Evaluation of artificial intelligence clinical applications: detailed case analyses shows value of healthcare ethics approach in identifying patient care issues

Journal:	*Bioethics*
Manuscript ID	BIOT-3415-07-20-ART.R1
Manuscript Type:	Original Article
Keywords:	artificial intelligence, ethical evaluation, ethics of new technologies, AI applications in healthcare, ethical frameworks, healthcare ethics

Evaluation of artificial intelligence clinical applications: detailed case analyses show value of healthcare ethics approach in identifying patient care issues

Running header: **Healthcare ethics and artificial intelligence**

Abstract

This paper is one of the first to analyse the ethical implications of specific healthcare artificial intelligence (AI) applications, and the first to provide detailed analysis of AI-based systems for clinical decision support. AI is increasingly being deployed across multiple domains. In response, a plethora of ethical guidelines and principles for general AI use have been published, with some convergence about which ethical concepts are relevant to this new technology. However, few of these frameworks are healthcare specific and there has been limited examination of actual AI applications in healthcare.

Our ethical evaluation identifies context- and case-specific healthcare ethical issues for two applications, and investigates the extent to which the general ethical principles for AI-assisted healthcare expressed in existing frameworks capture what is most ethically relevant from the perspective of healthcare ethics. We provide a detailed description and analysis of two AI-based systems for clinical decision support (Painchek® and IDx-DR). Our results identify ethical challenges associated with potentially deceptive promissory claims, lack of patient and public involvement in healthcare AI development and deployment, and lack of attention to the impact of AIs on healthcare relationships.

Our analysis also highlights the close connection between evaluation and technical development and reporting. Critical appraisal frameworks for healthcare AIs should include explicit ethical evaluation with benchmarks. However, each application will require scrutiny across the AI life cycle to identify ethical issues specific to healthcare. This level of analysis requires more attention to detail than suggested by current ethical guidance or frameworks.

Key words

Artificial Intelligence

AI applications in healthcare

Ethical frameworks

Ethics of new technologies

Healthcare ethics

Ethical evaluation

Introduction and Background

Development of artificial Intelligence (AI)—particularly machine learning—is occurring rapidly in many fields, including healthcare. These developments have been matched by an equally rapid proliferation of AI ethics guidelines.[1] Most of the ethical guidance consists of high-level principles, with few detailed ethical analyses of actual use-cases.

Current and emerging ethical principles and frameworks for AI ethics provide normative guidance for the development and use of AI across a wide range of applications and settings. Despite variability, there is some convergence around a central set of ethical issues. Floridi and Cowls, for example, claim that four principles from bioethics (beneficence, non-maleficence, autonomy, justice), plus a new AI-specific principle regarding explicability (how does it work and who is responsible) are sufficient to evaluate AI development and use.[2] A more comprehensive scoping review identifies eleven commonly recurring principles.[3] Five of these occur in over 50% of the 84 documents the authors analysed: transparency, justice and fairness, non-maleficence, responsibility and privacy. Despite this apparent convergence around key principles, the authors observe that there are "significant semantic and conceptual divergences" regarding interpretation of concepts, scope of applicability, ethical underpinnings and actions required by these principles.[4] For example, the authors note,

---

[1] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. *1*(9), 389–399; Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review,* 1(1). https://doi.org/10.1162/99608f92.8cd550d1.

[2] Floridi & Cowls, o*p. cit.* note 1, p. 5.

[3] Jobin, A. et al. *op. cit*. note 1, pp. 389–399.

[4] Ibid: p. 391.

transparency is used to refer to technical aspects regarding the explainability of an AI and/or to communications aimed at ensuring that those affected by the operation of the AI are aware that AI is being used. Similarly, appeals to justice and fairness range from avoiding bias and discrimination to redress for those adversely affected by the decisions of an AI to equitable access to AI-assisted services. We have summarised the principles and their interpretations in table 1.

Table 1 around here

As yet, guidelines and ethical principles focused on healthcare AI are less common.[5] In terms of formal principles, the UK Government has promulgated a *Code of conduct for data-driven health and care technology*[6] and the Royal Australian and New Zealand College of Radiologists (RANZCR) has developed its own guidance.[7] Bodies such as the UK Academy of

[5] Nuffield Council on Bioethics. (2018). *Artificial Intelligence (AI) in healthcare and research*. Retrieved from https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research [Accessed 21 July, 2020]; Academy of Royal Medical Colleges. (2019). *Artificial Intelligence in Healthcare*. Retrieved from https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/ [Accessed 21 July, 2020]; U.K. Government. (2019). *Code of conduct for data-driven health and care technology*. Retrieved from https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology [Accessed 21 July, 2020]; The Royal Australian and New Zealand College of Radiologists [RANZCR]. (2019). *Ethical Principles for Artificial Intelligence in Medicine*. Retrieved from https://www.ranzcr.com/documents/4952-ethical-principles-for-ai-in-medicine/file [Accessed 21 July, 2020].

Medical Royal Colleges[8] and the Nuffield Council on Bioethics[9] have published relevant discussion documents. Alongside these documents, there is an emerging literature exploring the ethical issues raised by the application of AI to healthcare.[10]

There has been little ethical analysis of AI use-cases in healthcare. This paper responds to calls in the literature to cease producing abstract principles and frameworks, and instead produce detailed analysis of concrete and currently deployed AI applications to better understand ethical tensions and identify any novel ethical issues,[11] particularly in the context of healthcare.

---

[6] U.K. Government, *op. cit.* note 2.

[7] RANZCR, *op. cit.* note 5.

[8] Academy of Royal Medical Colleges, *op. cit.* note 5.

[9] Nuffield Council on Bioethics, *op. cit.* note 5.

[10] See for example, Future Advocacy. (2018). *Ethical, social, and political challenges of artificial intelligence in health*. London: Wellcome Trust. Retrieved from https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf [Accessed 21 July, 2020]; NHS England. (2019, February). *The Topol Review. Preparing the healthcare workforce to deliver the digital future*. Retrieved from https://topol.hee.nhs.uk/the-topol-review/ [Accessed 21 July, 2020]; Ho, A., & Quick, O.

We provide the first detailed analysis of AI-based systems for clinical decision support, one of the rapidly growing areas of healthcare AI. Our two use-cases, Painchek® and IDx-DR,[12] are examples of AI-decision-support applications that have received regulatory approval and are currently used in the clinical care of patients, making them 'real world' exemplars. IDx-DR is an autonomous AI (i.e. it reaches a diagnosis and recommendation without human intervention) while Painchek® is assistive (i.e. the AI aids human decision-making by

---

(2018). Leaving patients to their own devices? Smart technology, safety and therapeutic relationships. *BMC Medical Ethics*. *19*(1), 18; Ho, A. (2019). Deep Ethical Learning: Taking the Interplay of Human and Artificial Intelligence Seriously. *Hastings Center Report*. *49*(1), 36–39; Carter, S.M., Rogers, W., Win, K.T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*. *49*, 25–32; Fenech, M.E., & Buston, O. (2020). AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges. *Frontiers in Cardiovascular Medicine*. *7*, 54; Braun, M., Hummel, P., Beck, S., & Dabrock, P. (2020). Primer on an ethics of AI-based decision support systems in the clinic. *Journal of Medical Ethics*. Published Online First: 03 April 2020. doi: 10.1136/medethics-2019-105860.

[11] Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for researc*h. London: Nuffield Foundation; Leslie, D. (2020). Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction. *Harvard Data Science Review*. Retrieved from https://hdsr.mitpress.mit.edu/pub/as1p81um [Accessed 21 July, 2020].

[12] Painchek® *Intelligent Pain Assessment*. (n.d.). Retrieved from https://painchek.com/ [Accessed 8 Feb, 2021]; *IDx-DR Overview*. (2018). Retrieved from https://dxs.ai/products/idx-dr/idx-dr-overview/ [Accessed 8 Feb, 2021].

automating some processes), thereby capturing some of diversity in AI applications in healthcare.

Section 1 describes our aims and methods. Section 2 describes the two use-cases, Painchek® and IDx-DR. Section 3 reports our analysis of ethical issues arising from these cases. In Section 4, we discuss the implications of our findings.

