Health Equity in Clinical Research Informatics

Sigurd Maurud, Silje H. Henni, Anne Moen
University of Oslo, Oslo, Norway

Summary

Objectives: Through a scoping review, we examine in this survey what ways health equity has been promoted in clinical research informatics with patient implications and especially published in the year of 2021 (and some in 2022).

Method: A scoping review was conducted guided by using methods described in the Joanna Briggs Institute Manual. The review process consisted of five stages: 1) development of aim and research question, 2) literature search, 3) literature screening and selection, 4) data extraction, and 5) accumulate and report results.

Results: From the 478 identified papers in 2021 on the topic of clinical research informatics with focus on health equity as a patient implication, 8 papers met our inclusion criteria. All included papers focused on artificial intelligence (AI) technology. The papers addressed health equity in clinical research informatics either through the exposure of inequity in AI-based solutions or using AI as a tool for promoting health equity in the delivery of healthcare services. While algorithmic bias poses a risk to health equity within AI-based solutions, AI has also uncovered inequity in traditional treatment and demonstrated effective complements and alternatives that promotes health equity.

Conclusions: Clinical research informatics with implications for patients still face challenges of ethical nature and clinical value. However, used prudently—for the right purpose in the right context—clinical research informatics could bring powerful tools in advancing health equity in patient care.

Keywords

International Medical Informatics Association Yearbook, clinical research informatics, health equity, structural bias, discrimination, artificial intelligence, prediction models, algorithmic bias

1 Introduction

Clinical research informatics (CRI) is a sub-discipline within biomedical and health informatics that focuses on the analysis, interpretation, and presentation of clinical knowledge generated through informatics [1]. This definition by Embi and Payne [1], dates back over a decade ago and has previously been mentioned in this journal by Solomonides [2]. Acknowledging and including topics that have flourished since Embi and Payne's definition [2], the most notable addition is Artificial Intelligence (AI), referring in a broad sense to the ability of technology to resemble functions and processes of human beings. Machine Learning (ML) [3], and Natural Language Processing (NLP) [4, 5], are notable subgroups of AI technologies using relevant clinical data sets to advance the representation and understanding of a problem. Increased availability of clinical data in digital form and expanding computational capacity enable more complex and sophisticated processing of clinical data in CRI [6]. This appears as prevalent use of AI techniques in CRI, seeking to advance clinical practice through decision-support and prediction capabilities to support health practitioners in different specialties [7-14]. However, extensive use of AI algorithms has also revealed potential risks with implications for patients and their prospects of best possible treatment [15]. An example is the "black box" challenge, which delineates the necessity of making complex AI operations transparent and comprehensible to end-users [16]. Vast databases are being queried by algorithms, as researchers and clinicians are seeking patterns that can guide decisions in clinical care and result in opaque explanations for how the algorithms reach their guidance [17]. The challenge here may not be a result of secrecy or inadequate knowledge, but rather ML outputs

without regard for human comprehension and careful consideration of clinical relevance [18]. In other words, "black box" approaches to decision-making in patient care may incur significant risk, as neither practitioner nor patient can fully comprehend the steps leading to recommendations [3, 19]. In addition to well-known problems being addressed in the development of AI services in health [20], issues like access to and ownership of clinical data, and possible exacerbation of health inequity [21, 22] are important ethical issues of concern. The focus on these topics seems further accelerated by the unprecedented deployment of digital solutions within healthcare during the COVID-19 pandemic. Consequently, several issues were illuminated, including dependence on digital health, how to enable digital solutions to provide better healthcare, as well as existing inequalities and structural discrimination [23].

Health is associated with a non-medical social gradient, where those on the lower socioeconomic end often have the least chance for good health. The circumstances in which an individual is born, ages, lives, and works, usually referred to as "social determinants of health" (SDoH), can be exacerbated by discrimination, prejudice, and stereotyping [24]. Health equity, as defined by the World Health Organization, is 'the absence of unfair, avoidable and remediable differences in health status among groups of people', and requires actions that even out differences in health outcomes between populations with different socioeconomic foundations [24]. While holding great promise for the use of AI in health care, CRI can pose risk of reflecting and reproducing analytical and algorithmic biases that potentially increase health inequities that come with SDoH [21, 25]. Algorithmic bias that discriminates based on characteristics integral to the

person, such as race and ethnicity, has received particular attention in this context [26-30]. Considering 'race' and 'ethnicity', 'ethnicity' comes with similar interpretation in the literature, but it is necessary to address the conflicting use of the term 'race' in European and American contexts. The US Census Bureau and the Office of Management and Budget (OMB) refer to race as a socially constructed way of separating humans into different sociocultural and ancestral groups [31], while in Europe, the term 'race' is avoided due to its association with the wrong notion of biological different races among human beings, preceding the linked historical and ideological associations. Racism is, on the other hand, an acknowledged term in Europe, referring to discrimination based on the notion of biological differences [32]. Recognizing this distinction, this article uses the term 'race' when referring to the cited sources that apply this in line with the definition of the US Census Bureau and OMB.

A key concern in Real World Data (RWD) based studies is representativeness. Using such data sets for training algorithms poses a risk for algorithmic biases in AI [33], originating from, e.g., lack of inclusion of underrepresented population groups in samples [34, 35], and subjective assessments [36, 37] within the data material. An example of this problem is the socially inconsistent and intermixing use of the terms 'ethnicity' and 'race', possibly affecting the creation and collection of data [30]. Another key concern is to delineate which circumstances one is to discern between different populations as it may be of relevance to some conditions and completely irrelevant to others [38].

