

BIOETHICS AND INFORMATION TECHNOLOGY

# The Ethics of Algorithms in Healthcare

Christina Oxholm<sup>1</sup>, Anne-Marie S. Christensen<sup>1\*</sup> and Anette S. Nielsen<sup>2</sup>

<sup>1</sup>Department for the Study of Culture, Faculty of Humanities, University of Southern Denmark, 5230 Odense, Denmark

<sup>2</sup>Clinical Alcohol Research Unit, Department of Clinical Research, Faculty of Health, University of Southern Denmark, 5230 Odense, Denmark

\*Corresponding author. Email: [amsc@sdu.dk](mailto:amsc@sdu.dk)

## Abstract

The amount of data available to healthcare practitioners is growing, and the rapid increase in available patient data is becoming a problem for healthcare practitioners, as they are often unable to fully survey and process the data relevant for the treatment or care of a patient. Consequently, there are currently several efforts to develop systems that can aid healthcare practitioners with reading and processing patient data and, in this way, provide them with a better foundation for decision-making about the treatment and care of patients. There are also efforts to develop algorithms that provide suggestions for such decisions. However, the development of these systems and algorithms raises several concerns related to the privacy of patients, the patient–practitioner relationship, and the autonomy of healthcare practitioners. The aim of this article is to provide a foundation for understanding the ethical challenges related to the development of a specific form of data-processing systems, namely clinical algorithms.

**Keywords:** ethics; clinical decision support algorithms; clinical decision-making algorithms; data; privacy; professional autonomy; black box; patient-centered; patient–practitioner relationship

## Introduction

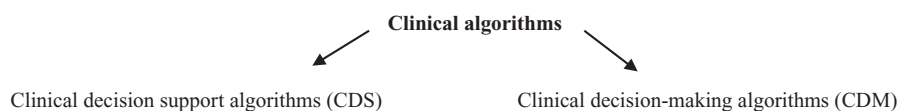
The healthcare sector is undergoing a significant evolution with the development and implementation of digital technology in healthcare practice. This evolution has the potential to fundamentally change the way medicine is practiced, as computer science methods are becoming more and more integrated in healthcare. The digitalization of the healthcare system together with an increasing demand for documentation and the development of new electronic methods for storing data have given rise to a rapid growth in the amount of patient data available, particularly in the form of electronic health records (EHR). However, the increase in available patient data is becoming a problem for healthcare practitioners, as they are often unable to fully survey and process the data relevant for the treatment or care of a patient. The reasons for this problem are sometimes practical in nature, such as a lack of time or resources, but the reasons can also be fundamental in nature, such as when the available data is too numerous and complicated for a human to cognitively assess it.

This growing amount of healthcare data has significantly expanded the amount of health data available to utilize in the development, training, and customization of a variety of automatic algorithmic systems designed to improve healthcare services. While digital technology has the potential to improve healthcare quality, it is important to determine and discuss the ethical challenges related to implementing this technology, particularly data-processing algorithms, in healthcare practice.

Formally, an algorithm is a purely mathematical construction designed to solve a prespecified problem. In other words, an algorithm is the electronic version of an advanced mathematical formula constructed to solve a certain problem in a predetermined number of steps. However, the most common way an algorithm is used does not focus so much on the algorithm's *mathematical* content but rather on

the problem it is implemented to solve<sup>1</sup>—the algorithm’s *what* rather than the algorithm’s *how*. Algorithms can thus be viewed from two different perspectives: (1) the formal, mathematical perspective and (2) the purpose-oriented perspective. In this article, we will examine algorithms from the perspective of ordinary discourse and focus on the applications and purposes of algorithms in healthcare (the *what* of the algorithms). The algorithms’ formal, mathematical content, and the ethical issues related to the technical development and implementation of software will not be examined.

In clinical practice, algorithms are used for clinical-related problem solving. These algorithms are called clinical algorithms.<sup>2,3,4,5</sup> The layout and scope of a clinical algorithm can vary depending on the problem it is designed to solve. However, all clinical algorithms serve a common purpose: to improve health and the healthcare system.<sup>6</sup> This article distinguishes between two types of clinical algorithms: (1) clinical decision support algorithms (CDS algorithms) and (2) clinical decision-making algorithms (CDM algorithms). CDS algorithms refer to algorithms that provide health professionals “(...) knowledge and person-specific information, intelligently filtered and presented at appropriate times, to enhance health and health care.”<sup>7</sup> The knowledge provided by CDS algorithms can vary depending on the purpose of the algorithm. A CDS algorithm can (1) give recommendations and instructions regarding medication and treatment,<sup>8,9,10,11</sup> (2) conduct preventive screening for particular diseases,<sup>12,13</sup> or (3) assist the diagnostic process.<sup>14,15,16</sup> CDM algorithms are, in principle, the same type of algorithm as CDS algorithms except for one essential aspect; CDM algorithms do not offer recommendations or assist practitioners but *make clinical decisions* on their own. We make this distinction because the different functions of CDS and CDM algorithms have different ethical implications in practice. While algorithms that can make decisions on their own are not yet implemented in healthcare practice (as CDS algorithms are), the ethical implications of these next-level algorithms should be investigated because of rapid advances in AI technology. So, while CDM algorithms may not currently be used in healthcare practice, they can influence healthcare in ethically problematic ways, which must be considered when developing these algorithms. The ethical issues associated with CDM algorithms are already at the center of the debate regarding whether digital technology will make health practitioners obsolete in the future.<sup>17</sup>