Section 1: Aims and methods

As methods for AI use-case analysis are not well established, we constructed our own with the goal of providing comprehensive descriptions of Painchek® and IDx-DR (section 2), together with an inductive analysis of the ethical issues identified. We based the case descriptions and analysis on publicly available materials reporting on the development, evidence-generation and deployment phases of Painchek® and IDx-DR in academic articles, regulatory documents and websites.

We used a strategy developed by an expert librarian to search relevant databases (Medline, Embase, PsycINFO, Scopus, and Web of Science); regulatory sites (US Food and Drug Administration [FDA], Australian Therapeutic Goods Administration) [TGA]; product websites for Painchek® and IDx-DR; and ancillary sites identified from the product sites and/or Google searching using the names of the products. We stopped searching on 30 March 2020. Our final data set for Painchek® was five academic papers,[13] one regulatory document,[14] the

[13] Atee, M., Hoti, K., Parsons, R., & Hughes, J.D. (2017). Pain Assessment in Dementia: Evaluation of a Point-of-Care Technological Solution. *Journal of Alzheimer's Disease*. *60*(1), 137–150; Atee, M., Hoti, K., Parsons, R., & Hughes, J. (2018a). A novel pain assessment tool incorporating automated facial analysis: interrater reliability in advanced dementia. *Clinical Interventions in Aging*. *13*, 1245–1258; Atee, M., Hoti, K., & Hughes, J.D. (2018b). A Technical Note on the PainChek™ System: A Web Portal and Mobile Medical Device for Assessing Pain

Painchek® website,[15] and one media report.[16] For IDx-DR, the final set was three academic articles,[17] one regulatory document,[18] the IDx-DR webpage,[19] a You Tube video,[20] and an FDA media release.[21]

WR wrote up the case studies, which were then reviewed by HD, SMC and a research assistant familiar with the cases. The case studies follow a template describing the aims of the AIs, how they work, the evidence base, regulation and data handling, and funding and other issues. These were written at a level of detail such that readers could understand the basis for claims made in the analysis. Regarding the analysis, we worked inductively informed by our existing expertise. We did not construct a deductive coding frame as our goal was to identify particular issues raised by these use-cases. All authors are experienced

---

in People With Dementia. *Frontiers in Aging Neuroscience*. *10*, 117; Atee, M., Hoti, K., & Hughes, J.D. (2018c). Psychometric Evaluation of the Electronic Pain Assessment Tool: An Innovative Instrument for Individuals with Moderate-to-Severe Dementia. *Dementia and Geriatric Cognitive Disorders*. *44*(5–6), 256–267; Hoti, K., Atee, M., & Hughes, J. (2018). Clinimetric properties of the electronic Pain Assessment Tool (ePAT) for aged-care residents with moderate to severe dementia. *Journal of Pain Research*. *Volume 11*, 1037–1044.

[14] Therapeutic Goods Administration. (2018). *Public Summary: Painchek Ltd*. Retrieved from http://search.tga.gov.au/s/search.html?collection=tga-artg&profile=record&meta_i=302794 [Accessed 21 July, 2020].

[15] Painchek®, *op cit*. note 12.

bioethicists with interests in clinical ethics and have been involved in research on AI ethics[22] and through this are familiar with the existing AI ethics literature. We relied on multiple coders and discussion to ensure the reliability of our analysis. Each use-case was analysed independently by two authors. WR analysed both cases; HD and SMC analysed one case

---

[16] Anon. (2019, December 1). The Non-Executive Chairman of PainChek Ltd. (ASX:PCK), John Murray, Just Sold 50% Of Their Holding. *Simply Wall St*. Retrieved from https://simplywall.st/stocks/au/healthcare/asx-pck/painchek-shares/news/the-non-executive-chairman-of-painchek-ltd-asxpck-john-murray-just-sold-50-of-their-holding/ Accessed 8 Feb 2021].

[17] Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., & Folk, J.C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*. *1*(1), 39; Van Der Heijden, A.A., Abràmoff, M.D., Verbraak, F., Hecke, M.V., Liem, A., & Nijpels, G. (2018). Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmologica*. *96*(1), 63–68; Verbraak, F.D., Abràmoff, M.D., Bausch, G.C.F., Klaver, C., Nijpels, G., Schlingemann, ReinierO., & van der Heijden, A.A. (2019). Diagnostic Accuracy of a Device for the Automated Detection of Diabetic Retinopathy in a Primary Care Setting. *Diabetes Care*. *42*(4), 651–656.

[18] US Food and Drug Administration. 2018. Device Classification Under Section 513(f)(2)(De Novo). Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN180001 [Accessed 8 Feb 2021].

[19] Digital Diagnostics. (2020). IDx-DR Overview: Close Care Gaps, Prevent Blindness. Retrieved from https://dxs.ai/products/idx-dr/idx-dr-overview/ [Accessed 8 Feb 2021].

each. Each author prepared a detailed memo, noting the ethical issues that they held to be important in each case, and provided evidence and argumentation to support each issue. These memos were then compared, commonalities noted, and any differences discussed until they could be resolved. This produced the final list of issues for each use-case, presented in chronological order across the AI lifecycle (in Section 3).

Section 2: Case studies

PainChek®

Aims of PainChek®

PainChek® was developed with the aim of improving the quality of pain management for non-verbal individuals such as those with severe dementia. Non-verbal patients cannot self-report pain, making it difficult for carers to assess their pain-relief needs. Under-treatment of pain can lead to adverse consequences including suffering, psychological trauma, behavioural disturbances and poor quality of life. There is a reported lack of accredited

---

[20] Abramoff. 2019. IDx-DR- How it works. Retrieved from https://www.youtube.com/watch?v=dWiF8THxf7Q [Accessed 21 July 2020]

[21] US Food and Drug Administration. 2018. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [Accessed 21 July 2020].

[22] Carter et al., *op. cit*. note 10; Draper, H., Schwartz, L., Racoceanu, D., Rogers W.A. Ethical Futures and AI Medicine, Workshop funded by CIFAR (Canadian Institute for Advanced Research) at the University of Warwick, 25-6 September 2019. All the authors gave invited plenaries at this workshop.

measures for quantifying pain, lack of uptake of available pain assessment tools, and potential variations between carers' subjective assessments about patients' levels of pain, leading to risk of under-treatment of pain for these patients.[23] PainChek® developers postulate that automation of pain assessment processes will make assessments more objective and less prone to error.[24] PainChek® is currently being used in Australian residential aged care facilities for the management of pain in residents with moderate to severe dementia who cannot verbalise, with plans to develop a version for use in assessing pain in pre-verbal children.

How does it work?

PainChek® is an assistive AI that generates a pain score for individuals based on 42 items in a pain scale, collected into 6 domains. AI is used in to generate the data for nine items in domain one, using automated facial recognition technology. The person doing the pain assessment records a short video of the resident's face and uploads this to a cloud-hosted web application. The algorithm identifies nine facial micro-expressions derived from a classification of pain-relevant expressions called the Facial Action Coding System.[25] The facial expressions are validated indicators of pain in both patients with dementia and cognitively

[23] Corbett, A., Husebo, B., Malcangio, M., Staniland, A., Cohen-Mansfield, J., Aarsland, D., & Ballard, C. (2012). Assessment and treatment of pain in people with dementia. *Nature Reviews Neurology*. *8*(5), 264–274.

[24] Atee et al., (2017), *op cit*. note 13.

[25] Ekman, P., Friesen, W., & Hager, J. (1978). The Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, CA.: Consulting Psychologists Press.

unimpaired individuals.[26] The data from the facial analytics comprise domain 1. Domains 2-6 are made up of 33 items derived from American Geriatric Society [AGS] Indicators of Persistent Pain [2002] to assess pain at the point of care.[27] Data for domains 2-6 are manually entered into the smart phone app by the assessor based on observations of the resident's appearance and activities, in the form of binary (yes/no) responses (see Box 1). There is no information on the weighting of the items.

[26] Kunz, M., Seuss, D., Hassan, T., Garbas, J.U., Siebers, M., Schmid, U., … Lautenbacher, S. (2017). Problems of video-based pain detection in patients with dementia: a road map to an interdisciplinary solution. *BMC Geriatrics*. *17*(1), 33.

[27] Atee et al., (2018a), *op cit*. note 13.