CRI hold much promise for improving clinical practice [1], but needs to incorporate assessment of its impact on health equity to provide healthcare to all patients, regardless of SDoH [16, 21, 39]. Illustrating this concern is the establishment of the High-Level Expert Group on Artificial Intelligence (AI HLEG) by the European Commission [15], to accommodate for the implementation of the Commission's vision for ethical AI [40]. As an output, seven requirements for Trustworthy AI have been published from this group: 1)

Human agency and oversight, 2) Technical robustness and safety, 3) Privacy and data governance, 4) Transparency, 5) Diversity, non-discrimination and fairness, 6) Societal and environmental wellbeing, and 7) Accountability [15]. Taking this into account, CRI should go beyond monitoring, controlling, and guarding against unintentional outcomes that may exacerbate structural health inequality, to actively address and hence, improve health equity [21, 35, 41]. Experience gained during the COVID-19 pandemic has highlighted the need for a more systematic approach to ensure that digital health and CRI promotes health equity and the goal towards universal health coverage [23]. With this aspiration and inspired by the topic of the 2022 IMIA Yearbook: "Inclusive Digital Health: Addressing Equity, Literacy, and Bias for Resilient Health Systems" [42], the aim of this scoping review was to examine in what ways research in CRI, published in the year of 2021, has included health equity to promote patient health and care.

2   Method

This scoping review applied methods as outlined in the Joanna Briggs Institute (JBI) Manual [43]. The process of the review proceeded as follows: 1) identify research question, search terms and keywords, 2) search for literature, 3) screening and selecting relevant literature, 4) extract data from selected literature and, 5) summarize and present the results. A protocol defining the research question, aim, screening process, search terms and criteria for inclusion and exclusion was developed in advance of the literature search. The approach is illustrated in a PRISMA flow diagram (see Figure 1) [44].

2.1   Search Strategy

A medical librarian guided our search in September 2022, using the following databases: Medline, Embase, ACM library and Epistemonikos. In line with the JBI Manual [43], a PCC framework was applied for the literature search:

Population – Clinical Research Informatics
Concept – Health Equity
Context – Patient Implications

The documentation of the search and the overview of identified literature in the databases are available upon request.

2.2   Screening and Selection of Literature

Ahead of the screening process, we screened the titles and abstracts of 25 randomly selected sources from the search results to reach a general agreement on Inclusion and Exclusion criteria before the selection of sources (see Table 1). All sources were then screened by the first and second author using the predetermined criteria for inclusion and exclusion. The sources were screened in two subsequent rounds supported by Covidence, a web-based collaboration software platform used for screening and data extraction in literature reviews [45]. The first round of screening extracted literature based on titles and abstract, while the second round of screening extracted literature through full-text reading. The first round of screening resulted in conflicts on 86 sources (18 % of the total sources screened), all of which were resolved through a plenary review. The second screening round resulted in one conflict among the 58 sources that underwent full-text review. The conflict was resolved through a plenary review of the source. Specific quality assessment of the literature was not carried out as this is generally not a priority in scoping reviews [43].

2.3   Data Extraction

A spreadsheet with the data material was created to extract information on the study reference, population characteristics and key findings that relate to the aim of the scoping review. The first and second authors read the full text sources with the purpose to identify and extract aspects of clinical research informatics, aspects of health equity, and aspects of patient implications.

Table 1 Inclusion and exclusion criteria for the screening of literature.

Inclusion Criteria	Exclusion criteria
Research literature published in 2021: ▪ Articles Full paper conference contributions	Research literature published in 2021: ▪ Editorials ▪ Letters ▪ Protocols ▪ Conference abstracts
Literature published in Danish, English, Norwegian and Swedish	Literature that is not directly focused on how clinical research informatics can drive health equity
Literature that focuses on clinical research informatics and health equity	
Literature that provides research with patient implications	