Like all other algorithms, clinical algorithms<sup>18</sup> depend on a dataset (input) to perform problem solving. Most often, clinical algorithms are developed to make use of extensive amounts of health data. This health data can be categorized into the following categories: (1) observational data, (2) laboratory data, and (3) administrative data.<sup>19</sup> Observational data is data that health professionals enter into EHRs (e.g., health professionals’ observations about factors relating to the patient’s health, diseases, and courses of treatment). Laboratory data is typically biological material that can be subject to testing (e.g., blood, genes, chromosomes, DNA, MR scans, and X-rays). Administrative data is data *about* a patient’s contact with the healthcare system (e.g., attendance, type of consultation, and frequency of visits).<sup>20</sup>

Clinical algorithms thus use data that has already been collected and stored by the healthcare entity, such as data in the existing EHR. This means that for the use of the algorithm to be justifiable, the data in use must be collected and stored according to existing legal and formal standards within the health system, such as the legal framework of General Data Protection Regulation (GDPR) in the European Countries,<sup>21</sup> national law, and national and local rules concerning informed consent for the collection of data. These questions concerning the collection and storage of data must be answered before clinical algorithms are used; however, because such questions are not particular to clinical algorithms, they will not be discussed further in this article.

## Ethical Approach

In our view, clinical algorithms are in of themselves not ethically good or bad. Rather, ethical benefits and challenges arise in relation to specific uses of these algorithms. Therefore, it is key to investigate how *using* algorithms in certain ways and contexts can cause potential ethical challenges that must be considered when developing and implementing algorithms.

Methodologically, we work in accordance with a bottom-up, rather than a top-down, approach in our ethical analysis. We have therefore refrained from choosing a specific ethical theory prior to and independently from our discussion of the specific ethical challenges arising from the use of clinical algorithms, although the discussion is conducted within the overall framework of the four basic principles of biomedical ethics: autonomy, beneficence, nonmaleficence, and justice.<sup>22</sup> However, without specification, the principles are too crude to account for the specific challenges that arise with the use of clinical algorithms in healthcare.

We identified three groups of ethical challenges related to the use of clinical algorithms in healthcare. First, the use of clinical algorithms causes a flow of data and information about patients through the healthcare system that may threaten patients' privacy. Second, if the output of clinical algorithms is presented or perceived not as recommendations but as obligatory instructions that must be followed, then the output may unduly influence or even restrict health practitioners' professional autonomy and discretion. Finally, the interventions of clinical algorithms may affect both the current and future patient-practitioner relationship, as a third agent in this relationship.

In this article, we analyze, from the bottom-up, the important ethical considerations involved in these challenges and suggest how CDS and CDS algorithms have the potential to influence fundamental elements in healthcare. The challenges identified are not exhaustive but are only some of the central ethical considerations that must be addressed when developing and implementing clinical algorithms.

## Clinical Algorithms and Patient Privacy

Clinical algorithms<sup>23</sup> draw on health data that are already stored in electronic systems, and, as we noted above, there are obvious gains to patient beneficence through the use of automated clinical algorithms. However, potential breaches of privacy related to data use in clinical algorithms arise because data flows from one context in healthcare to another. This means that there are constant changes in the circumstances, people, and type of healthcare relations in which the data is used. This flow of data is a general problem with many of the digital tools currently implemented in healthcare. As Fairweather and Rogerson note, "the use of databases is diminishing the practical extent of privacy and confidentiality. It may be that the *right* has not been diminished, but evidence suggests that the extent of *respect* of the right is declining."<sup>24,25</sup> The ethical tension between patient beneficence and privacy related to the use of clinical algorithms arises because clinical algorithms enable healthcare practitioners to discover information about a patient that may be useful in the care and treatment of that patient, but that information may not have been disclosed by the patient to that particular healthcare professional, or the patient may not find that information relevant or appropriate in that specific healthcare context.<sup>26</sup>

There is widespread agreement that the ethical concern for privacy is a concern for "the ability to determine for ourselves when, how, and to what extent information about us is communicated to others."<sup>27,28,29,30</sup> Privacy is concerned with the collection, storage, and use of personal information, and in the context of health, the concern for privacy is pertinent because much of the collected information is personal and/or sensitive in nature. The rules and practices of medical confidentiality and informed consent are established to allow for the use of sensitive data in healthcare, which can infringe upon ordinary privacy norms but are important for providing quality treatment and care to patients. Still, the notion of privacy is pivotal to examine questions not just about when data can be collected, but also about "the justifications, if any, under which data collected for one purpose can be used for another (secondary) purpose. An important issue in privacy analysis is whether the individual has authorized particular uses

of his or her personal information.”<sup>31</sup> However, questions about what justifies clinical reuse of data are varied and complicated.