Box 1: Domains in the PainChek® pain scale

Domain 1 (assessed by AI): Facial expression (brow lowering, cheek raising, tightening of eyelids, wrinkling of nose, raising of upper lip, pulling at corner lip, horizontal mouth stretch, parting lips, closing eyes)
Domain 2: Voice (pain sounds e.g. ouch, groaning, moaning, crying, screaming, talking)
Domain 3: Movement (altered or random arm or leg movement, restlessness, freezing, guarding/touching body parts, moving away, abnormal standing or walking)
Domain 4: Behavior (changes in interpersonal, mental status changes, aggression, confused, distressed, dislike of touch, fear)
Domain 5: Activity (resisting care, altered sleep cycle, prolonged resting)
Domain 6: Physical signs (fever, rapid breath, red face/flushed, painful injuries or painful medical conditions)

Based on the presence or absence of these 42 indicators, PainChek® produces a pain score that falls within one of four categories reflecting no pain (score 0-6), mild pain (score 7-11), moderate pain (score 12-15) and severe pain (score 16-42). The assessor uses the score to make decisions about administering pain relief, and by repeating the assessment, records the individual's response to any treatment.

Evidence-base

The evidence-base for PainChek® relies on a series of papers published by the team that developed the initial electronic Pain Assessment Tool (ePAT) and later PainChek® system

consisting of the mobile application and web portal.[28] The research compares ePAT with a paper-based pain evaluation tool called the Abbey Pain Score (APS). The studies report that PainChek® scores correlate with APS scores indicating validity, and that there is good inter-rater reliability and high internal consistency (i.e. PainChek® items that are designed to measure the same construct generate similar scores, suggesting that it measures what it purports to measure). However, the research is at high risk of bias for several reasons. Raters were largely familiar with the tool and the residents; they were not blinded to residents' diagnoses or management; and the studies were small (maximum of 40 participants). There is no reported piloting of the tool with cognitively intact individuals; the APS was used as the comparator despite the ePAT allegedly being developed to overcome the limitations of the APS; and all but the first paper received funding from the companies created to commercialise the product (EPAT Technologies Ltd, ePat Pty Ltd, PainChek Ltd). There has been no independent evaluation of PainChek® and no published evidence of evaluations performed after the initial research in 2017.

Regulation, approvals and data handling

PainChek® received regulatory clearance in 2017 from the Australian Therapeutic Goods Administration (TGA) and the Conformité Européene (CE) mark as a 'Medical Device Included Class 1' (the lowest risk category, which also includes tongue depressors and surgical retractors). It was approved with the intended purpose of being "used to assess and monitor pain in people who cannot verbalise such as people with dementia or communication difficulties".[29]

[28] Ibid.; Atee et al., (2017), *op cit*. note 13; Atee et al., (2018b), *op cit*. note 13; Atee et al., (2018c), *op cit*. note 13; Hoti et al., *op cit*. note 13.

[29] Therapeutic Goods Administration, *op cit*. note 14.

The data collected by PainChek® are managed via a web administration portal hosted on Amazon Web Services using the Amazon Elastic Compute Cloud.[30] The PainChek® website states that operators of the app must have the consent of the person they are assessing, provides a link to the privacy policy and lists potential uses of the anonymised data collected by the app including trend analysis, app usage, clinical reviews, content for educational programmes and/or content for publication.[31]

Funding and other issues

In 2019, the Australian Federal government gave $5m in grant funding under the Dementia and Aged Care Services program to PainChek®. The funds comprise $500,000 for training materials and an evaluation report, $4.4m for 100,000 PainChek® licenses for use with people living with dementia across residential aged care in Australia, and $100,000 for a report at the end of the contract term.[32] In late 2019, the non-executive chairman and the CEO both sold large quantities of shares in PainChek Ltd (ASX:PCK). One analyst interpreted this to indicate a possible lack of confidence in the company,[33] although sellers might have a variety of reasons other than lack of confidence for selling stock. After the sell off, the share price dropped from a high of around 0.370, reaching a low of 0.061 in March 2020.

[30] Attee et al. (2018b), *op. cit*. note 13, p. 4.

[31] *Painchek consent information*. (2018). Retrieved from http://painchek.com/data-consent-policy/ [Accessed 21 July, 2020].

[32] *PainChek ASX Announcements*. (n.d.). Retrieved from https://painchek.com/asx-announcements/ [Accessed 21 July, 2020].

[33] Anon., *op cit*. note 16.

IDx-DR

Aims of IDx-DR

IDx-DR is an autonomous AI system for the automatic detection of early signs of eye disease in diabetic patients.[34] IDx-Dr aims to bring specialist level diagnostics to primary care, thereby increasing access to, and decreasing the cost of, diabetic eye care based on the premise that improved access will facilitate better eye care, and ultimately fewer cases of blindness. Diabetic retinopathy (DR) is a known complication of diabetes and a leading cause of blindness worldwide. For early detection and optimal management of diabetic retinopathy, patients require regular eye examinations. Those with no or mild DR can be followed with annual screening, while those with more than mild DR and/or diabetic macular edema (DME) require specialist evaluation for management to avoid damage to vision. Despite the risk of becoming visually impaired, fewer than 50% of diabetic patients receive the recommended screening. Screening usually requires a separate appointment, potentially distant from the primary care provider, and may involve dilatation of the pupils, which can be uncomfortable and affect vision for an hour or longer. IDx-DR offers a specialist-level diagnostic service in primary care. It is currently in use in multiple locations around the United States.[35]

How does it work?

IDx-DR analyses retinal images to provide a diagnosis that classifies patients according to the presence or absence of more than mild DR and/or DME (referred to as mtmDR). Patients who are negative for mtmDR are recommended annual screening. Those with more than

[34] Abràmoff et al., *op cit*. note 17.

[35] IDx-DR, *op. cit*. note 12.

mtmDR are recommended for referral for specialist eye care. The diagnosis is delivered within one minute.

IDx-DR is deployed in primary care. Operators with no previous relevant experience require a single four-hour training session to operate IDx-DR. The operator uses a non-mydriatic retinal camera (i.e. patients do not require dilatation of pupils) to take two images of each retina. The images are immediately evaluated by the first of two algorithms for image quality. Images of adequate quality are analysed by the second algorithm. In the testing protocol, the operator could make two further attempts at imaging followed by three attempts with dilatation to obtain an adequate image.

The second, diagnostic algorithm is described as "clinically inspired" as it is trained to detect the characteristic lesions of DR that a clinician would look for.[36] The training of algorithm two is described in some detail. It has independent validated detectors to identify relevant lesions including microaneurysms, hemorrhages and lipoprotein exudates. Detectors have been implemented as multilayer convolutional neural networks (CNN). CNNs are a type of deep learning algorithm that can differentiate and assign importance to various aspects of an image. Each of the CNNs was independently trained and validated to detect its assigned lesions from a region of a retinal image, using a total of over 1 million lesion patches from retinal images from people with and without DR. The AI fuses the outputs of the detectors into a disease-level diagnosis. As the algorithm was trained to identify the visual indicators that human clinicians rely on, the authors describe it as transparent and interpretable. It is a locked algorithm so its diagnostic processes cannot change. The source codes for IDx-DR are copyrighted by the parent company IDx LLC and are not publicly available.

[36] Abràmoff et al., *op. cit.* note 17, p. 4.

Evidence-base

The foundational evidence-base for IDx-DR is an observational study of 900 participants published in 2018.[37] Each participant underwent two examinations. The first, by the IDx-DR involved 4 retinal images analysed by the algorithm. The second involved pupil dilatation, retinal photos, images of the anterior chamber of the eye, and optical coherence tomography performed by independent photographers certified by the Wisconsin Fundus Photography Reading Centre (FPRC). Three experienced FPRC-validated readers then graded the latter images according to a protocol. The readers were blinded to the AI diagnoses. The diagnostic outputs of IDx-DR were compared with outputs from the human readers, which were used as the gold standard or ground truth. The AI system had sensitivity of 87.2% (i.e. 87.2% of the images classified by the human gold standard process as *having* mtmDR received the same classification from the AI) and specificity of 90.7% (i.e. 90.7% of the images classified by the human gold standard process as *not having* mtmDR received the same classification from the AI). These figures indicate that the AI had acceptable accuracy in identifying both those who had mtmDR and those who did not. The imageability rate was 96.1% (i.e. 4% of patients were not successfully imaged).