Reasons for exclusion of full-text sources
No focus on clinical research informatics .... n = 28
No focus on health equity .... n = 7
No focus on implications for patients .... n = 1
General data suitability and ethical considerations for AI research .... n = 5
Letter .... n = 8
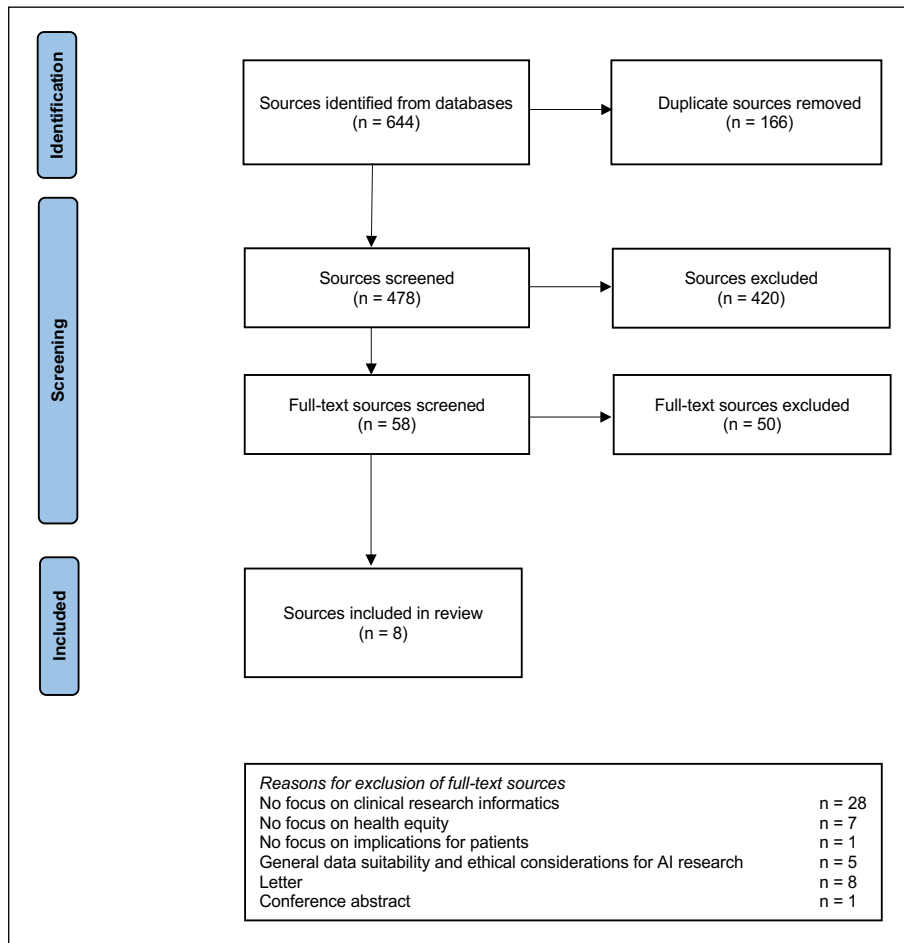Conference abstract .... n = 1

Fig. 1 PRISMA Flow Diagram

3 Results

Of the 58 sources that underwent full-text review, eight studies were included in this study. The reasons for exclusion are listed in Figure 1. Although five sources focused on health equity in CRI, they did not focus on what ways CRI can drive health equity and was therefore excluded under the reason "General data suitability and ethical considerations for AI research". Among the eight included studies, three were reviews and therefore controlled for duplicates. Patra *et al.* [46] and Pham *et al.* [47] are both citing Hazlehurst *et al.* [48]. Patra *et al.* [46] also share an article with Craig *et al.* [49], both citing Navathe *et al.* [50]. However, as the included sources have different focus and have contributed different findings, we did not consider the reviews as overlapping. No overlapping articles were identified between Pham *et al.* [47] and Craig *et al.* [49]. With the exception of one Canadian study [47], all studies were conducted in the United States. The focus on CRI in these papers was consistently on AI technology with two themes identified on how CRI with patient implications can drive health equity:
1) Exposing health inequity in CRI and addressing the need for adequate measures;
2) Promoting health equity through CRI.

The included publications are presented in Table 2 with descriptions of how health equity in CRI with patient implications is addressed in each paper.

4 Discussion

Health equity, CRI and AI are topics in a global setting. It is therefore interesting that all the included papers in this review are of North American origin. Except for one Canadian study, all included studies are conducted in the US It may be that this simply reflects a greater focus on health equity in CRI and AI in the US compared to other countries, responding to a policy that focuses on promoting equity and justice for all [57, 58]. Furthermore, low- and middle income countries still appear to face challenges considering the implementation

Table 2 Included sources for study. The articles are listed in alphabetical order of the first author's surname.

