powered decision-making must ensure that affected individuals can understand the decision that the tool made, have access to all necessary information (including the criteria used by the decision tool), and have the ability to appeal decisions they disagree with.[76] The appeals process should be clear and accessible, including clear timelines, an opportunity to examine the outcome and justification for the outcome, the ability to show a necessary service outside of the tools' guidance, and a timeline for a decision.[77]

   NHeLP writes that "the impact of not having the information necessary to challenge the decision of an ADS can be incredibly significant. A denial of health care services can lead to poorer health outcomes including death."[78]

4. CMS should study or encourage study by other groups such as the Government Accountability Office (GAO) or the Office of the Inspector General (OIG) regarding the use of AI by plans and providers in making coverage decisions.

   To confirm that AI-powered decision-making tools do not conflict with Medicare guidance, CMS or another federal entity should conduct research on how these tools are developed and used in patient care. Further, this research should illuminate the processes by which seemingly more restrictive coverage guidelines are used by such tools. The findings of this research should be available for review by the public as the proprietary nature of these tools results in limited-to-no availability of public information. While it appears that UM-focused AI-powered decision-making tools do not complement, but rather replace, the decisions of clinicians, this new research can inspect the unclear and apparently lax regulatory process that these tools undergo. Diagnostic AI-powered decision-making tools generally complement the decisions of clinicians and thus, developers of these tools often do not seek FDA approval. **Why is it that UM-focused tools which seem to fully replace the judgment of clinicians might be subject to even less regulation than complementary diagnostic tools?** UM AI tools should be subject to FDA approval to ensure that they make decisions that conform with Medicare coverage law.

5. CMS should require plans and providers to disclose to individuals and report to CMS when: 1) they rely on AI-powered clinical decision-making tools, 2) if such tools are used, for what services, 3) denial rates when tools are used, and 4) subsequent appeal rates.

14

The apparent under-regulation of UM-focused AI-powered decision-making tools warrants a more rigorous approach, especially when it comes to the information that affected patients can access. Informing both patients and CMS of the four topics listed above will allow patients to access important information that can be useful in the appeals process. It will also allow CMS to better inform its guidance and to market the new technology used in its programs to beneficiaries.

Conclusion

UM-focused AI-powered decision-making tools are increasingly used by plans and providers to make coverage decisions, without an opportunity for affected individuals to analyze the tools' coverage criteria. The proprietary nature of these tools leaves plans, providers, and patients with many questions unanswered regarding the development and use of such tools. This paper finds that the use of AI-powered decision-making tools in UM processes may prematurely terminate care and coverage for Medicare beneficiaries, often in violation of the standard for coverage set forth in *Jimmo v. Sebelius*. In the Center's experience, when the recommendations made by such tools are actually visible, their coverage criteria are often more restrictive than Medicare law and guidance. Medicare guidance often requires an individualized assessment of each beneficiary's qualification for coverage, yet AI-powered decision-making tools offer decisions that are based on previous patient experiences, ignoring the nuance and individuality of the current patient. These general, blanket "rule of thumb" decisions are used to determine beneficiaries' access to coverage and care and often conflict with clinical judgment, harming both beneficiaries and providers. Currently, providers often treat tools' coverage criteria and decisions as law when they are in fact not law. As more is learned, it will be important for policymakers to manage and, where appropriate, rein in the influence of AI-powered decision-making tools in Medicare.

January 2022, Lyla Saxena

Lyla Saxena (B.A., Dec. 2021), was a health policy intern at the Center for Medicare Advocacy from May 2021 through December 2021.

Endnotes:

[1] No. 5:11-CV-00017-cr (D. Vt.) (settlement approved Jan. 24, 2013); settlement agreement available at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/Downloads/Jimmo-Settlement-Agreement.pdf.

[2] "What Is Machine Learning? A Definition." expert.ai, May 6, 2020, https://www.expert.ai/blog/machine-learning-definition/.

[3] "The State of AI in 2020." McKinsey & Company, November 17, 2020, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020.

[4] Bradford H. Gray and Marilyn J. Field. *Controlling Costs and Changing Patient Care? The Role of Utilization Management.* Washington (DC): National Academies Press (US), 1989. doi:10.17226/1359. https://www.ncbi.nlm.nih.gov/books/NBK235000/.

[5] Addressing Commercial Health Plan Abuses to Ensure Fair Coverage for Patients and Providers. (American Hospital Association, December 2020). https://www.aha.org/white-papers/2020-12-01-addressing-commercial-health-plan-abuses-ensure-fair-coverage-patients-and.

[6] American Hospital Association, "Addressing Commercial Health Plan Abuses."

[7] "Medicare Advantage in 2021: Premiums, Cost Sharing, out-of-Pocket Limits and Supplemental Benefits." Kaiser Family Foundation, June 21, 2021, https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2021-premiums-cost-sharing-out-of-pocket-limits-and-supplemental-benefits/.

[8] "Medical Review and Education." Centers for Medicare & Medicaid Services, November 11, 2020, https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/Medical-Review.