In the development and use of clinical algorithms, it is necessary to weigh the concern for patient beneficence against the concern for patient privacy. In healthcare, the concern for the right to privacy is most often acknowledged through an obligation on the part of the healthcare practitioners to obtain informed consent from the patient. However, the increased flow of information has caused serious challenges to the existing practices of informed consent, and these challenges can be used to argue for two fundamental changes in these practices.

First, informed consent practices must be changed so that the patient is aware of the flow of data through the healthcare system. As Cato et al. note, “the patient is obligated to know that even though they might have chosen not to disclose a characteristic (e.g., transgender status, history of drug use or mental illness, or other potentially sensitive issues), software may still uncover it and present that information to their clinician.”<sup>32</sup> Informed consent must live up to a principle of openness in relation to the collection and reuse of patient data, because knowledge of these processes is necessary if patients are to make autonomous decisions about the use of their data in health care.

Second, practices of informed consent must be changed so that patients can distinguish between *different forms of uses of data* and *uses of data in different contexts* within the healthcare system. One suggestion is to model practices of informed consent on principles of “fair information,” which, in addition to openness, are principles of limitation of collection, limitation of disclosure, and limitation of use, security, and access.<sup>33</sup> The goal is, in Fairweather’s and Rogerson’s words, that “[a] patient should normally have effective control over his/her data and the ability to prevent any casual distribution that might be harmful.”<sup>34</sup> This would be in line with a background notion of privacy as concerned with the patient’s ability to determine when, how, and to what extent information and data are accessible to others—in this context, to healthcare practitioners.<sup>35</sup> However, these measures must be complemented with safeguards that ensure patient beneficence to prevent critical patients’ interests from being harmed by previous limitations established in patients’ consent to data use. Examples of such safeguards could be that health professionals will be licensed to use all clinical algorithms—even those blocked by restrictions in reuse of data—in situations where (1) a patient is unconscious, or (2) a patient is in some form of immediate and serious danger, as both situations make it impossible to ask for renewed consent from the patient.

A remaining problem is that changes in the practices of informed consent may not be enough to secure patient privacy in relation to the flow of data caused by using clinical algorithms. As noted by Mittelstadt et al., “[o]paque decision-making by algorithms [...] inhibits oversight and informed decision-making concerning data sharing. Data subjects cannot define privacy norms to govern all types of data generically because their value or insightfulness is only established through processing.”<sup>36</sup> The use of clinical algorithms is currently giving rise to a situation in which it is impossible for both professionals and patients to fully assess and predict the possible uses of patient data within the healthcare system. In such a situation, it is impossible for patients to fully explicate which uses of data they do and do not find acceptable in terms of privacy. In a similar vein, it will also sometimes be impossible for patients to foresee when previously recorded data may become vital for their treatment. As summed up in the *Nuffield Council on Bioethics* report, “[c]onsent provides a mechanism to make controlled exceptions to an existing privacy norm for specific purposes. However, consent does not itself ensure that all of the interests of the person giving consent are protected nor does it set aside the moral duty of care owed to that person by others who are given access to the information. On its own, consent is not always necessary, nor always sufficient for ethical extensions of data access.”<sup>37</sup>

Questions concerning ethically justified uses of data cannot simply be reduced to questions of consent. This is also reflected in the fact that EU’s GDPR-regulation under certain circumstances allows for access and processing of patient data without patient consent, for example, if a patient is in a state of coma or suffers from severe dementia and is therefore unable to give consent. In such cases, use of data without consent seems not only legally permissible but also morally justified. Another circumstance in which the GDPR-regulation allows for access and processing of patient data without consent is when this

is “necessary for reasons of substantial public interest.”<sup>38</sup> In such cases, a burden lies on the ones invoking this regulation to put forward strong arguments for the substantial public interest that can justify breaches of privacy of individual patients.

Consequently, the concern for patient privacy must be considered in the construction of clinical algorithms, just as it must be a part of the ongoing ethical considerations of healthcare practitioners. First, in the development of clinical algorithms, one should try to address the question of how to define “morally reasonable expectations about how data will be used in a data initiative, giving proper attention to the morally relevant interests at stake.”<sup>39</sup> The answer to this question will depend on context. That is, what is morally reasonable will depend on the data processes conducted by the clinical algorithms and the form of intervention they offer as well as the healthcare practices in which they are used. As a consequence, it will vary what norms of privacy are applicable to specific clinical algorithms.<sup>40</sup>

Second, healthcare practitioners using clinical algorithms will sometimes have to face complicated choices about when to access patient data. In some situations, the best choice may be to refrain from accessing available data even if it has been the object of informed consent, either because the practice of informed consent is not suitably adapted to new uses of data or because the patient has been unable to foresee this specific use of data. In other situations, healthcare practitioners may need to override the limitations laid out by the informed consent of the patient if the potential benefits to the patient outweigh the concern for privacy in a way that justify such a transgression. The notion of privacy as the explicit control of access to information about oneself<sup>41</sup> is not well suited to guide considerations in these situations. Instead, healthcare practitioners must keep in mind that different norms of privacy and of uses of information will apply to different types of relationships and contexts. Consideration of context is therefore crucial and may sometimes mean that healthcare practitioners should refrain from using or responding to recommendations from specific clinical algorithms when the recommendation is of little benefit to the patient or may lead to serious breaches of privacy. Below, we further examine the ethical significance of clinical algorithms on the relationship between patients and healthcare practitioners.