Reference	Health Equity	Clinical Research Informatics	Patient implications
Coley et al., [51]	The study evaluated racial/ethnic differences in the performance of two prediction models for death by suicide.	The object under assessment was a risk algorithm, trained on health record data, with the aim of predicting suicide within 90 days after an outpatient visit in mental healthcare.	Health record data may be inadequate for the prediction of suicide in some racial/ethnic groups, due to insufficient data on death by suicide for racial/ethnic groups with low prevalence, but also systemic barriers, practitioner bias, institutionalized discrimination, outcome misclassification, and cultural differences. Currently accessible methods do not address subgroups of the entire sample. Clinical prediction models need to be assessed for performance within the population group of relevance to bridge the gap to underserved populations.
Craig et al., [49]	Through a literature review, this study examined how SDoH data and health technology are utilized to enhance the management of population health.	AI holds the possibility to increase the acquisition and utilization of SDoH data from electronic health records (EHR).	The study concluded that technologies with advanced analytical methods holds promise for considerate, inclusive, and effective personalized care and may provide positive contributions to decision-making processes.
Hammarlund [52]	The study examines whether differential health factors can explain racial treatment disparity in cardiac surgery.	Conventionally, the explicability of health risk disparity in treatment disparity is tested with adjusted observed rates for patient-level health differences to optimize patient comparability. The study suggests an alternative, using a ML-based method that enhances the capture of clinical practice to adjust for the critical predictors in acute myocardial infarction surgery.	The suggested ML-based method decreased the standard adjusted acute myocardial infarct surgery disparity by 45-55%. After adjusting for predictive controls, a smaller, but still present surgery disparity remains, indicating a differential racial treatment explained by other factors than health risk differences. Physician experience with treating black patients appeared to reduce stereotyped statistical discrimination.
Lu et al., [53]	With the aim of improving equity in healthcare access throughout the COVID-19 pandemic, the study developed and tested a ML-based model that matched predicted patient language with contact tracer language in communities highly represented with language barriers.	The ML-based model used random forests, integrating laboratory reports data with census data to predict the language of any incoming patient. To ensure interpretability for the stakeholders, the study based the model on a small feature set, optimizing transparency and comprehension.	The study demonstrates how a small ML-based measure can enhance effectiveness of infectious disease response and at the same time support health equity by matching cases to contact tracers based on language.
Patra et al., [46]	The review examined different approaches for indexing and providing SDoH data through natural language processing (NLP).	Matching of keyword and classification were the most prevalent approaches to identify SDoH, either through supervised or unsupervised algorithms. Due to insufficient data quantity for certain SDoHs, the review reports lower performance of neural networks compared to traditional ML-algorithms.	The study promotes an NLP-based SDoH identification to tackle the increasing acknowledgment of nonclinical factors' influence on health. The study argues that NLP-based methods to index and develop outcome analysis in SDoH may optimize clinical decision-making without greater interference in current workflows.
Pham et al., [47]	The potential for AI technology to be algorithmically biased in disfavour for underserved populations led this case study to review and propose ethnoracial considerations in AI-based diabetes management tools.	The authors express particular concern about distribution shift in AI models; discrepancy between the training population and the end-users leading to inaccurate predictions.	The authors identified a small degree of ethnoracial data in the reviewed articles (10 out of 141 articles on AI-based diabetes tools). The study calls for measures that facilitate inclusive algorithms that provide diversity considerate healthcare innovations.
Pierson et al., [54]	Through a ML-based method, the study demonstrates how standard radiographic measures of osteoarthritis severity ignore undiagnosed features that inordinately affect underserved populations with knee pain.	As an alternative to the standard radiographic measure Kellgren-Lawrence Grade (KLG), the study trained a convolutional neural network on a dataset of radiographs to predict reported pain score for each knee. Predictions were generated in an independent validation set separated from the training set.	The algorithmic pain prediction measure accounts to a larger degree for disparities in osteoarthritis compared to KLG and is proposed by the authors as a tool to enhance access to treatment for underserved patients with osteoarthritis.
Thompson et al., [55]	The study assessed the predictions of an existing NLP opioid misuse classifier, using principles of fairness, accountability, transparency, and ethics to tackle the possible impact of historical and structural bias on disadvantaged subpopulations.	The study adapted a python toolkit for reviewing fairness and bias in ML-models and the Local Interpretable Model-Agnostic Explanations [56] to assess face validity. The post-hoc results demonstrated disparities in patient characteristics across subgroups.	The authors argue that collaboration with data scientists and ethic experts in advance of model training, could have addressed disparities and barriers in the model, although inequities present in the model may only be addressed in post-hoc recalibrations. The study demonstrates the result of disproportionately overlooking potential biases in the validation of convolutional neural networks.

of AI in health [59]. The US, compared to other high-income countries, are ranking well below in health care system performance. Among the domains that pull the performance down are less access and equity, despite the significant amount of the US gross domestic product that are spent on health care [60].

Although it appears the COVID-19 pandemic has highlighted established disparities in health and digital health [23], interestingly only one study identified in the survey as published in 2021 acknowledged the COVID-19 pandemic as a catalyst [53]. However, the recognition of social inequities and its influence on health or digital health in all included publications may rather reflect the COVID-19 pandemic's role as a catalyst in illuminating and driving the focus on health equity [23].

4.1 Exposing Health Inequity in CRI

In line with concerns articulated in previously cited literature [21, 25-30], three of the included sources discussed risk for health inequity in AI-based health technology [47, 51, 55]. Coley *et al.* [51] assessed differences in the performance of two prediction models. They found that these addressed subpopulations of ethnic or racial minorities in an inadequate manner, identifying a smaller proportion of anticipated suicides in patients who report race/ethnicity to be Black and Alaskan Native/American Indian, and in patients who do not report race/ethnicity. The secondary analysis conducted by Pham *et al.* [47] of 141 articles from a literature review, looked into ethnoracial considerations in AI diabetes tools, to propose a strategy for equity for such technology. As the creators of an NLP opioid misuse classifier, Thompson *et al.* [55] evaluated the impact of bias against historically and structurally disadvantaged groups. All three CRI studies acknowledged the challenges for health equity within AI, mainly expressed through algorithmic bias [47, 51, 55]. This is echoing the fact that RWD used to train algorithms risk reproducing bias in technological solutions [33], possibly through lower accuracy for underrepresented samples of underserved groups

[34, 35], or through subjective assessments [36], reinforcing possible judgemental biases from healthcare providers [37]. Thompson *et al.* [55] acknowledge their previous lack of consideration for disadvantaged populations in the creation of their instrument, and thus mirror the concern of Coley *et al.* [51] for insufficient attention to the clinical usefulness or utility of AI technology to disadvantaged subpopulations. Besides, Pham *et al.* [47] identified only 10 out of 141 papers on AI diabetes tools that inconsistently addressed race or ethnicity, or both (race/ethnicity), pointing to a lack of reliable data and a lack of focus for ensuring adequate training algorithms for ethnic or racial minority populations. Even if assessed for algorithmic bias, the "black box" nature of AI will still challenge transparency, potentially including unintended bias, or withhold information underlying the performance of a model [17, 18, 55]. This is crucially important [21, 22], as it emphasizes the responsibility of CRI to acknowledge and act upon this in digital health technology [46, 55], and incorporate principles for ethical AI, as outlined by the AI HLEG [15].

