Artificial intelligence for health insurance: A proposed framework for FDA oversight

Renée Sirbu¹, Jessica Morley¹, Luciano Floridi¹,²

1. Yale Digital Ethics Center, Yale University, 85 Trumbull Street, New Haven, CT 06511
2. Department of Legal Studies, University of Bologna, Via Zamboni 22, Bologna, Bo 40100

Abstract

Despite mounting enthusiasm regarding the introduction of artificial intelligence (AI) software as a medical device (SaMD) to clinical care and, consequently, the development of a new regulatory proposal for the federal oversight of AI/ML medical devices, little attention has been paid to the oversight of AI tools used by large insurers. The U.S. Food and Drug Administration (FDA) has advanced an “Action Plan” for clinical AI (CAI) governance. However, the U.S. healthcare system remains threatened by the unregulated application of insurance AI (IAI). In this article, we use IAI tools in the Medicare Advantage (MA) prior authorization pathway as an illustrative case to argue that these technologies require further regulatory attention by the FDA. Specifically, we propose a redefinition of “medical device” under the 21st Century Cures Act as necessarily inclusive of IAI and advance an actionable framework for FDA oversight in the approval of IAI tools for deployment by large healthcare insurers.

Introduction

According to the Population Reference Bureau, the proportion of Americans over 65 years of age has grown at a rate “unprecedented in U.S. history.”¹ An aging population imposes greater demands upon the healthcare system and, consequently, upon healthcare insurers. Both providers and insurers are struggling to cope with this increased demand and are turning to artificial intelligence (AI) as a solution.²³ While significant attention has been paid to the governance, ethical, legal, and social implications (GELSI) of clinical AI (CAI),⁴ far less attention has been paid to the GELSI of healthcare insurance AI (IAI). Since IAI tools are not considered “software as a medical device” (SaMD), they are not subject to regulatory oversight by the Food and Drug Administration (FDA), nor are they covered by upcoming and proposed governance pathways targeted at CAI, such as the 2021 AI/ML SaMD Action Plan.

Justification for the lack of regulatory oversight surrounding IAI may arise from the mistaken assumption that IAI harbors no direct impact on patient care. However, the use of IAI in the Medicare Advantage (MA) prior authorization pathway, for example, may directly determine the treatment coverage to which a patient is entitled. IAI may, therefore, raise just as many ethical concerns as CAI. The literature suggests that the use of IAI has already increased MA denials to
the significant disadvantage of vulnerable patient groups. Thus, in this article, we argue that the same level of scrutiny exercised over AI-SaMD (i.e., CAI) should also be applied to IAI. We use IAI deployed in the MA prior authorization pathway as an illustrative case. Specifically, we propose a redefinition of “medical device” under the 21st Century Cures Act as necessarily inclusive of IAI and advance an actionable framework for FDA oversight in the approval of IAI tools for use by large healthcare insurers.

Medicare (Dis)advantage
Medicare is the U.S. federal health insurance program for those aged 65 and older, younger people with disabilities, and individuals with end-stage renal disease or ALS. Traditional Medicare (TM) covers hospital care (Part A) and outpatient care (Part B) and allows beneficiaries to opt-in to prescription drug coverage (Part D) for an additional premium. TM covers most, but not all, of the costs associated with approved and undeniably necessary (i.e., not preventative) healthcare services, and there is no cap on “out-of-pocket” costs to patients. Medicare Advantage (MA, or Part C) plans, provided by private insurers contracted with the federal government, “bundle” extra benefits—like hearing, vision, and dental coverage—and prescription charges otherwise not included in TM. MA plans also provide spending caps on out-of-pocket expenditures. Therefore, it is unsurprising that MA plans are becoming increasingly popular; according to the Center for Medicare Advocacy, Medicare Advantage (MA) enrollment now accounts for greater than 50% of the total Medicare population, and overpayments to MA plans have surpassed $75 billion annually.