An independent contract research organization managed this study, and an algorithm integrity provider locked the system and blocked access by the sponsor to all results until the trial ended. Two further studies (of the IDx-DR-EU-2.1) in the Netherlands with over 3000 patients have validated and replicated these results.[38]

[37] Ibid.

[38] Van Der Heijden et al., *op cit*. note 17; Verbraak et al., *op cit*. note 17.

Regulation, approvals and data handling

IDx-DR received FDA approval as an autonomous medical device in April 2018, the first to be so approved.[39] IDx-DR was granted Breakthrough Device designation and approved through the FDA's De Novo premarket review pathway for some low- to moderate-risk devices that are novel and for which there is no prior legally marketed device. The FDA provided intensive guidance to ensure the project met regulatory requirements regarding statistical design and study endpoints. In addition, IDx-DR has clearance as a Class IIa Medical Device (other Class IIa devices include dental drills and ultrasound machines) for sale in the European Union and received the CE mark from Underwriter's Laboratory in 2013.[40] There is no information on the IDx-DR website about data storage and handling.

Funding and other issues

All three studies to date have been funded by IDx LLC. Several authors, including Abràmoff, have close connections with IDx LLC such as holding shares, holding patents or being an employee.

Section 3: Ethical analysis of the two cases

Our aim was to inductively identify ethical issues arising in the context of these AIs. We present these issues as they arise sequentially in the AI system lifecycle for each use-case, together with the relevant underlying ethical concepts (see Table 2 for a summary).

---

[39] US Food and Drug Administration. (2018). *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [Accessed 21 July, 2020].

[40] IDx-DR, *op. cit*. note 12.

Promissory claims

AI 'hype' refers to promissory claims made in professional and public facing materials. These raise issues of veracity, transparency and trustworthiness, depending upon the extent to which they are backed by evidence. The front page of the Painchek® website has the tagline "Intelligent Pain Assessment" and claims "PainChek® uses AI, facial recognition and smartphone technology to intelligently automate the pain assessment process at the point of care".[41] However, the AI element of PainChek® is modest (one domain out of six; nine out of 42 items on the pain scale), and there is no information about how the pain score is generated to support the claim that it is "intelligently automated". Presenting PainChek® as a novel AI tool fosters what has been called the technological imperative in healthcare, where technical innovations such as AI are seen as superior practice, with the implication that not using the novel intervention signals dated and perhaps inferior care.[42] IDx-DR promotes the verifiable claim that it was the first autonomous AI to receive FDA approval. This claim may also invoke the technological imperative but as it is supported by evidence, does not raise questions of potential deception and lack of transparency raised by the PainChek® hype.

Value of stated goals

The goals of AI use-cases speak directly to questions of beneficence. Healthcare interventions are justified to the extent that they promote the patient's overall interests and contribute to decreasing avoidable morbidity and mortality. Both PainChek® and IDx-DR attempt to address this point. The PainChek® materials do not include evidence that

[41] Painchek®, *op. cit*. note 12.

[42] Burger-Lux, M.J. & Heaney, R.P. (1986) For better and worse: the technological imperative in health care. *Social Science and Medicine*. 22(12),1313e20.

diagnostic failures cause under or overtreament of pain in the relevant population, although the implication is prima facie plausible. IDx-DR materials include evidence of potentially avoidable morbidity from inadequate diabetic eye examinations and follow up. The IDx-DR authors explicitly state that a major motivation for developing the AI is to increase access and reduce costs, thereby appealing to social justice issues of access and equity as well as beneficence.

Information about the algorithm

Information about the algorithm and its development is ethically relevant for reasons including its understandability or interpretability, and hence the extent to which physicians might trust it. IDx-DR provides evidence to support its claim that it is clinically inspired. For PainChek®, this issue is hard to assess as there is no information about the contribution of the AI elements to the overall pain score, which is however based on clinical indicators used in other pain assessment tools. Technical details about the development of algorithms (such as the training sets used) have justice implications as they indicate the potential for unfair bias. The pivotal IDx-DR paper contains a description of the development of the diagnostic algorithm and addresses issues of potential bias with evidence of attention to gender and racial diversity. There is no information about the training set used in PainChek and no discussion of potential bias.

Evidence of efficacy

Evidence of efficacy is essential for ensuring that interventions are beneficent: healthcare interventions provide benefit to patients to the extent that they achieve the relevant goals. High quality research is necessary to underpin claims of efficacy. This was lacking in the case of PainChek®. Its trials had low participant numbers and were at high risk of bias due to lack of blinding and lack of independent operators of PainChek® during the research. The research underpinning IDx-DR seemed rigorous and provided proof of efficacy. Neither

PainChek® nor IDx-DR included patient-reported outcome measures (PROMs), which are increasingly used to ensure that healthcare interventions achieve outcomes that are valued by patients.[43]

Potential for harm

Providing information about the potential harms of healthcare interventions is essential to meet the ethical requirement of non-maleficence. Patients screened by IDx-DR may be harmed compared to screening by an ophthalmologist if they have as yet undetected conditions that IDx-DR is not trained to identify, such as glaucoma or macular degeneration. However, they may be better off overall if the comparator is no eye exam. Patients whose pain is managed by use of PainChek® may be benefited if it accurately identifies a need for more pain relief, or harmed if their subsequent pain relief is less adequate than that provided by normal care, or leads to overtreatment. This comparative information is not provided.

Regulation, legal liability and managing interests

Regulatory approvals of new interventions indicate a level of independent assessment and can therefore foster both practitioner and patient trust. Both AI devices have regulatory approvals, but the degree of oversight varies between them. PainChek® is approved as a class 1 device which has very little oversight regarding safety or efficacy. IDx-DR is approved as a low to moderate risk De Novo device which does require a specific evidence base. As patients may not be aware of the different classes of regulatory approval and associated guarantees of safety and efficacy, clarifying the evidence required for different types of approval would enhance transparency.

[43] Black, N. (2013). Patient reported outcome measures could help transform healthcare. *BMJ*. 346, f167.

Providing information on legal liability for harm ensuing from use of the AI demonstrates responsibility and is required by justice. This information was not available for either AI on 30 March when we closed data collection. Since then, the company responsible for IDx-DR has indicated that it is legally liable for the accuracy of the diagnosis.[44]

Conflicts of interest undermine patient care when they lead to bias about the efficacy of and need for interventions, with potential detriment to patients' interests and wellbeing. In both these case studies, companies with a financial interest in the success of the AIs funded relevant research. With PainChek®, most of the authors are commercially involved with the product and were directly involved in the research. No measures to guard against the effects of the financial conflicts were reported. Attempts were made to manage and minimise conflicts of interest related to company funding of IDx-DR research through the use of a contract research organisation and Algorithm Integrity Provider.

Management of information and data

AI applications in healthcare rely on patient data, such as images, test results or biometric data. Patient autonomy can be supported by requiring informed consent for data collection and management, and controlling access to collected and stored patient information in ways that protect patients' privacy. PainChek® uses facial images and pain scores which are highly personal, making any data breach a potentially serious breach of privacy. It is unclear how consent was obtained for participants in the PainChek® trials, given that many had cognitive impairments and may not have been able to provide consent themselves. The PainChek® authors claim that data are held securely but their data are hosted on the Amazon cloud

[44] Abràmoff, M.D., Tobey, D., & Char, D.S. (2020). Lessons Learned About Autonomous AI: Finding a Safe, Efficacious, and Ethical Path Through the Development Process. *American Journal of Ophthalmology*. *214*, 134–142.

which has been subject to security breaches.[45] The IDx-Dr materials indicate that an independent Algorithm Identity Provider managed the research data, implying data security, but no further details are provided. Patients gave informed consent for participation in the research. There is no information about where the data are stored or potential secondary uses.