4.2 Promoting Health Equity in CRI

The remaining five papers examined ways in which CRI may enable and promote health equity. Craig *et al.* [49] and Patra *et al.* [46] did so in an indirect manner, through literature reviews examining and promoting the utility of AI to actively include and use SDoH data from electronic records. However, it appears to be beyond the scope of both reviews [46, 49], to discuss value in subjective data, as well as potential bias introduced by the source of data. As clinical text include subjective data in EHR [36], this illustrates the issue of possibly overlooked subjective bias in algorithmic performance [30, 36, 37]. Indeed, algorithmic bias appears in general as a difficult barrier for health equity to overcome [26-30, 33-37, 51, 55]. The current conditions where CRI demonstrates promotion of health equity appears admittedly to be those where CRI is used specifically to address inequity in health; not only in the evaluation

of AI-based healthcare instruments ability to promote health equity [51, 55], but through AI-based methods demonstrating and addressing disparities in the delivery of healthcare services [52-54].

Building from observations that underserved populations experience a greater amount of pain in osteoarthritis, Pierson et al. [54] used a deep learning approach on radiographs to predict the pain level of the individual patient, finding that the approach significantly reduced unexplained racial pain disparities compared to traditional methods. Through a ML-based method, Hammarlund [52] demonstrated disparities between black and white patients in acute myocardial infarction treatment, beyond that explained by health risk differences. As a response to how language discordances limited the contact tracing of a non-English speaking population in California, already disproportionately affected by COVID-19, Lu et al. [53] used an ML-based approach to predict the language of an incoming patient and match this to the language of the contact tracer. In contrast to the other included sources in this scoping review, these sources address health equity by directly responding to existing health inequities. Pierson et al. [54] and Hammarlund [52] both have in common the use of AI in exposing health inequity in clinical practices and providing alternative solutions, while Lu et al. [53] uses AI to promote health equity in a setting known to be characterized by disparities in health and access to health. All the included sources of this scoping review address health equity in CRI with patient implications, either by exposing health inequality in AI-based solutions or by examining possibilities for AI to extract data of importance to address health equity [46, 47, 49, 51, 55]. However, Pierson et al. [53], Lu et al. [53], and Hammarlund [52] all stand out in their application of AI to drive health equity. The accomplishments of these three studies appear to stem from how they approach the issue of health equity. Instead of illuminating health inequity present in CRI-driven solutions, such as algorithmic bias within AI-based prediction models [51, 55], they use AI to promote and improve health equity in the deliverance of existing treatments and health care services [52-54].

4.3 The Way Forward

To further assess the result of our findings, we performed a similar search for the year of 2022 to discern if more recent literature would add to the significance for this study. We identified at least 21 papers [61-81] that met our inclusion criteria, including results from the 2022 IMIA Yearbook [64, 68, 78]. Reading through these articles, we did not identify additional thematic areas than those we have included for the year of 2021. However, it appears to be a change in terms of attention to the topic. The focus on health equity in CRI seems to increase considering the 21 studies we found published in 2022 compared to the 8 from 2021. Following this again, three studies were also identified, just for the first month of 2023 [82-84]. The interest for the topic are expanding from North America based study reports to other parts of the world, including Europe [63, 65, 67, 77] and Asia [84]. The scope of health equity in CRI also appears to have expanded and evolved. Primarily centred on challenges considering race and/or ethnicity in 2021 [47, 51-55], health equity in CRI has extended to diagnosis bias in rural populations [75], age [64], and gender or sex-specific bias [63, 64, 67, 70, 73, 77].

5 Conclusion

Several of the studies on Clinical Research Informatics presented here highlight algorithmic bias as a factor in the promotion of health equity in digital solutions [47, 51, 55]. It appears to be a considerable challenge for CRI to provide AI-based solutions without algorithmic bias that prove counterproductive to the intention and goal of the solutions. Carefully selecting and appropriately balancing different characteristics may reduce algorithmic bias and adjust outcomes in some cases, but bias can also remain hidden which make correction nearly impossible [38]. Based on the findings in this scoping review, our impression is that the field of CRI, here exemplified by AI as the focus of the recent publications found, is more aware of the challenges at hand, which is an important starting point to find solutions that remedy this challenge. This way CRI will increase capability to promote and improve health equity. This review illustrates that when the right form of digital technology is correctly adapted to the population in question at the right time, AI-based CRI-solutions hold a promise to drive equity in health. Recent publications, in 2022 and beyond, illustrate advancements and endeavour to improve AI algorithms that leverage and combine efforts to reduce and eliminate algorithmic bias. Further progress and full incorporation into CRI require thorough assessment and improvement for equitable and ethical distribution of health care services that respect patient autonomy and dignity.

Going forward, CRI holds opportunities for novel patient- focused digital tools that stimulate engagement and promote health equity. This requires tools that do not exacerbate structural inequalities, incorporate ethical consideration to avoid harm, and mitigate risks related to sub-populations already exposed to disparities in society and health.

References

1. Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc 2009;16(3):316-27. doi: 10.1197/jamia.M3005.

2. Solomonides A. Review of Clinical Research Informatics. Yearb Med Inform 2020;29(1):193-202. doi: 10.1055/s-0040-1701988.

3. Begley K, Begley C, Smith V. Shared decision-making and maternity care in the deep learning age: Acknowledging and overcoming inherited defeaters. J Eval Clin Pract 2021 Jun;27(3):497-503. doi: 10.1111/jep.13515.

4. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. BMC Med Res Methodol 2021;21(1):158. doi: 10.1186/s12874-021-01347-1.

5. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: An introduction. J Am Med Inform Assoc 2011;18(5):544-51. doi: 10.1136/amiajnl-2011-000464.

6. Chute CG. From Notations to Data: The Digital Transformation of Clinical Research. In: Richesson RL, Andrews JE, editors. Clinical Research Informatics. Cham: Springer International Publishing; 2019. p. 17-25. doi: 10.1007/978-1-84882-448-5_2

7. Senders JT, Arnaout O, Karhade AV, Dasenbrock HH, Gormley WB, Broekman ML, et al. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. Neurosurgery 2018;83(2):181-92. doi: 10.1093/neuros/nyx384.

8. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neurosci Biobehav Rev 2017;80:538-54. doi: 10.1016/j.neubiorev.2017.07.004.

9. Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy artificial intelligence for the diagnosis of melanoma. Br J Dermatol 2009;161(3):591-604. doi: 10.1111/j.1365-2133.2009.09093.x.

10. van den Heever M, Mittal A Haydock M, Windsor J. The use of intelligent database systems in acute pancreatitis – A systematic review. Pancreatology 2013;14(1):9-16. doi: 10.1016/j.pan.2013.11.010.

11. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017;124(7):962-9. doi: 10.1016/j.ophtha.2017.02.008.

12. Dallora AL, Eivazzadeh S, Mendes E, Berglund J, Anderberg P. Machine learning and microsimulation techniques on the prognosis of dementia: A systematic literature review. PLoS One 2017;12(6):e0179804-e.

13. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. Comput Math Methods Med 2016;2016:1-8. doi: 10.1155/2016/8708434.

14. McCoy TH, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. JAMA Psychiatry 2016;73(10):1064-71. doi: 10.1001/jamapsychiatry.2016.2172.

15. European Commission. Ethics Guidelines for Trustworthy AI; 2019.

16. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Cai JC, MalhotraN, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Medical Ethics 2021;22(1):14. doi: 10.1186/s12910-021-00577-8.

17. Price WN. Medical Malpractice and Black-Box Medicine. Cambridge University Press; 2018. p. 295-306.

18. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big data & society 2016;3(1):2053951715622251. doi: 10.1177/2053951715622512.

19. Bjerring JC, Busch J. Artificial Intelligence and Patient-Centered Decision-Making. Philosophy & technology 2021;34(2):349-71. doi: 10.1007/s13347-019-00391-6.

20. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA); 2018. p. 80-9. doi:10.1109/DSAA.2018.00018.

21. Clark CR, Wilkins CH, Rodriguez JA, Preininger AM, Harris J, DesAutels S, et al. Health Care Equity in the Use of Advanced Analytics and Artificial Intelligence Technologies in Primary Care. J Gen Intern Med 2021;36(10):3188-93. 10.1007/s11606-021-06846-x.

22. Matheny M, Thadaney Israni S, Ahmed M, Whicher D. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington DC: National Academy of Medicine; 2019.

23. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, et al. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. Lancet 2021;398(10312):1727-76. doi: 10.1016/S0140-6736(21)01824-9.

24. World Health Organization. Health equity and its determinants 2021 [updated Apr 6. Available from: https://www.who.int/publications/m/item/health-equity-and-its-determinants].

25. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science 2017;356(6334):183-6. doi: 10.1126/science.aal4230.

26. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med 2018;378(11):981-3. doi: 10.1056/NEJMp1714229.

27. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-53. doi: 10.1126/science.aax2342.

28. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med 2020;383(9):874-82. doi: 10.1056/NEJMms2004740.

29. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. JAMA Dermatol 2018;154(11):1247-8. doi: 10.1001/jamadermatol.2018.2348.

30. Singh S. Racial biases in healthcare: Examining the contributions of Point of Care tools and unintended practitioner bias to patient treatment and diagnosis. Health (London) 2021:13634593211061215. doi: 10.1177/13634593211061215.

31. United States Census Bureau. About the Topic of Race United States Census Bureau; 2022 [updated Mar 1, 2022; cited 2022 Nov 8. Available from: https://www.census.gov/topics/population/race/about.html].

32. Bell M. 'Race', Ethnicity, and Racism in Europe. Oxford: Oxford: Oxford University Press; 2009.

33. Chiang S, Picard RW, Chiong W, Moss R, Worrell GA, Rao VR, et al. Guidelines for Conducting Ethical Artificial Intelligence Research in Neurology: A Systematic Approach for Clinicians and Researchers. Neurology 2021;97(13):632-40. doi: 10.1212/WNL.0000000000012570.

34. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Intern Med 2018;178(11):1544-7. doi: 10.1001/jamainternmed.2018.3763.

35. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. JAMA 2019;322(24):2377-8.

36. Mullainathan S, Obermeyer Z. Does Machine Learning Automate Moral Hazard and Error? Am Econ Rev 2017;107(5):476-80. doi: 10.1001/jama.2019.18058.

37. Blumenthal-Barby JS, Krieger H. Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy. Med Decis Making 2015;35(4):539-57. doi: 10.1177/0272989X14547740.

38. Starke G, De Clercq E, Elger BS. Towards a pragmatist dealing with algorithmic bias in medical machine learning. Med Health Care Philos 2021;24(3):341-9. doi: 10.1007/s11019-021-10008-5.