Unlike TM, which is provided by the federal government, MA plans are offered by private insurers, which usually require patients to gain approval through prior authorization before start-of-care. Prior authorization requirements often prove problematic, delaying or denying vulnerable populations access to care that they would have received under TM. According to a 2022 report by the Inspector General of the Department of Health and Human Services, in 2019, 13% of MA plan prior authorization denials would not have been denied under TM. It is within this prior authorization pathway that MA insurers are most keen to deploy IAI. For example, IAI can be used to predict the length and intensity of treatment that a patient is likely to require, and, therefore, to guide insurers in making coverage decisions. In theory, IAI may expedite the review of MA claims, increasing accuracy and cost-effectiveness, and leading to improved quality of care and better patient outcomes.

Rather than improving the quality of care provided to MA beneficiaries by ensuring that covered treatments are appropriately tailored to individual patient interests, IAI may increase the
extent to which care is prescriptive (i.e., dictated by inflexible rules) rather than personalized. For instance, a popular tool deployed by one large insurer uses data compiled from a pool of six million patients to map the outcomes of people most similar to the individual submitting the claim. When a claim is submitted, the tool compares the patient to one of its “representative” cases and approves only the treatment that resulted in a positive outcome for the case most similar to the claimant. To clarify with an illustration, if Claimant A matches Case B in terms of condition and basic demographics, and Case B was hospitalized for ten days, then the IAI tool will suggest that Claimant A should only be approved for ten days of coverage in-hospital. Thus, case outcomes are applied by the insurer to predict when they can cut off treatment payments.

IAI tools may make epistemically flawed recommendations that could directly impact individual patient care, based on generic averages that do not account for differences in individual patient clinical characteristics, demographics, or even clinical judgment, when providing recommendations, thereby producing unfair or discriminatory outcomes. Treatment predictions proposed by IAI tools based on misguided evidence may not only increase claims denials but also decrease the overall quality of care afforded to MA patients.

Given the lack of both regulatory oversight and transparency in the development of IAI tools, there is presently no robust federal mechanism to ensure that tools used to assess claims have been externally and rigorously validated, nor is there any mechanism to guarantee that IAI tools are accurate (in terms of specificity and sensitivity) or fair (i.e., not biased). These are not just hypothetical concerns. In 2020, the purchase of a proprietary algorithm by one of the largest Medicare Advantage insurers in the United States correlated with a significant increase in denial of care claims. Providers that attempted to appeal denial decisions claimed that the proprietary algorithm relied on “rigid,” black-box criteria that did not appear to reflect public Medicare criteria for denial. No scientific studies assessing the performance of this algorithm were published—it is unclear to what extent (or if at all) these tests were performed. Furthermore, the proprietary nature of this algorithm limited the extent to which practitioners could appeal decisions successfully, resulting in significant tensions between payers and providers, and leaving patients stranded.

If these problems are allowed to persist, patients will shoulder the burden. Protecting patient interests requires mitigating these negative side-effects of IAI, meaning that IAI must be recognized for what it is: software as a medical device. IAI tools must be subject to rigorous independent verification, validation, and evaluation that guarantees accuracy (in terms of specificity and sensitivity) and fairness (in terms of access and outcomes). As a result, we argue that IAI must be encompassed within the legal purview of the FDA.
Towards an IAI SaMD Action Plan