### Professional Autonomy

From an Aristotelian perspective, an important aspect of being a professional is exercising and cultivating *professional practical wisdom*, which is “...a matter of complex reflection and deliberation concerning professional ends to be realized in action, adequate scrutiny of the contextual complexities, and imaginative exploration of the most effective means to these ends.”<sup>42</sup> Professional practical wisdom is therefore an ability that makes good, professional discretion possible. This requires that professional practitioners consider and explore a wide range of possible issues, decisions, and actions when they deliberate about the course of action to take. The exercise and cultivation of professional practical wisdom are thus a central part of the professional autonomy of the healthcare practitioner, and this autonomy is necessary in order to make a considered decision about what would be the right course of action in relation to an individual patient. It is therefore relevant to examine how clinical algorithms influence healthcare professionals’ opportunities to exercise and cultivate professional practical wisdom.

The distinction between CDS algorithms and CDM algorithms is highly relevant to discussions of professional autonomy. Because the function and purpose of CDS algorithms and CDM algorithms are very different, they have very different effects on practitioner autonomy. The purpose of CDS algorithms is to perform decision *support*, which is a beneficial tool for the cultivation and exercise of professional practical wisdom, because it can bring to attention a consideration or possible course of action that the practitioner otherwise would have not considered. In this way, CDS algorithms support professional practical wisdom and can compensate for constitutional constraints such as time pressure by kickstarting a decision process and ensuring that all relevant information available about a patient is being taken into consideration. CDM algorithms, on the other hand, do not offer decision *support* but make decisions about the best course of treatment. CDM algorithms may pose a problem for professional autonomy

because they become an authority in the decision-making process and thus restrict the practitioners' exercise of discretion in the assessment of the right course of action. In such cases, CDM algorithms may restrict professional autonomy.

When it comes to CDS algorithms, an important aspect as to whether they pose a threat to practitioner discretion is how the algorithms are *perceived* by the practitioner.<sup>43</sup> There is a fine line between when a CDS algorithm is perceived as supportive and when it is perceived as authoritative. To elucidate this delicate balance, it is important to examine a CDS algorithm's *design* and *framework* and how it presents itself. For example, a warning or reminder to the practitioner that says, "the patient is at risk of having high alcohol consumption" presents itself very different from a warning or reminder that says, "you should speak with the patient about his/her alcohol consumption." The first is informative and leaves more room for practitioner discretion than the latter, which can be perceived as more authoritative and therefore not as supportive of practitioner discretion. The responsibility for designing CDS algorithms effectively lies primarily with the algorithm developers and often demands a thorough implementation process in cooperation with relevant health professionals.

The institutional framework of CDS algorithms also influences whether CDS algorithms are perceived as authoritative or not. For example, time pressure can result in CDS algorithms being followed blindly and without discretion because these algorithms provide health professionals with a quick decision. In such cases, CDS algorithms are no longer exclusively supportive but rather authoritative; they are *de facto* used as CDM algorithms, and they thereby limit professional autonomy. In these cases, the responsibility lies with institutions to allow time for professional judgement and with professional to retain autonomy and authority by exercising professional practical wisdom regarding the suggestions of the CDS algorithms.

Professional autonomy can also be challenged in cases in which clinical algorithms involve elements that serve as a "black box," that is, if the practitioner is not able to access or oversee what serves as the justification of the decision support or the decision-making of the clinical algorithm. Maddox et al. note that black boxing is a major challenge for implementing AI-based algorithms into clinical practice, because "use of deep learning and other analytic approaches in AI [...] generate insights via unobservable methods, clinicians cannot apply the face validity available in more traditional clinical decision tools. [...] This 'black box' nature of AI may thus impede the uptake of these tools into practice."<sup>44</sup> International guidelines recommend that developers of clinical algorithms avoid black box software,<sup>45</sup> but this is not always possible when developing clinical algorithms for two reasons. Issues with black boxing can arise if health practitioners do not have the right to access all the available data concerning their patients and thus cannot review the data utilized in the algorithm. In such cases, even if the result of the algorithm is, in principle, understandable by the practitioner, it will *de facto* be "black boxed." Issues with black boxing can also arise because of the complexity of the clinical algorithm's code, or the amount of data processed by the algorithm is impossible for healthcare practitioners to survey.<sup>46</sup>

The black box elements in clinical algorithms can be restrictive for professional autonomy if relevant and important information about a suggestion or decision is black boxed, making it difficult for the practitioner to exercise professional discretion to evaluate the suggested or chosen treatment. In this way, black boxing is thus another way that clinical algorithms can challenge or restrict professional autonomy.