[9] "Medicare Fee-for-Service Recovery Audit Program." Centers for Medicare & Medicaid Services, December 1, 2021, https://www.cms.gov/research-statistics-data-and-systems/monitoring-programs/medicare-ffs-compliance-programs/recovery-audit-program.

[10] Erin Brodwin and Casey Ross. *Promise and Peril: How Artificial Intelligence Is Transforming Health Care.* (STAT and The Commonwealth Fund, 2021). https://www.statnews.com/wpcontent/uploads/2021/04/STAT_Promise_and_Peril_2021_Report.pdf.

[11] Nenad Tomašev et al., "A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney Injury." *Nature* 572, no. 7767 (2019): 116-19. https://doi.org/10.1038/s41586-019-1390-1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6722431/.

[12] Brodwin and Ross, *Promise and Peril*, 164.

[13] Ibid, 163-165.

[14] Ibid, 36.

[15] *Highmark Post-Acute Care Management for Medicare Advantage Members Frequently Asked Questions.* (Highmark, September 6, 2019), 1, https://content.highmarkprc.com/Files/Region/PA-WV/CareMgmtProg/PostAcuteCareMgmt/navihealth-faqs.pdf; *naviHealth: Post-Acute Care Management for Blue Medicare Members Frequently Asked Questions for Providers*, 1, (Blue Cross and Blue Shield of North Carolina, December 18, 2014). https://www.bluecrossnc.com/document/navihealth-provider-faqs; *naviHealth Coordinating Prior Authorizations for Highmark Members*, 1, (Anthem Blue Cross, July 21, 2016). https://bluecrosscamedicarerx.com/ca/shared/f2/s2/t1/pw_g254931.pdf?refer=provider; "Priority Health's Post-Acute Strategy Led to Shorter SNF Stays, Fewer Readmits." *Medicare Advantage News*, July 27, 2017. https://www.navihealth.com/wp-content/uploads/MedicareAdvantageNews_July27.pdf.

[16] "Cite AutoAuth." MCG Health, 2021, https://www.mcg.com/how-we-help/health-plans/cite-evidence-based-clinical-decision-support-for-payers/cite-autoauth/.

[17] Brodwin and Ross, *Promise and Peril*, 169.

[18] Brodwin and Ross, *Promise and Peril*, 116; Agostina K. Larrazabal et al. "Gender Imbalance in Medical Imaging Datasets Produces Biased Classifiers for Computer-Aided Diagnosis." Proceedings of the National Academy of Sciences 117, no. 23 (2020): 12592-93. https://doi.org/10.1073/pnas.1919012117. http://www.pnas.org/content/117/23/12592.abstract.

[19] Brodwin and Ross, *Promise and Peril*, 104.

[20] "Disability and Health." World Health Organization, December 1, 2020, https://www.who.int/news-room/fact-sheets/detail/disability-and-health.

[21] Elizabeth Edwards et al., *RE: RFI on the Use of Clinical Algorithms That Have the Potential to Introduce Racial/Ethnic Bias into Healthcare Delivery.* (National Health Law Program, May 4, 2021), 14, https://healthlaw.org/wp-content/uploads/2021/06/NHeLP-AHRQ-RFI-Comments.pdf.

[22] Henry Claypool, email to Lyla Saxena, October 19, 2021.

[23] "The AI Black Box Explanation Problem." European Research Consortium for Informatics and Mathematics, January 22, 2019, https://ercim-news.ercim.eu/en116/special/the-ai-black-box-explanation-problem; "Managing the Black Box of Artificial Intelligence (AI)." Deloitte Touche Tohmatsu Limited, https://www2.deloitte.com/us/en/pages/advisory/articles/black-box-artificial-intelligence.html.

[24] European Research Consortium for Informatics and Mathematics, "The AI Black Box."

[25] Edwards et al., *RE: RFI on the Use of Clinical Algorithms*, 2.

[26] Ibid, 4.

[27] Ibid, 7.

[28] Ibid, 4.

[29] Lydia X.Z. Brown et al., *Challenging the Use of Algorithm-Driven Decision-Making in Benefits Determinations Affecting People with Disabilities*. (Center for Democracy & Technology, October 21, 2020), 6, https://cdt.org/insights/report-challenging-the-use-of-algorithm-driven-decision-making-in-benefits-determinations-affecting-people-with-disabilities/.

[30] Brown et al., *Challenging the Use of Algorithm-Driven Decision-Making*, 7.

[31] Ibid, 8, 10, 13.

[32] Ibid, 9, 11.

[33] Ibid, 9, 17.

[34] Ibid, 9, 14.

[35] Judith A. Stein, Alfred J. Chiplin, and Kata M. Kertesz. *Medicare Handbook, 2021 Edition*. Wolters Kluwer, 2020, (3-4).

[36] "Verifying Potential Errors and Taking Corrective Actions." Chap. 3 In *Medicare Program Integrity Manual*, 1-113: Centers for Medicare & Medicaid Services, September 30, 2021, 8, § 3.2.1.

[37] "CMS to Continue Use of Interqual Criteria." Business Wire, December 19, 2016, https://www.businesswire.com/news/home/20161219005102/en/CMS-to-Continue-Use-of-InterQual-Criteria; "What You Need to Know About the Utilization Review Process." Simplify Compliance LLC June 5, 2020, https://www.psqh.com/analysis/what-you-need-to-know-about-the-utilization-review-process/.