The FDA’s AI/ML-SaMD Action Plan,\textsuperscript{13} published in 2021, aims to address stakeholder concerns regarding the governance of AI/ML SaMD, which were advanced in a 2019 FDA discussion paper.\textsuperscript{14} The Action Plan incorporates elements of the IMDRF Framework for Risk Categorization of Software as a Medical Device,\textsuperscript{15} the Good Machine Learning Principles (GMLP),\textsuperscript{16} the total product lifecycle (TPLC) framework introduced in the Digital Health Software Precertification (PreCert) Pilot Program,\textsuperscript{17} and the FDA’s risk-benefit framework and principles outlined in the software modifications guidance.\textsuperscript{18} To show potential avenues of inclusion for IAI, in Table 1 we highlight the applicable provisions in each of these documents and make a series of suggested amendments to these provisions to ensure that the FDA: (i) expands the multi-fold regulatory definition of “medical device” within the 21st Century Cures Act to include IAI; (ii) considers creating a new premarket clearance pathway for IAI used in the denial of MA claims; and (iii) recognizes the role that IAI maintains in both informing and driving clinical management. We include a rationale for each of the recommended amendments.
<table>
<thead>
<tr>
<th>Guidance Document</th>
<th>Applicable Provisions</th>
<th>Amendments and Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD&amp;C Act (21 USC 321(h))</td>
<td>Section 201(h) statutory definition of a medical device, under (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.”</td>
<td>Include IAI under the statutory definition of a “medical device.” Do not include IAI under exemptions to information blocking.</td>
</tr>
<tr>
<td>21st Century Cures Act (21CCA)</td>
<td>Section 513(f)(2), De Novo Classification (see Pre-Cert Pilot Program). Section 520(o)(1)(E), on Clinical Decision Support Software. Section 3060(a) (21 U.S.C. § 360j(o)6) excludes five categories of SaMD from the FD&amp;C Act “medical device” definition, including: (E) “a subset of Clinical Decision Support (CDS) software.” CDS software is excluded from the definition of a “medical device” and exempt from FDA regulation if “it is not the intent that such health care professional rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient.”</td>
<td>Despite the intended use of IAI to act as a CDS tool, the actualized use of IAI has proven to directly impact the course of treatment by dictating coverage. Regardless of their intended use, the current regulatory definition of a “medical device” under the FD&amp;C Act does not accurately capture the application of IAI tools. Given that IAI: (i) is used in the diagnosis and treatment of disease in man, (ii) is not a subset of CDS software, and (iii) is primarily relied upon by providers and health systems to make significant decisions regarding diagnosis, outcomes, and treatment, IAI should be reclassified as a medical device and subject to appropriate federal oversight.</td>
</tr>
<tr>
<td>IMDRF risk categorization framework for SaMD</td>
<td>Provides recommendations to identify the risk category of a SaMD based on the impact of its output on healthcare decisions. Includes four classes (I-IV) corresponding to the risk apparent to the health of an individual or the public, depending on whether the information provided by the SaMD is being used to treat or diagnose (II-IV), drive clinical management (I-III), or inform clinical management (I-II), and according to the severity of the health state or situation.</td>
<td>Categorize information the IAI provides as “[driving] clinical management.” The highest risk category would correspond to level III, under a critical state, or level I, under a non-serious state. This categorization follows from the above amendments to the 21CCA.</td>
</tr>
<tr>
<td>GMLP guiding principles</td>
<td>Ten guiding principles jointly identified by the FDA, Health Canada, and the MHRA in the promotion and effective use of AI/ML SaMD. The FDA encourages the “harmonization of Good Machine Learning Practice development” via extensive collaboration with international guidance communities and the Agency’s own internal Medical Device Cybersecurity Program.</td>
<td>Promote in advancement of TPLC approach (see below).</td>
</tr>
<tr>
<td>FDA benefit-risk framework</td>
<td>Principles considered in weighing benefits and risks during the premarket review of medical devices.</td>
<td>Promote in advancement of TPLC approach (see below).</td>
</tr>
<tr>
<td>FDA software modifications guidance risk management principles</td>
<td>Guidance in determining at which point changes or updates to SaMD might require the developer to submit a new 510(k).</td>
<td>Promote in advancement of TPLC approach (see below).</td>
</tr>
<tr>
<td>FDA Pre-Cert Pilot Program TPLC Approach</td>
<td>Address limitations that the iterative nature of SaMD imposed on premarket clearance by introducing an expedited review process for companies with a “demonstrated culture of quality.” Implemented under the De Novo classification process, instead of requiring Class III premarket approval; participants may be evaluated under an Excellence Appraisal. SaMD safety and effectiveness evaluated across the total product lifecycle (TPLC).</td>
<td></td>
</tr>
<tr>
<td>Redefine what it means for an insurer to demonstrate a “culture of quality” in the context of a TPLC framework. Insurers using IAI must: (i) maintain GMLP in development and deployment and report on how these standards are being met, (ii) update algorithms on the basis of appeals data (which requires active post-market surveillance by the insurer). (i) While the algorithm itself may be proprietary, companies must be transparent in identifying the parameters used in outcome prediction (i.e., training data, Medicare data, and the patient’s own MA plan characteristics). (ii) Urges greater collaboration between physicians and insurers; active monitorization of appeals by companies as a requirement for algorithmic updating should, in theory, shorten the length of the appeals process.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Proposed IAI as SaMD Action Plan
Conclusion and Future Considerations