Sometimes these black box problems can be mitigated if the clinical algorithms can be given transparency using models or visualization. The designers of clinical algorithms can therefore, to a great extent, counteract potential black boxing elements of clinical algorithms by including explanatory models to the design. In that way, "input > processing > output"-models (which are at high risk of resulting in black boxing) can be replaced by "input > processing > output + explanation"-models,<sup>47</sup> which are more likely to counter the threats to professional autonomy that black boxing can produce.

However, black boxing cannot always be avoided in practice, for example when the structure of the algorithm is too complex to allow for modeling or visualization. This is often the case with clinical algorithms based on machine learning algorithms. Machine learning algorithms develop a mathematical model based on interactions with a set of sample or training data in order to "automatically detect patterns in data without being explicitly programmed."<sup>48</sup> As a result, even developers are often unable to

describe how a particular algorithm works, or why it results in the most optimal set of predictions or decisions. Machine learning algorithms thus raise problems concerning both understandability (of how the algorithm works) and explainability (of how the algorithm arrives at a result), and many such algorithms are in this way essentially black boxes.<sup>49</sup> Furthermore, the black box problems connected to machine learning algorithms are becoming more pressing as the number of clinical algorithms based on this technology is rising.<sup>50</sup>

According to Watson et al., the main ethical challenge raised by machine learning algorithms is to develop model-centric explanations allowing for an overall understanding of the working of an algorithm, and subject-centric explanations providing a local understanding of the individual application of an algorithm that can be utilized by the healthcare professional and the patient.<sup>51</sup> However, before these two modes of explanation are available, it seems doubtful whether black box machine learning algorithms can be implemented in healthcare practices in an ethically acceptable manner.

### The Effect of Algorithms on the Patient–Practitioner Relationship

The patient–practitioner relationship is a fundamental part of any healthcare practice. To determine the effect or influence algorithms can have on this relationship, it is necessary to understand what type of relationship the patient–practitioner relationship is. Currently, the patient-centered patient–practitioner relationship is the most widely accepted ideal for the patient–practitioner relationship in the Western world.<sup>52</sup> By reviewing what characterizes the patient-centered relationship, we will be able to investigate the potential effects of clinical algorithms on the patient–practitioner relationship.<sup>53</sup>

According to Nicola Mead and Peter Bower (2000), the patient-centered relationship is characterized by the following five dimensions: (1) the biopsychosocial perspective, (2) the “patient-as-person,” (3) shared power and responsibility, (4) the therapeutic alliance, and (5) the “practitioner-as-person.”<sup>54</sup> The biopsychosocial perspective means that the practitioner ought to consider the situation of patients from a broader perspective than those of biology and pathology; the practitioner must also take into account the psychological and sociological perspectives, because patients and their health-related issues are affected in a psychosocial manner.

The second dimension of the patient-centered relationship is that the patient is regarded as a person. In practice, this means taking into account the *individual patients’ experiences* of their disease and the effects these experiences have on the patients’ *unique context* when deciding on the right course of treatment. Mead and Bower give the example of a broken leg.<sup>55</sup> A broken leg does not affect all patients in the same way. If you work in an office, a broken leg might not affect your everyday life as much as if you work as a carpenter. Therefore, the psychological, sociological, and biological perspectives of the illness are not sufficient, because *how* the disease is experienced (according to all perspectives) by the individual patient can vary across these perspectives.

The dimension of shared power and responsibility is best described in opposition to paternalism.<sup>56</sup> The paternalistic relationship between healthcare professional and patient is centered around an asymmetrical relationship in which the practitioner takes on the role of expert and is considered the authority on what is the best decision on behalf of the patient. In contrast, the practitioner and patients are equal in the patient-centered relationship in the sense that patients are acknowledged as experts of their own life and therefore also have autonomy and authority in the relationship.<sup>57</sup> This means that both patient and practitioner possess expert knowledge—the doctor about the medical aspects and the patient about their own needs and preferences—that must be included in a shared decision process in which the patient, to a large extent, is the one who must decide on what (if any) treatment to pursue.<sup>58</sup>

Developing a therapeutic alliance is a fundamental requirement for the patient-centered relationship. The therapeutic alliance focuses on the importance of the personal relationship between the practitioner and patient and the need to establish a common understanding and agreement about the goals and requirements of the specific treatment.<sup>59</sup> The central elements necessary to achieve such an alliance are

both cognitive and affective, ensuring that the patient perceives the doctor as caring, empathic, and sensitive and enabling a personal bond between doctor and patient. The therapeutic alliance is important for the patient-centered relationship exactly because the center of the relationship is collaboration (the shared power and responsibility) between patient and practitioner, which allows them to work together to determine the best course of action.