39. Chauhan C, Gullapalli RR. Ethics of AI in Pathology: Current Paradigms and Emerging Issues. Am J Pathol 2021;191(10):1673-83. doi: 10.1016/j.ajpath.2021.06.011.

40. European Commission. Communication from the Commission to The European Parliament, The European Council, The Council, The European Economic and Social Committee and The Comittee of the Regions: Coordinated Plan on Artificial Intelligence. Brussels: European Commission; 2018. p. 10.

41. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866-72. doi: 10.7326/M18-1990.

42. Mougin F, Fultz Hollis K, Soualmia LF. Inclusive Digital Health. Yearb Med Inform 2022;31(01):2-6. doi: 10.1055/s-0042-1742540.

43. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews. 2020 [cited 16 feb 2022]. In: JBI Manual for Evidence Synthesis. [cited 16 feb 2022]. [Available from: https://synthesismanual.jbi.global].

44. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

45. Covidence systematic review software. Melbourne: Veritas Health Innovation; [Available from: www.covidence.org.

46. Patra BG, Sharma MM, Vekaria V, Adekkanattu P, Patterson OV, Glicksberg B, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. J Am Med Inform Assoc 2021 Nov 25;28(12):2716-27. doi: 10.1093/jamia/ocab170.

47. Pham Q, Gamble A, Hearn J, Cafazzo JA. The Need for Ethnoracial Equity in Artificial Intelligence for Diabetes Management: Review and Recommendations. J Med Internet Res 2021;23(2):e22320. doi: 10.2196/22320.

48. Hazlehurst B, Green CA, Perrin NA, Brandes J, Carrell DS, Baer A, et al. Using natural language processing of clinical text to enhance identification of opioid-related overdoses in electronic health records data. Pharmacoepidemiol Drug Saf 2019;28(8):1143-51. doi: 10.1002/pds.4810.

49. Craig KJT, Fusco N, Gunnarsdottir T, Chamberland L, Snowdon JL, Kassler WJ. Leveraging Data and Digital Health Technologies to Assess and Impact Social Determinants of Health (SDoH): a State-of-the-Art Literature Review. Online J Public Health Inform 2021;13(3):E14. doi: 10.5210/ojphi.v13i3.11081.

50. Navathe AS, Zhong F, Lei VJ, Chang FY, Sordo M, Topaz M, et al. Hospital Readmission and Social Risk Factors Identified from Physician Notes. Health Serv Res 2018;53(2):1110-36. doi: 10.1111/1475-6773.12670.

51. Coley RY, Johnson E, Simon GE, Cruz M, Shortreed SM. Racial/Ethnic Disparities in the Performance of Prediction Models for Death by Suicide after Mental Health Visits. JAMA Psychiatry 2021;78(7):726-34. doi: 10.1001/jamapsychiatry.2021.0493.

52. Hammarlund N. Racial treatment disparities after machine learning surgical risk-adjustment. Health Serv Outcomes Res Methodol 2021;21(2):248-86. doi: 10.1007/s10742-020-00231-7.

53. Lu L, Anderson B, Ha R, D'Agostino A, Rudman SL, Ouyang D, et al. A language-matching model to improve equity and efficiency of COVID-19 contact tracing. Proc Natl Acad Sci U S A 2021;118(43):e2109443118. doi: 10.1073/pnas.2109443118.

54. Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nat Med 2021;27(1):136-40. doi: 10.1038/s41591-020-01192-7.

55. Thompson HM, Sharma B, Bhalla S, Boley R, McCluskey C, Dligach D, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: Detection and mitigation of electronic health record data disadvantages across racial subgroups. J Am Med Inform Assoc 2021;28(11):2393-403. doi: 10.1093/jamia/ocab148.

56. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA; 2016. p. 1135–44. doi: 10.1145/2939672.2939778.

57. Biden J. Presidential Actions [Internet]. Washington D. C. : The White House; 2021. [Available from: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/].

58. The White House. Advancing Equity and Racial Justice Through the Federal Government Washington DC: The White House; 2021 [cited 2023 Jan 20]. [Available from: https://www.whitehouse.gov/equity/].

59. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. NPJ Digit Med 2022;5(1):162. doi: 10.1038/s41746-022-00700-y.

60. Schneider E, Shah A, Doty M, Tikkanen R, Fields K, Williams II R. Mirror, Mirror 2021 – Reflecting Poorly: Healthcare in the U.S. Compared to Other High-Income Countries. The Commonwealth Fund; 2021 August 4. [Available from: https://www.commonwealthfund.org/publications/fund-reports/2021/aug/mirror-mirror-2021-reflecting-poorly].

61. Adam H, Yang MY, Cato K, Baldini I Senteio C, Celi LA, et al. Write It Like You See It: Detectable Differences in Clinical Notes by Race Lead to Differential Model Recommendations. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society; Oxford, United Kingdom: Association for Computing Machinery; 2022. p. 7–21. doi: 10.1145/3514094.3534203.

62. Bear Don't Walk OJ, Reyes Nieva H, Lee SS, Elhadad N. A scoping review of ethics considerations in clinical natural language processing. JAMIA Open 2022;5(2):ooac039. doi: 10.1093/jamiaopen/ooac039.