Justice and care

The impact of these AI tools on the delivery of care raises questions about the relationships in healthcare and about social justice and equity issues. Neither of the cases provide information about the impact on delivery of care, such as the amount of time it takes to use PainChek® compared to the alternative paper-based tool, or the impact of providing IDx-DR in a primary care setting. In both cases, use of the AI changes the carer and patient relationship. With IDx-DR, the ophthalmologist is eliminated from the examination process if the screening is negative with the risk of missing incidental findings. In both the specialist and AI situation, a human takes the retinal images, but there is no information on patients' experiences of and views about being screened by IDx-DR. PainChek® does not eliminate a human from the care process. The impact on the relationship between carer and patient of the use of a smart phone compared to a paper-based tool is not reported and may be variable. Looking at a video of the resident's face may lead to the carer paying more attention to them, but on the other hand, the carer may be inclined to leave the decision to

[45] See e.g. Scroxton, A. (2020). Exposed AWS buckets again implicated in multiple data leaks. Retrieved from https://www.computerweekly.com/news/252476870/Exposed-AWS-buckets-again-implicated-in-multiple-data-leaks [Accessed 10 February 2021).

the app and administer pain relief based on the pain score generated by PainChek® rather than anything the resident says or does, reflecting automation bias.[46]

Healthcare interventions have social justice implications as their nature and delivery may exacerbate or mitigate health inequities. Although a stated goal of IDx-DR is to improve access to diabetic eye care and reduce costs, increased access and lower cost have not been demonstrated in practice.[47] There is no information about, or discussion of, the impact of PainChek® on equity of access to healthcare or reducing health inequalities. As mentioned above, the development and training of the algorithm may introduce, exacerbate or minimise racial and gender biases.

Costs and opportunities

Finally, ethical questions about cost and opportunity arise because new interventions tend to be more expensive than existing ones and should therefore be required to demonstrate improved health outcomes and/or greater efficiencies compared with existing care. There is no information about the cost of using PainChek® compared to usual care or, if outcomes are improved, the incremental cost of these improvements. PainChek® is currently offering free twelve-month licenses, funded by the Australian Government, for all Australian aged care residential facilities. The cost of the license is not revealed on its website but it represents an additional cost to usual care unless use of the app significantly reduces carer time, or the cost is warranted by improved outcomes for residents regarding their pain management.

---

[46] Gretton, C. (2017). The dangers of AI in health care: risk homeostasis and automation bias. Retrieved from https://towardsdatascience.com/the-dangers-of-ai-in-health-care-risk-homeostasis-and-automation-bias-148477a9080f [Accessed 10 February 2021].

[47] Savoy, M. (2020). Diagnostic Tests: What Physicians Need to Know IDx-DR for Diabetic Retinopathy Screening. *American Family Physician*. 101(5), 307-308.

There is no information about the cost of using IDx-DR compared to usual care, or the incremental cost of any improvements in health outcomes.

Table 2 around here

Section 4: Discussion

Attending to the context of both development and implementation of AI use-cases in healthcare has led us to identify a range of ethical issues, some of which have had little emphasis to date in general AI ethics frameworks. Here we focus on three of these: veracity and deception, public and patient involvement (PPI) and healthcare relationships. Veracity and deception arise in the context of AI hype. Honest presentation of AI in healthcare matters because both respect for autonomy and acting in patients' best interests require a commitment to honesty, which is a fundamental value in the practitioner-patient relationship. Healthcare algorithms however, are often developed in the context of competitive venture capitalism, the values of which differ from, and may be incompatible with, the values of healthcare. This observation suggests the need to critically evaluate new healthcare AI technologies in their social, legal and economic contexts as well as in the clinic. While veracity and deception relate to the broader concepts of transparency and trustworthiness, both of which appear in the AI ethics literature, the particular issue of hype has not previously been emphasised in AI ethics frameworks.

Any lack of involvement of patients and/or the public in developing AI for healthcare raises concerns.[48] Given the potential impact of AI-assisted healthcare on patients' experiences and outcomes, there is a strong ethical mandate for PPI in the commissioning, design, deployment and evaluation of healthcare AIs. The absence of PPI risks the development of AI to address problems that are amenable to AI solutions and/or likely to be profitable, rather

[48] Ho, *op. cit*. note 10, p. 36; Future Advocacy, *op. cit.* note 10, p. 36.

than addressing issues that are important for patients. We found no evidence of PPI in the development and deployment of IDx-DR or PainChek®.

Clinical AI applications have consequences for healthcare relationships. Potential adverse effects include impaired communication, loss of trust, or conflicted decision making, especially if there are discrepancies between advice from the AI and from the relevant clinician. Potential benefits for patients include freeing clinicians from routine and largely administrative tasks to focus more on humanistic aspects of care including communication and healing.[49] Addressing this point requires attention to detail as the potential impact varies with the type of AI. There was no information about patient experiences of and responses to care provided using IDx-DR or PainChek®. Including patient reported outcomes measures (PROMs) in the research base for new healthcare AIs is critical to address this point.

Other issues that we identify have been flagged in existing AI-ethics frameworks. However, the process of identifying issues at each stage of development and deployment ensures that no issues are neglected. Additonally, this approach has practical utility as it can inform attribution of responsibility for attending to the issues where they arise in the AI life-cycle. In contrast, top-down principles such as transparency or trustworthiness have little clarity or effect unless they are explicitly linked to particular features of the AI. Our findings support the claim that there is wide scope for interpretation of ethical concepts, how and where they apply, and actions required to support ethical practice.[50]

[49] Topol, E. (2019). *Deep Medicine: how Artificial Intelligence can make Medicine Human again*. New York: Basic Books; Ho, *op. cit*. note 10, p. 37; NHS England, *op. cit*. note 10.

[50] Jobin et al., op cit. note 1: p. 391.

Finally, our analysis points to the close connection between ethical evaluation for healthcare and technical reporting. Details about the training set, the way the algorithm is constructed, sensitivity and specificity, data storage and so forth are essential for making ethical evaluations. For example, while data privacy is a well-recognised concern in the use of AI interventions, the level of concern relates to the nature of the data. PainChek® collects and stores facial images which reveal who the person is, their dementia diagnosis, and that they are being assessed for pain management. It is not possible to anonymise data like facial images. Retinal images are also potentially identifying and revelatory of health information,[51] but without access to IT resources it is considerably harder to recognise an individual from their retinal scan than their facial image. This point suggests that ethicists must contribute to multidisciplinary evaluations of healthcare AIs, to ensure accuracy in interpreting the ethical implications of technical specifications.

Conclusion

Our detailed analysis of use-cases illustrates the value of fine-grained examination of specific AI applications in identifying and addressing relevant ethical issues. Further detailed use-cases are required to develop an inventory of ethical issues in practice, supplemented with empirical research to ascertain impacts of specific AIs on patients, publics, healthcare providers and other stakeholders. We note that addressing ethical issues arising from healthcare AIs requires engagement with the values of healthcare throughout the AI development process in order to meet patient and practitioner expectations. General ethical

[51] Poplin, R., Varadarajan, A., Blumer, K., Liu, Y., McConnell, M., Corrado, G., … Webster, D. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*. *2*(3), 158–164.

frameworks for AI may not adequately address healthcare-specific expectations such as beneficence.

Without further work to specify ethical standards, the force of ethical frameworks in domains such as healthcare is unclear. For example, to meet the injunction for healthcare AIs to be beneficent, the meeting of certain outputs or thresholds should be specified (e.g. in terms of reduced morbidity/mortality or increased equity). How these should be developed is an open question but points to the need for PPI involvement as well as studies of AI applications in healthcare to be reported in ways that facilitate specialist healthcare ethics review as well as technical analysis. Proposals for critical appraisal frameworks for healthcare AIs offer opportunities for, and could be strengthened by, including ethical benchmarks,[52] but as we have shown, ethical reflexivity and attention to particulars will also be necessary to make a full evaluation of any application.

There is a need for stronger communication between all stakeholders involved in developing and implementing AIs in healthcare, to gain a shared understanding of both technical limits and context-specific ethical obligations, and work towards solutions. While acting in the patients' best interests and avoiding harm are fundamental to healthcare, these principles may seem out of scope to developers who consequently disregard some relevant benefits and harms in developing their AIs. Inclusion of PROMs is a critical part of this process.

[52] Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J.P.A., & Shah, N.H. (2020). MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *Journal of the American Medical Informatics Association. 27(12),* 2011-2015.

Our paper provides a potential blueprint for further use-case analyses. It is a contribution towards developing robust ethical evaluation of healthcare AI that can be integrated with other appraisal tools. However, much work remains to support ethically robust AI-assisted healthcare.