The practitioner-as-person dimension concerns the importance of the doctor being a subject who has some personal qualities that can influence the way the patient–practitioner relationship develops.<sup>60</sup> In other words, it highlights how it is impossible to separate the practitioner’s professional and personal qualities and to replace a practitioner (even for another practitioner with the same professional qualities) and achieve the same patient–practitioner relationship.

To investigate how algorithms may or may not affect the doctor–patient relationship, we draw on the distinction between CDS and CDM algorithms. The CDM algorithm poses some problems for the biopsychosocial perspective and the patient-as-person perspective, which are based on a holistic perspective of patients and consider the patients’ own experiences of their disease. This is because the data that forms the basis of the decisions of CDM algorithms does not include the social and psychological perspectives of the patients and the patients’ own experience of their disease, which are all central to the patient-centered relationship. As the decisions made by CDM algorithms are based solely on quantified (biological/pathological) health data, these algorithms cannot consider what social and/or psychological effects a specific treatment will have on an individual patient. This type of decision-making, which is entirely based on the biological and pathological perspectives, has more in common with a paternalistic practitioner–patient relationship than with the patient-centered relationship. This means that the CDM algorithm has the potential to undermine both the biopsychosocial perspective and the patient-as-person perspective which are central dimensions of the patient-centered relationship.

On the other hand, CDS algorithm can, to a high extent, support and promote the biopsychosocial and the patient-as-person perspectives included in the patient-centered relationship because CDS algorithms do not make clinical decisions but provide suggestions that the practitioner can consider together with social and psychological perspectives and patients’ experiences. Furthermore, CDS algorithms might save the practitioner some time-consuming work and, in this way, ensure that the practitioner has extra time to better understand the biopsychosocial perspective of the patient and how the disease and different treatments might affect their lives. This possibility of CDS algorithms being promotive to the biopsychosocial and the patient-as-person perspectives is, of course, conditioned by the lack of black boxing elements in CDS algorithms.

Both CDS and CDM algorithms have the potential to undermine some aspects of the shared power and responsibility dimension of the patient-centered relationship. The autonomy of the patient is, to a high extent, built upon the idea that the patient is the expert of his/her own life and therefore is an important and active part of the decision-making process, which is an exchange of the practitioners’ and patients’ expert knowledge; in this exchange, the practitioner and patient, in consultation with each other, agree on the best course of action. In other words, the patient’s autonomy is greatly based not only on being *heard* and but also on *receiving* expert knowledge from the practitioner, which enables the patient to make informed decisions. The practitioner’s autonomy, on the other hand, is greatly based on *having* this medical expert knowledge that the patient is lacking. The problem with both CDS and CDM algorithms is that these algorithms have the potential to damage this exchange of knowledge if the practitioner does not understand the decisions or suggestions made by the algorithm because of issues with black boxing.

In a situation where a CDS algorithm suggests a specific treatment for the patient and the practitioner does not understand why the treatment is suggested and how the algorithm came to that result, the algorithm has the potential to restrict the practitioner’s autonomy, which is based on having and sharing expert knowledge with the patient. In the same scenario, the patient’s autonomy is restricted because the practitioner is not able to inform the patient sufficiently to make an informed decision.

In the same scenario but with a CDM algorithm, the algorithm has the potential to restrict practitioner and patient autonomy even further and possibly deprive them of autonomy altogether. If



an algorithm has the power to *decide* what the best treatment for a patient is without the practitioner being able to understand this decision, then the decision-making process that characterizes the patient-centered relationship becomes redundant. This means that both practitioner and patient autonomy de facto would be transferred to the algorithm, changing the patient-centered relationship to a more paternalistic relationship in which—instead of the practitioner—the algorithm becomes the sole authoritative element in the relationship. This would make both practitioner and patient passive partners in the relationship. Black boxing—caused by both CDS and CDM algorithms—can thus have a negative effect on the shared power and responsibility dimension of the patient-centered relationship and potentially change the nature of the relationship altogether.

The therapeutic alliance can—like the shared power and responsibility dimension—be challenged by the potential of black boxing that both CDS and CDM algorithms have. While the power and responsibility dimension highlight the autonomy of both practitioner and patient, the focus of the therapeutic alliance is the personal relationship between the practitioner and patient. The algorithm can interfere as a third agent in the alliance and potentially be the cause of conflict if the practitioner is not able to explain the relevance or importance of what the algorithm suggests or decides (e.g., because of black boxing). This can create conflict in the therapeutic alliance and obstruct the possibility of a common ground that enables a shared understanding between the practitioner and the patient.

The practitioner-as-person dimension of the patient-centered relationship is not fundamentally changed or challenged by either the CDS or CDM algorithm. This dimension is not affected because practitioners' personal qualities still play a vital role even though the algorithm makes suggestions or decisions. The individual practitioner's personal quality will still make a difference if one practitioner is replaced by another, because their personal qualities influence the way they handle and explain the results of the algorithms to their patient. In this way, the introduction of CDS or CDM algorithms does not take away the importance of practitioners' personal qualities in the patient–practitioner relationship, also highlighted as a central dimension in the patient-centered relationship.