63. Bermudez-Lopez M, Marti-Antonio M, Castro-Boque E, Bretones MDM, Farras C, Torres G, et al. Development and Validation of a Personalized, Sex-Specific Prediction Algorithm of Severe Atheromatosis in Middle-Aged Asymptomatic Individuals: The ILERVAS Study. Front Cardiovasc Med 2022;9: 895917. doi: 10.3389/fcvm.2022.895917.

64. Chaunzwa TL, Del Rey MQ, Bitterman DS. Clinical Informatics Approaches to Understand and Address Cancer Disparities. Yearb Med Inform 2022;31(1):121-30. doi: 10.1055/s-0042-1742511.

65. D'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: A systematic scoping review and framework. Fam Med Community Health 2022; 10(Suppl 1):e001670. doi: 10.1136/fmch-2022-001670.

66. Das S, Shi X. Offspring GAN augments biased human genomic data. Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; Northbrook, Illinois: Association for Computing Machinery; 2022. p. 1-10. doi: 10.1145/3535508.3545537.

67. Delgado J, de Manuel A, Parra I, Moyano C, Rueda J, Guersenzvaig A, et al. Bias in algorithms of AI systems developed for COVID-19: A scoping review. J Bioeth Inq 2022, 19(3):407-19.

68. Dixon BE, Holmes JH. Special Section on Inclusive Digital Health: Notable Papers on Addressing Bias, Equity, and Literacy to Strengthen Health Systems. Yearb Med Inform 2022;31(1):100-4.

69. Estiri H, Strasser ZH, Rashidian S, Klann JG, Wagholikar KB, McCoy TH, et al. An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes. J Am Med Inform Assoc 2022;29(8):1334-41. doi: 10.1093/jamia/ocac070.

70. Golder S, O'Connor K, Wang Y, Stevens R, Gonzalez-Hernandez G. Best Practices on Big Data Analytics to Address Sex-Specific Biases in Our Understanding of the Etiology, Diagnosis, and Prognosis of Diseases. Annu Rev Biomed Data Sci 2022 Aug 10;5:251-67. doi: 10.1146/annurev-biodatasci-122120-025806.

71. Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and Mitigation of Racial Bias in Clinical Machine Learning Models: Scoping Review. JMIR Med Inform 2022;10(5):e36388. doi: 10.2196/36388.

72. Li Y, Wang H, Luo Y. Improving Fairness in the Prediction of Heart Failure Length of Stay and Mortality by Integrating Social Determinants of Health. Circ Heart Fail 2022;15(11):E009473. doi: 10.1161/CIRCHEARTFAILURE.122.009473.

73. Minot JR, Cheney N, Maier M, Elbers DC, Danforth CM, Dodds PS. Interpretable Bias Mitigation for Textual Data: Reducing Genderization in Patient Notes While Maintaining Classification Performance. ACM Trans Comput Healthcare 2022;3(4):1-41. doi: 10.1145/3524887.

74. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open 2022;5(9):e2233946. doi: 10.1001/jamanetworkopen.2022.33946.

75. Seker E, Talburt JR, Greer ML. Preprocessing to Address Bias in Healthcare Data. Stud Health Technol Inform 2022;294:327-31. doi: 10.3233/SHTI220468.

76. Sikstrom L, Maslej MM, Hui K, Findlay Z, Buchman DZ, Hill SL. Conceptualising fairness: three pillars for medical algorithms and health equity. BMJ Health and Care Informatics 2022; 29(1):e100459. doi: 10.1136/bmjhci-2021-100459.

77. Straw, I., Wu, H. Investigating for bias in healthcare algorithms: A sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health Care Inform 2022;29(1) 29(1):e100457. doi: 10.1136/bmjhci-2021-100457.

78. Veinot TC, Clarke PJ, Romero DM, Buis LR, Dillahunt TR, Vydiswaran VVG, et al. Equitable Research PRAXIS: A Framework for Health Informatics Methods. Yearb Med Inform 2022;31(1):307-16. doi: 10.1055/s-0042-1742542.

79. Velichkovska B, Gjoreski H, Denkovski D, Kalendar M, Mamandipoor B, Celi LA, et al. Vital signs as a source of racial bias. medRxiv 2022;04. doi: 10.1101/2022.02.03.22270291.

80. Xu J, Xiao Y, Wang WH, Ning Y, Shenkman EA, Bian J, et al. Algorithmic fairness in computational medicine. EBioMedicine 2022 Oct;84:104250. doi: 10.1016/j.ebiom.2022.104250.

81. Yan M, Pencina MJ, Boulware LE, Goldstein BA. Observability and its impact on differential bias for clinical prediction models. J Am Med Inform Assoc 2022;29(5):937-43. doi: 10.1093/jamia/ocac019.

82. Hong C, Pencina MJ, Wojdyla DM, Hall JL, Judd SE, Cary M, et al. Predictive Accuracy of Stroke Risk Prediction Models Across Black and White Race, Sex, and Age Groups. JAMA 2023;329(4):306-17. doi: doi: 10.1001/jama.2022.24683.

83. Park JI, Bozkurt S, Park JW, Lee S. Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer. BMJ Health Care Inform 2023;30(1):e100666. doi: 10.1136/bmjhci-2022-100666.

84. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023 Jan 13;23(1):7. doi: 10.1186/s12911-023-02103-9.

Correspondence to:
Sigurd Maurud, RN, PhD(c)
Department of Public Health Science
Institute of Health and Society
University of Oslo
Pb 1089 Blindern
0318 Oslo
Norway
E-mail: sigurd.maurud@medisin.uio.no