It may be argued that IAI is not fundamentally SaMD, because it is not directly deployed for medical purposes, given that insurance is considered administrative, rather than clinical. We suggest that IAI is considered either SaMD, or an entirely novel category of software that influences the course of treatment, and, therefore, must be regulated through a new premarket approval pathway. In either case, it should not be slotted under the existing AI/ML SaMD framework.

It may also be objected that IAI tools are intended for internal improvements in efficiency (i.e., in handling large volumes of prior authorization requests), and not for primary decisions in clinical settings. This argument is unconvincing because the predictions generated by these tools may be—and indeed often are—used prescriptively in-clinic to dictate treatment coverage. Thus, despite their intended use, their actual use may directly affect care delivery.

A further complaint may be that more stringent regulation would stifle innovation. However, as historically emphasized, innovation is not always (or necessarily) good; Zeppelins, for example, were once lauded as the innovative future of aviation. Innovation in itself should never be privileged over patients’ health. The locus of the efficiency issue in insurance claims review lies in the volume of prior authorization requests; the long-term solution is to lighten the back-end burden on reviewers by reducing the number of requests, not by creating a system that denies these requests more quickly.

A final consideration that may be offered to resist the identification of IAI as SaMD involves an additional “human-in-the-loop” safety mechanism to review predictions generated by IAI, in a bid to avoid regulatory oversight. This objection, too, does not hold. Existing SaMD may already include a “human-in-the-loop,” and yet, do not eschew federal governance. Most importantly, this “human-in-the-loop” would have to be responsible for making the final decision regarding treatment coverage. The appeals process could no longer rely on the safety of its “proprietary” nature; therefore, accountability over these predictions would fall upon the human decision-maker.

The use of AI-SaMD to improve efficiency will likely keep growing, given the burden placed upon the U.S. healthcare system by an increasingly elderly, multi-morbid population. It may be a welcome trend. When novel technologies can save lives or improve patient outcomes, it is ethically imperative to investigate their potential and eventually invest in their development and deployment. However, implementation must not occur prematurely or inappropriately, at the cost of patient safety and quality of care. Encouragingly, the FDA recognizes that contemporary medical device regulation concerning adaptive technologies standards is insufficient, and the
Agency is taking positive action to ensure that the use of AI-SaMD is appropriately regulated. This is a necessary first step to protecting patients—and the broader healthcare system itself—from the harms that may arise from adopting unregulated AI-SaMD. To be comprehensive, the FDA must expand its AI/ML-SaMD Action Plan to encompass regulatory oversight over proprietary IAI tools. Expanding the Action Plan, as we have suggested, is critical to ensure better outcomes for patients. The FDA cannot afford to wait, as it did in the early 20th century, when multiple fraudulent and harmful devices and drugs (including height-stretching machines and elixir sulfanilamide) entered the market, unregulated. Regulatory oversight of proprietary IAI is urgent and critical to ensure better care outcomes. The rapid development of these tools requires a timelier response by federal agencies. Better care through technology is perfectly achievable with the proper regulation in place.
References

15. IMDRF Software as a Medical Device (SaMD) Working Group. ‘Software as a Medical Device’: Possible Framework for Risk Categorization and Corresponding Considerations. (2014).
23. FDA. Deciding When to Submit a 510(k) for a Software Change to an Existing Device: Guidance for Industry and Food and Drug Administration Staff. (2017).