## Conclusion

As we have shown, there are ethical challenges to implementing and using clinical algorithms in healthcare that touch upon some of the most fundamental aspects of how medicine is practiced today. These challenges involve some of the core elements of medicine: beneficence, privacy, autonomy, and the nature of the patient–practitioner relationship. We have shown that there is tension between the possible patient beneficence arising from using clinical algorithms and the possible breaches of privacy these algorithms give rise to; that the use of CDM algorithms and clinical decision-supportive algorithms may challenge professional autonomy in situations characterized by pressure or lack of time; and that clinical algorithms, especially CDM algorithms, may in different ways challenge the ideal of a patient-centered patient–practitioner relationship. It is important to take these ethical challenges seriously in the development and implementation of clinical algorithms in healthcare to ensure that quality healthcare is still provided, and that fundamental rights in healthcare are preserved. This responsibility lies, to a large extent, on the stakeholders and the developers of systems that are based on clinical algorithms. Clinical algorithms have great potential to improve the healthcare system, but it is important not to be blinded by their potential and address and regard the challenges and consequences they might have on some of the fundamental elements of the practice of healthcare professionals.

## Notes

1. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. *Big Data & Society* 2016;3:1–21.
2. Green G, Defoe EC. What is a clinical algorithm. *Clinical Pediatrics* 1978;17(5):457–63.

3. Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, *et al.* A pharmacogenetic versus a clinical algorithm for warfarin dosing. *The New England Journal of Medicine* 2013;**369**:2283–93.
4. Jacobson TA. Toward ‘pain-free’ statin prescribing: Clinical algorithm for diagnosis and management of Myalgia. *Mayo Clinic Proceedings* 2008;**83**(6):687–700.
5. De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: A weighted genetic risk score. *The Lancet. Neurology* 2009;**8**(12):1111–19.
6. Berner ES. Clinical decision support systems: State of the art. *Agency for Healthcare Research and Quality*. AHRQ Publication No. 09–0069; 2009 June:4.
7. See [note 6](#), Berner 2009, at 4.
8. Silber MH, Ehrenberg BL, Allen RP, Buchfuhrer MJ, Earley CJ, Hening WA, *et al.* An algorithm for the management of restless leg syndrome. *Mayo Clinic Proceedings* 2004;**79**(7):916–22.
9. Bousquet J, Schünemann HJ, Hellings PW, Arnavielhe S, Bachert C, Bedbrook A, *et al.* MACVIA clinical decision algorithm in adolescents and adults with allergic rhinitis. *Journal of Allergy and Clinical Immunology* 2016;**138**(2):367–74.
10. Hughes J. An algorithm for choosing among smoking cessation treatments. *Journal of Substance Abuse Treatment* 2008;**34**(4):426–32.
11. Schurink C, Lucas PJF, Hoepelman IM, Bonten MJM. Computer-assisted decision support for diagnosis and treatment of infectious diseases in intensive care units. *The Lancet Infectious Diseases* 2005;**5**(5):305–12.
12. Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children with Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics* 2006;**118**:405–20.
13. Berger JS, Jordan CO, Lloyd D, Blumenthal RS. Screening for cardiovascular risk in asymptomatic patients. *Journal of the American College of Cardiology* 2010;**55**(12):1169–77.
14. Nelson SJ, Blois MS, Tuttle MS, Erlbaum M, Harrison P, Kim H, *et al.* Evaluating RECONSIDER – A computer program for diagnostic prompting. *Journal of Medical Systems* 1985;**9**(5–6):379–88.
15. Uzoka FME, Osuji J, Obot O. Clinical decision support systems (DSS) in the diagnosis of malaria: A case comparison of two soft computing methodologies. *Expert Systems with Applications* 2011;**38**(3):1537–53.
16. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. *Journal of General Internal Medicine* 2008;**23**(1):37–40.
17. Goldhahn J, Rampton V, Spinaz GA. Could artificial intelligence make doctors obsolete? *BMJ* 2018;**363**:k4563.
18. From here on and for the sake of brevity, we use the term “clinical algorithm” to refer to both CDS and CDM algorithms. We will address the two types of algorithms separately when their different implications are relevant to the discussion.
19. Nuffield Council on Bioethics. *The Collection, Linking and Use of Data in Biomedical Research and Healthcare: Ethical Issues*. London: Nuffield Council on Bioethics; 2015, at Chap. 1.
20. See [note 19](#), Nuffield Council on Bioethics 2015, at Chap. 1.
21. Regulation (EU) 2016/679 of the European Parliament, General Data Protection Regulation (GDPR); available at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e1797-1-1> (last accessed 15 Nov 2019).
22. Beauchamp T, Childress J. *Principles of Biomedical Ethics*. New York: Oxford University Press; 2013.
23. The distinction between CDS and CDM algorithms is not relevant in the context of the challenge of patient privacy because the implications of CDS and CDM algorithms both concern data flow.
24. Fairweather NB, Rogerson S. A moral approach to electronic patient records. *Medical Informatics and the Internet in Medicine* 2001;**26**(3):219–34.