[38] Alice Bers and Kathleen Holt. Interview by Lyla Saxena. Conversational. Zoom, July 16, 2021; Alexander v. Azar, --- F. Supp. 3d ---, ---, 2020 WL 1430089, *19-*20 (D. Conn. Mar. 24, 2020).

[39] "Our Purpose: Enabling Seniors to Live a More Fulfilling Life." naviHealth, Inc., 2021, https://navihealth.com/about/our-purpose/.

[40] "About." myNEXUS Inc., 2021, https://www.mynexuscare.com/about/; "Windrose Health Investors Completes Sale of MyNEXUS to Anthem." Cision US Inc., April 29, 2021, https://www.prnewswire.com/news-releases/windrose-health-investors-completes-sale-of-mynexus-to-anthem-301279991.html.

[41] *Alexander*, 2020 WL 1430089, at *14-*15; "Home Health Services." Chap. 7 In *Medicare Benefit Policy Manual*, 1-106: Centers for Medicare & Medicaid Services, November 6, 2020, 22, § 20.3.; Centers for Medicare & Medicaid Services, "Verifying Potential Errors," 8, § 3.2.1.

[42] Centers for Medicare & Medicaid Services, "Home Health Services," 22, § 20.3.

[43] Brown et al., *Challenging the Use of Algorithm-Driven Decision-Making*, 21.

[44] Jimmo v. Sebelius, No. 5:11-CV-00017-cr, Doc. # 82-1 at 11-13 (D. Vt.) (settlement approved Jan. 24, 2013), (hereinafter "Jimmo Settlement").

[45] Centers for Medicare & Medicaid Services, "Home Health Services," § 40.2.

[46] 42 C.F.R. § 409.32(c).

[47] Jimmo Settlement at 11-13.

[48] 42 C.F.R. § 406.2 (Scope).

[49] Stein, Chiplin, and Kertesz, Medicare Handbook, 2021 Edition, (3-3).

[50] 42 U.S.C. § 11395d(a)(2)(A), 42 U.S.C. § 1395x(h).

[51] 42 U.S.C. § 1395f(a)(2)(B); 42 C.F.R. § 409.31(b).

[52] 42 C.F.R. §§ 409.32(c) and 409.33(c)(5), 42 C.F.R. §§ 409.32 and 409.33(a), (b); "Coverage of Extended Care (SNF) Services under Hospital Insurance." Chap. 8 In *Medicare Benefit Policy Manual*, 1-57: Centers for Medicare & Medicaid Services, October 4, 2019, 32, § 30.3.

[53] Centers for Medicare & Medicaid Services, "Coverage of Extended Care," 33-34, § 30.4.

[54] Ibid, 36-37, § 30.4.1.2.

[55] Centers for Medicare & Medicaid Services, "Home Health Services," 42, § 40.1.1.

[56] Ibid, 22, § 20.3.

[57] Ibid, 68, § 40.2.1.

[58] Ibid, 73, § 40.2.3, 74, § 40.2.4.

[59] Brown et al., *Challenging the Use of Algorithm-Driven Decision-Making*, 21.

[60] Fox v. Bowen, 656 F. Supp. 1236, 1248 (D. Conn. 1986).

[61] Vorster v. Bowen, 709. F. Supp. 934, 942-47 (C.D. Cal. 1989).

[62] Hooper v. Sullivan, No. H-80-99 (PCD), 1989 WL 107497, *1 (D. Conn. 1989).

[63] *Hooper*, 1989 WL 107497, at *1.

[64] Alice Bers and Kathleen Holt, Interview with Lyla Saxena.

[65] "Anthem, Inc. To Acquire myNEXUS, Home Health Benefits Manager." Business Wire, March 24, 2021, https://www.businesswire.com/news/home/20210324005316/en/Anthem-Inc.-To-Acquire-myNEXUS-Home-Health-Benefits-Manager; Alice Bers and Kathleen Holt, Interview with Lyla Saxena.

[66] Alice Bers and Kathleen Holt, Interview with Lyla Saxena.

[67] Brodwin and Ross, *Promise and Peril*, 208.

[68] Edwards et al., *RE: RFI on the Use of Clinical Algorithms*, 3.

[69] Ibid, 31.

[70] Ibid, 6-7.

[71] Ibid, 7.

[72] "National and Local Coverage Determinations - What You Need to Know." American Association of Clinical Endocrinology, https://pro.aace.com/practice-management/national-and-local-coverage-determinations-what-you-need-know; Stein, Chiplin, and Kertesz, *Medicare Handbook, 2021 Edition*, (6-62), (6-63).

[73] Stein, Chiplin, and Kertesz, *Medicare Handbook, 2021 Edition*, (6-63), (6-64).

[74] Brown et al., *Challenging the Use of Algorithm-Driven Decision-Making* 10.

[75] Ibid, 10.

[76] Edwards et al., *RE: RFI on the Use of Clinical Algorithms*, 5.

[77] Ibid, 34.

[78] Ibid, 33.