Table 1. Principles for ethical AI summarised from Jobin et al.

Concept	**Scope of applicability**	**Ethical underpinnings**	**Actions required to uphold principle**
Transparency	Refers to: explainability or interpretability; nature and scope of communication; disclosures	To minimise or avoid harm, improve AI, foster trust, enable engagement/debate	Disclosure of information by those developing or deploying AI (such as AI uses, source code, data use, evidence base, limitations etc.); provision of non-technical explanations; public and stakeholder interactions.
Justice, fairness and equity	Refers to: fairness, minimizing/preventing bias and discrimination, supporting diversity/equity; also procedural issues for appeal, and fair access to AI/data and fair benefits	To ensure equity in access to AI, data and the benefits of AI; minimise harms; provide fair access to redress and remedy; support social justice	Technical solutions (standards, explicit normative coding); transparency; testing and audit; ensuring laws/regulations that are fit for purpose; systemic changes for greater inclusivity
Non-maleficence	Refers to: safety and security; avoiding foreseeable or unintentional harms including to social	To avoid or minimise harms (understood as discrimination, privacy violations, bodily harms); to	Technical solutions and governance including in-built data quality evaluations, in-built security, privacy by design, appropriate

	wellbeing, infrastructure, and psychological, emotional and economic aspects	prevent loss of trust, skills, to prevent other negative impacts	oversight processes and practices, independent audits
Responsibility and accountability	Refers to: acting with integrity, transparent and clear chains of responsibility and legal liability; focusing on how harms may arise; promoting diversity and openness	To minimise or avoid harm; to promote diversity; to attribute responsibility and support justice in seeking redress for harms	Identify relevant actors (AI developers, designers, institutions, industry); clarify degree to which responsibility and liability can be attributed to AI or whether human actors are always responsible
Privacy	Refers to upholding privacy and protecting right to privacy with regard to data protection and security	To protect right to privacy; to respect autonomy (freedom); to build trust	Technical solutions (e.g. privacy by design); more research; improved awareness and regulatory approaches

Table 2. Summary of ethical issues raised by PainChek® and IDx-DR

Issue	**PainChek®**	**IDx-DR**	**Ethical underpinnings**
AI 'hype'	AI contributes only a small portion of the pain score generated.	Accurate claim that this was the first autonomous AI to receive FDA approval.	Veracity/avoiding deception Avoiding unwarranted uptake of new technologies (beneficence/non-maleficence) Building trust
Goal of AI	To improve pain management for patient cohort. No evidence provided that improved diagnosis will lead to better pain management.	To decrease avoidable visual impairment from diabetic retinopathy, to bring specialty level diagnostics to primary care, thereby increasing access and lowering cost.	Beneficence/non-maleficence Equity Social justice
Information about algorithm	Role of the algorithm in generating pain score is not explained. The training set used for the algorithm is not described.	Two algorithms and their roles are described. Training set used for the algorithm is described. Diagnostic algorithm is based on well-established 'racially invariant' biomarkers.	Veracity/avoiding deception Justice/avoiding discrimination Building trust
Evidence-base	Claims of efficacy not well supported by the published research which is at high risk	Claims of efficacy supported by the pivotal trial, and validated by later studies.	Beneficence Veracity/avoiding deception

	of bias. No information about the representativeness of the study participants. No reporting of patient-oriented outcome measures. No data on effectiveness of PainChek® in improving pain management.	Study participants are representative of the US diabetes population. No reporting of patient-oriented outcome measures. No data on effectiveness in reducing visual loss, or of cost-saving or improved access.	Fostering public and patient involvement
Potential for harm	No reports of adverse effects from using PainChek®. No information on adequacy of pain relief using PainChek® compared to usual care.	Patients with concurrent eye disease may feel falsely reassured and/or have delayed diagnosis of incidental conditions that IDx-DR does not identify (noted by authors).	Non-maleficence Building trust
Regulatory approval	Approved in Australia as a Class 1 device.	Approved by the FDA as a low to moderate risk De Novo device, and by CE as a Class II device.	Beneficence/non-maleficence Fostering trust
Legal liability for harm	No information on legal liability for harm ensuing from use of PainChek®.	IDx LLC is legally liable for accuracy of diagnosis, with some caveats.	Attributing responsibility Justice
Conflicts of interest	Most of the authors are commercially invested in the	Some of the authors have financial ties to the relevant	Beneficence/non-maleficence

	product. No evidence of strategies to minimise effects of commercial interests.	company IDx LLC which funded the research. Third parties engaged to minimise effects of commercial interests.	Building trust Impact on relationships of care Veracity/avoiding deception
Data uses	Little information provided about the consent process. Participants have dementia so may be unable to give informed consent. Data protection measures are described as 'secure' but data are hosted on Amazon cloud with potential privacy and security implications.	Participants gave informed consent for the study. No information provided about consent to future data uses. An Algorithm Integrity Provider managed the research data. No information about post-study data storage or any secondary uses other than that data will be available upon 'reasonable request'.	Consent and autonomy Privacy Building trust Veracity/avoiding deception
Impact on delivery of care	Not described. Operators had to be familiar with the patients to use PainChek® but the effect on care relationships is not described.	Not described. Impact on relationships is unclear.	Social justice Equity Impact on relationships of care
Impact on social justice	No information about impact on reducing health inequities.	No information about impact on equity of access to diabetic eye care or reducing health	Social justice Equity

	No evidence of transferability of the app to other racial or ethnic groups.	inequalities. Algorithm is effective in all racial groups.	
Cost and opportunity cost	No information about the cost of using PainChek® compared to usual care, or the incremental cost of any improvements. Overall opportunity cost is not stated.	No information about the cost of using IDx-DR compared to usual care, or the incremental cost of any improvements. Overall opportunity cost is not stated.	Beneficence Resource allocation

Evaluation of artificial intelligence clinical applications: detailed case analyses show value of healthcare ethics approach in identifying patient care issues

Running header: **Healthcare ethics and artificial intelligence**

Abstract

This paper is one of the first to analyse the ethical implications of specific healthcare artificial intelligence (AI) applications, and the first to provide detailed analysis of AI-based systems for clinical decision support. AI is increasingly being deployed across multiple domains. In response, a plethora of ethical guidelines and principles for general AI use have been published, with some convergence about which ethical concepts are relevant to this new technology. However, few of these frameworks are healthcare specific and there has been limited examination of actual AI applications in healthcare.

Our ethical evaluation identifies context- and case-specific healthcare ethical issues for two applications, and investigates the extent to which the general ethical principles for AI-assisted healthcare expressed in existing frameworks capture what is most ethically relevant from the perspective of healthcare ethics. We provide a detailed description and analysis of two AI-based systems for clinical decision support (Painchek® and IDx-DR). Our results identify ethical challenges associated with potentially deceptive promissory claims, lack of patient and public involvement in healthcare AI development and deployment, and lack of attention to the impact of AIs on healthcare relationships.

Our analysis also highlights the close connection between evaluation and technical development and reporting. Critical appraisal frameworks for healthcare AIs should include explicit ethical evaluation with benchmarks. However, each application will require scrutiny across the AI life cycle to identify ethical issues specific to healthcare. This level of analysis requires more attention to detail than suggested by current ethical guidance or frameworks.

Key words

Artificial Intelligence

AI applications in healthcare

Ethical frameworks

Ethics of new technologies

Healthcare ethics

Ethical evaluation

Introduction and Background

Development of artificial Intelligence (AI)—particularly machine learning—is occurring rapidly in many fields, including healthcare. These developments have been matched by an equally rapid proliferation of AI ethics guidelines.[1] Most of the ethical guidance consists of high-level principles, with few detailed ethical analyses of actual use-cases.

Current and emerging ethical principles and frameworks for AI ethics provide normative guidance for the development and use of AI across a wide range of applications and settings. Despite variability, there is some convergence around a central set of ethical issues. Floridi and Cowls, for example, claim that four principles from bioethics (beneficence, non-maleficence, autonomy, justice), plus a new AI-specific principle regarding explicability (how does it work and who is responsible) are sufficient to evaluate AI development and use.[2] A more comprehensive scoping review identifies eleven commonly recurring principles.[3] Five of these occur in over 50% of the 84 documents the authors analysed: transparency, justice and fairness, non-maleficence, responsibility and privacy. Despite this apparent convergence around key principles, the authors observe that there are "significant semantic and conceptual divergences" regarding interpretation of concepts, scope of applicability, ethical underpinnings and actions required by these principles.[4] For example, the authors note,

[1] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. *1*(9), 389–399; Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review,* 1(1). https://doi.org/10.1162/99608f92.8cd550d1.