25. Llandres N. Ethical problems caused by the use of informatics in medicine. In: Collste G, ed. *Ethics and Information Technology*. New Delhi, India: New Academic Publishers; 1998:76.
26. Cato KD, Bockting W, Larson E. Did I tell you that? Ethical issues related to using computational methods to discover non-disclosed patient characteristics. *Journal of Empirical Research in Human Research Ethics* 2016;**11**(3):214–19.
27. DeCew J. Privacy. *The Stanford Encyclopedia of Philosophy* (spring 2018 edition). Zalta EN, ed.; available at <https://plato.stanford.edu/entries/privacy/> (last accessed 4 Nov 2019).
28. See also Westin A. *Privacy and Freedom*. New York: Atheneum; 1967.
29. This is also the fundamental understanding of privacy motivating the discussion of Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. *Big Data & Society* 2016:1–21.
30. See also Winkelstein PS. Ethical and social challenges of electronic health information. In: Hsinchun C, Fuller SS, Friedman C, Hersh W, eds. *Medical Informatics. Integrated Series in Information Systems*. Boston, MA: Springer Science; 2005:144–5.
31. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*; available at <http://www.ncbi.nlm.nih.gov/books/NBK9579/>.
32. See note 26, Cato et al. 2016, at 215.
33. See note 24, Fairweather, Rogerson 2001. See also Kluge EHW. Health information, the fair information principles and ethics. *Methods of Information in Medicine* 1994;**33**:336–45.
34. See note 24, Fairweather, Rogerson 2001, at 224.
35. See note 27, DeCew 2018
36. See note 1, Mittelstadt 2006, at 10.
37. See note 19, Nuffield Council on Bioethics 2015, at 1–198.
38. See note 21, GDPR, chap. II, article 9(2) g.
39. See note 19, Nuffield Council on Bioethics 2015, at 46–56.
40. For a contextually sensitive understanding of norms of privacy, see Nissenbaum H. Privacy as contextual integrity. *Washington Law Review* 2004;**79**(1):119–58. Nissenbaum suggests two types of informational norms: norms of appropriateness and norms of distribution.
41. See note 26, Cato et al. 2016.
42. Christensen AMS. The institutional framework of professional virtue. In: Carr D, ed. *Cultivating Moral Character and Virtue in Professional Practice*. London: Routledge; 2018:124–34.
43. Zhiping W, Lopez MC. Physician acceptance of information technologies: Role of perceived threat to professional autonomy. *Decision Support Systems* 2008;**46**:206–15.
44. Maddox TM, Rumsfeld JS, Payne PRO. Questions for artificial intelligence in health care. *JAMA* 2019;**321**(1):31–2.
45. High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics guidelines for trustworthy AI. Brussels: European Commission 2019. Compare Shahriari K, Shahriari M. *IEEE Standard Review — Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Artificial Intelligence and Autonomous Systems*. 2017 IEEE Canada International Humanitarian Technology Conference (IHTC); 2017: 197–201.
46. See note 1, Mittelstadt 2006, at 6.
47. Turek M. *Explainable Artificial Intelligence (XAI)*. US Department of Defense Advanced Research Projects Agency; available at <http://www.darpa.mil/program/explainable-artificial-intelligence> (last accessed 6 Aug 2019).
48. Watson D, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: Beyond the black box. *BMJ* 2019;**364**:1886–90, at 1886.
49. We want to thank an anonymous reviewer for this important point.
50. Kim JT. Application of machine and deep learning algorithms in intelligent clinical decision support systems in healthcare. *Health & Medical Informatics* 2018;**9**(5):321–6.

51. See [note 48](#), Watson et al. 2019. For a discussion of validation and regulation of black box algorithms in healthcare, see Nicholson PW. Big data and black-box medical algorithms. *Science Translational Medicine* 2018;10(471):eaao5333.
52. Kaba R, Sooriakumaran P. The evolution of the doctor–patient relationship. *International Journal of Surgery* 2007;5(1):57–65.
53. For another way of investigating the influence of algorithms and AI on the patient–practitioner relationship, see LaRosa E, Danks D. *Impact on trust of healthcare AI*. AIES’18, Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society; 2018:210–15.
54. Mead N, Bower P. Patient-centeredness: A conceptual framework and review of the empirical literature. *Social Science & Medicine* 2000;51:1087–110.
55. See [note 54](#), Mead, Bower 2000.
56. Hershey PT. A definition for paternalism. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 1985;10(2):171–82.
57. See [note 54](#), Mead, Bower 2000.
58. See [note 52](#), Kaba, Sooriakumaran 2007, at 61.
59. See [note 54](#), Mead, Bower 2000, at 1090.
60. See [note 54](#), Mead, Bower 2000, at 1091.