[2] Floridi & Cowls, o*p. cit.* note 1, p. 5.

[3] Jobin, A. et al. *op. cit*. note 1, pp. 389–399.

[4] Ibid: p. 391.

transparency is used to refer to technical aspects regarding the explainability of an AI and/or to communications aimed at ensuring that those affected by the operation of the AI are aware that AI is being used. Similarly, appeals to justice and fairness range from avoiding bias and discrimination to redress for those adversely affected by the decisions of an AI to equitable access to AI-assisted services. We have summarised the principles and their interpretations in table 1.

Table 1 around here

As yet, guidelines and ethical principles focused on healthcare AI are less common.[5] In terms of formal principles, the UK Government has promulgated a *Code of conduct for data-driven health and care technology*[6] and the Royal Australian and New Zealand College of Radiologists (RANZCR) has developed its own guidance.[7] Bodies such as the UK Academy of

[5] Nuffield Council on Bioethics. (2018). *Artificial Intelligence (AI) in healthcare and research*. Retrieved from https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research [Accessed 21 July, 2020]; Academy of Royal Medical Colleges. (2019). *Artificial Intelligence in Healthcare*. Retrieved from https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/ [Accessed 21 July, 2020]; U.K. Government. (2019). *Code of conduct for data-driven health and care technology*. Retrieved from https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology [Accessed 21 July, 2020]; The Royal Australian and New Zealand College of Radiologists [RANZCR]. (2019). *Ethical Principles for Artificial Intelligence in Medicine*. Retrieved from

Medical Royal Colleges[8] and the Nuffield Council on Bioethics[9] have published relevant discussion documents. Alongside these documents, there is an emerging literature exploring the ethical issues raised by the application of AI to healthcare.[10]

There has been little ethical analysis of AI use-cases in healthcare. This paper responds to calls in the literature to cease producing abstract principles and frameworks, and instead produce detailed analysis of concrete and currently deployed AI applications to better understand ethical tensions and identify any novel ethical issues,[11] particularly in the context of healthcare.

---

https://www.ranzcr.com/documents/4952-ethical-principles-for-ai-in-medicine/file [Accessed 21 July, 2020].

[6] U.K. Government, *op. cit.* note 2.

[7] RANZCR, *op. cit.* note 5.

[8] Academy of Royal Medical Colleges, *op. cit.* note 5.

[9] Nuffield Council on Bioethics, *op. cit.* note 5.

[10] See for example, Future Advocacy. (2018). *Ethical, social, and political challenges of artificial intelligence in health*. London: Wellcome Trust. Retrieved from https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf

We provide the first detailed analysis of AI-based systems for clinical decision support, one of the rapidly growing areas of healthcare AI. Our two use-cases, Painchek® and IDx-DR,[12] are examples of AI-decision-support applications that have received regulatory approval and are

---

[Accessed 21 July, 2020]; NHS England. (2019, February). *The Topol Review. Preparing the healthcare workforce to deliver the digital future*. Retrieved from https://topol.hee.nhs.uk/the-topol-review/ [Accessed 21 July, 2020]; Ho, A., & Quick, O. (2018). Leaving patients to their own devices? Smart technology, safety and therapeutic relationships. *BMC Medical Ethics*. *19*(1), 18; Ho, A. (2019). Deep Ethical Learning: Taking the Interplay of Human and Artificial Intelligence Seriously. *Hastings Center Report*. *49*(1), 36–39; Carter, S.M., Rogers, W., Win, K.T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*. *49*, 25–32; Fenech, M.E., & Buston, O. (2020). AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges. *Frontiers in Cardiovascular Medicine*. *7*, 54; Braun, M., Hummel, P., Beck, S., & Dabrock, P. (2020). Primer on an ethics of AI-based decision support systems in the clinic. *Journal of Medical Ethics*. Published Online First: 03 April 2020. doi: 10.1136/medethics-2019-105860.

[11] Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for researc*h. London: Nuffield Foundation; Leslie, D. (2020). Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction. *Harvard Data Science Review*. Retrieved from https://hdsr.mitpress.mit.edu/pub/as1p81um [Accessed 21 July, 2020].

[12] Painchek® *Intelligent Pain Assessment*. (n.d.). Retrieved from https://painchek.com/ [Accessed 8 Feb, 2021]; *IDx-DR Overview*. (2018). Retrieved from https://dxs.ai/products/idx-dr/idx-dr-overview/ [Accessed 8 Feb, 2021].

currently used in the clinical care of patients, making them 'real world' exemplars. IDx-DR is an autonomous AI (i.e. it reaches a diagnosis and recommendation without human intervention) while Painchek® is assistive (i.e. the AI aids human decision-making by automating some processes), thereby capturing some of diversity in AI applications in healthcare.

Section 1 describes our aims and methods. Section 2 describes the two use-cases, Painchek® and IDx-DR. Section 3 reports our analysis of ethical issues arising from these cases. In Section 4, we discuss the implications of our findings.

Section 1: Aims and methods

As methods for AI use-case analysis are not well established, we constructed our own with the goal of providing comprehensive descriptions of Painchek® and IDx-DR (section 2), together with an inductive analysis of the ethical issues identified. We based the case descriptions and analysis on publicly available materials reporting on the development, evidence-generation and deployment phases of Painchek® and IDx-DR in academic articles, regulatory documents and websites.

We used a strategy developed by an expert librarian to search relevant databases (Medline, Embase, PsycINFO, Scopus, and Web of Science); regulatory sites (US Food and Drug Administration [FDA], Australian Therapeutic Goods Administration) [TGA]; product websites for Painchek® and IDx-DR; and ancillary sites identified from the product sites and/or Google searching using the names of the products. We stopped searching on 30 March 2020. Our final data set for Painchek® was five academic papers,[13] one regulatory document,[14] the

[13] Atee, M., Hoti, K., Parsons, R., & Hughes, J.D. (2017). Pain Assessment in Dementia: Evaluation of a Point-of-Care Technological Solution. *Journal of Alzheimer's Disease*. *60*(1), 137–150; Atee, M., Hoti, K., Parsons, R., & Hughes, J. (2018a). A novel pain assessment tool

Painchek® website,[15] and one media report.[16] For IDx-DR, the final set was three academic articles,[17] one regulatory document,[18] the IDx-DR webpage,[19] a You Tube video,[20] and an FDA media release.[21]

WR wrote up the case studies, which were then reviewed by HD, SMC and a research assistant familiar with the cases. The case studies follow a template describing the aims of the AIs, how they work, the evidence base, regulation and data handling, and funding and other issues. These were written at a level of detail such that readers could understand the basis for claims made in the analysis. Regarding the analysis, we worked inductively informed by our existing expertise. We did not construct a deductive coding frame as our goal was to identify particular issues raised by these use-cases. All authors are experienced bioethicists with interests in clinical ethics and have been involved in research on AI ethics[22]

---

incorporating automated facial analysis: interrater reliability in advanced dementia. *Clinical Interventions in Aging*. *13*, 1245–1258; Atee, M., Hoti, K., & Hughes, J.D. (2018b). A Technical Note on the PainChek™ System: A Web Portal and Mobile Medical Device for Assessing Pain in People With Dementia. *Frontiers in Aging Neuroscience*. *10*, 117; Atee, M., Hoti, K., & Hughes, J.D. (2018c). Psychometric Evaluation of the Electronic Pain Assessment Tool: An Innovative Instrument for Individuals with Moderate-to-Severe Dementia. *Dementia and Geriatric Cognitive Disorders*. *44*(5–6), 256–267; Hoti, K., Atee, M., & Hughes, J. (2018). Clinimetric properties of the electronic Pain Assessment Tool (ePAT) for aged-care residents with moderate to severe dementia. *Journal of Pain Research*. *Volume 11*, 1037–1044.

and through this are familiar with the existing AI ethics literature. We relied on multiple coders and discussion to ensure the reliability of our analysis. Each use-case was analysed independently by two authors. WR analysed both cases; HD and SMC analysed one case each. Each author prepared a detailed memo, noting the ethical issues that they held to be important in each case, and provided evidence and argumentation to support each issue. These memos were then compared, commonalities noted, and any differences discussed

---

[14] Therapeutic Goods Administration. (2018). *Public Summary: Painchek Ltd*. Retrieved from http://search.tga.gov.au/s/search.html?collection=tga-artg&profile=record&meta_i=302794 [Accessed 21 July, 2020].

[15] Painchek®, *op cit*. note 12.

[16] Anon. (2019, December 1). The Non-Executive Chairman of PainChek Ltd. (ASX:PCK), John Murray, Just Sold 50% Of Their Holding. *Simply Wall St*. Retrieved from https://simplywall.st/stocks/au/healthcare/asx-pck/painchek-shares/news/the-non-executive-chairman-of-painchek-ltd-asxpck-john-murray-just-sold-50-of-their-holding/ Accessed 8 Feb 2021].

[17] Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., & Folk, J.C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*. *1*(1), 39; Van Der Heijden, A.A., Abràmoff, M.D., Verbraak, F., Hecke, M.V., Liem, A., & Nijpels, G. (2018). Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmologica*. *96*(1), 63–68; Verbraak, F.D., Abràmoff, M.D., Bausch, G.C.F., Klaver, C., Nijpels, G., Schlingemann, ReinierO., & van der Heijden, A.A. (2019). Diagnostic Accuracy of a Device for the Automated Detection of Diabetic Retinopathy in a Primary Care Setting. *Diabetes Care*. *42*(4), 651–656.

until they could be resolved. This produced the final list of issues for each use-case, presented in chronological order across the AI lifecycle (in Section 3).

Section 2: Case studies

PainChek®

Aims of PainChek®

PainChek® was developed with the aim of improving the quality of pain management for non-verbal individuals such as those with severe dementia. Non-verbal patients cannot self-report pain, making it difficult for carers to assess their pain-relief needs. Under-treatment of pain can lead to adverse consequences including suffering, psychological trauma, behavioural disturbances and poor quality of life. There is a reported lack of accredited measures for quantifying pain, lack of uptake of available pain assessment tools, and

[18] US Food and Drug Administration. 2018. Device Classification Under Section 513(f)(2)(De Novo). Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/denovo.cfm?ID=DEN180001 [Accessed 8 Feb 2021].

[19] Digital Diagnostics. (2020). IDx-DR Overview: Close Care Gaps, Prevent Blindness. Retrieved from https://dxs.ai/products/idx-dr/idx-dr-overview/ [Accessed 8 Feb 2021].

[20] Abramoff. 2019. IDx-DR- How it works. Retrieved from https://www.youtube.com/watch?v=dWiF8THxf7Q [Accessed 21 July 2020]

[21] US Food and Drug Administration. 2018. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [Accessed 21 July 2020].

potential variations between carers' subjective assessments about patients' levels of pain, leading to risk of under-treatment of pain for these patients.[23] PainChek® developers postulate that automation of pain assessment processes will make assessments more objective and less prone to error.[24] PainChek® is currently being used in Australian residential aged care facilities for the management of pain in residents with moderate to severe dementia who cannot verbalise, with plans to develop a version for use in assessing pain in pre-verbal children.

How does it work?

PainChek® is an assistive AI that generates a pain score for individuals based on 42 items in a pain scale, collected into 6 domains. AI is used in to generate the data for nine items in domain one, using automated facial recognition technology. The person doing the pain assessment records a short video of the resident's face and uploads this to a cloud-hosted web application. The algorithm identifies nine facial micro-expressions derived from a

[22] Carter et al., *op. cit*. note 10; Draper, H., Schwartz, L., Racoceanu, D., Rogers W.A. Ethical Futures and AI Medicine, Workshop funded by CIFAR (Canadian Institute for Advanced Research) at the University of Warwick, 25-6 September 2019. All the authors gave invited plenaries at this workshop.

[23] Corbett, A., Husebo, B., Malcangio, M., Staniland, A., Cohen-Mansfield, J., Aarsland, D., & Ballard, C. (2012). Assessment and treatment of pain in people with dementia. *Nature Reviews Neurology*. *8*(5), 264–274.

[24] Atee et al., (2017), *op cit*. note 13.

classification of pain-relevant expressions called the Facial Action Coding System.[25] The facial expressions are validated indicators of pain in both patients with dementia and cognitively unimpaired individuals.[26] The data from the facial analytics comprise domain 1. Domains 2-6 are made up of 33 items derived from American Geriatric Society [AGS] Indicators of Persistent Pain [2002] to assess pain at the point of care.[27] Data for domains 2-6 are manually entered into the smart phone app by the assessor based on observations of the resident's appearance and activities, in the form of binary (yes/no) responses (see Box 1). There is no information on the weighting of the items.

Box 1: Domains in the PainChek® pain scale

Domain 1 (assessed by AI): Facial expression (brow lowering, cheek raising, tightening of eyelids, wrinkling of nose, raising of upper lip, pulling at corner lip, horizontal mouth stretch, parting lips, closing eyes)

Domain 2: Voice (pain sounds e.g. ouch, groaning, moaning, crying, screaming, talking)

Domain 3: Movement (altered or random arm or leg movement, restlessness, freezing, guarding/touching body parts, moving away, abnormal standing or walking)

Domain 4: Behavior (changes in interpersonal, mental status changes, aggression, confused, distressed, dislike of touch, fear)

Domain 5: Activity (resisting care, altered sleep cycle, prolonged resting)

Domain 6: Physical signs (fever, rapid breath, red face/flushed, painful injuries or painful medical conditions)

[25] Ekman, P., Friesen, W., & Hager, J. (1978). The Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, CA.: Consulting Psychologists Press.

[26] Kunz, M., Seuss, D., Hassan, T., Garbas, J.U., Siebers, M., Schmid, U., … Lautenbacher, S. (2017). Problems of video-based pain detection in patients with dementia: a road map to an interdisciplinary solution. *BMC Geriatrics*. *17*(1), 33.

[27] Atee et al., (2018a), *op cit*. note 13.

Based on the presence or absence of these 42 indicators, PainChek® produces a pain score that falls within one of four categories reflecting no pain (score 0-6), mild pain (score 7-11), moderate pain (score 12-15) and severe pain (score 16-42). The assessor uses the score to make decisions about administering pain relief, and by repeating the assessment, records the individual's response to any treatment.

Evidence-base

The evidence-base for PainChek® relies on a series of papers published by the team that developed the initial electronic Pain Assessment Tool (ePAT) and later PainChek® system consisting of the mobile application and web portal.[28] The research compares ePAT with a paper-based pain evaluation tool called the Abbey Pain Score (APS). The studies report that PainChek® scores correlate with APS scores indicating validity, and that there is good inter-rater reliability and high internal consistency (i.e. PainChek® items that are designed to measure the same construct generate similar scores, suggesting that it measures what it purports to measure). However, the research is at high risk of bias for several reasons. Raters were largely familiar with the tool and the residents; they were not blinded to residents' diagnoses or management; and the studies were small (maximum of 40 participants). There is no reported piloting of the tool with cognitively intact individuals; the APS was used as the comparator despite the ePAT allegedly being developed to overcome the limitations of the APS; and all but the first paper received funding from the companies created to commercialise the product (EPAT Technologies Ltd, ePat Pty Ltd, PainChek Ltd).

[28] Ibid.; Atee et al., (2017), *op cit*. note 13; Atee et al., (2018b), *op cit*. note 13; Atee et al., (2018c), *op cit*. note 13; Hoti et al., *op cit*. note 13.

There has been no independent evaluation of PainChek® and no published evidence of evaluations performed after the initial research in 2017.

Regulation, approvals and data handling

PainChek® received regulatory clearance in 2017 from the Australian Therapeutic Goods Administration (TGA) and the Conformité Européene (CE) mark as a 'Medical Device Included Class 1' (the lowest risk category, which also includes tongue depressors and surgical retractors). It was approved with the intended purpose of being "used to assess and monitor pain in people who cannot verbalise such as people with dementia or communication difficulties".[29]

The data collected by PainChek® are managed via a web administration portal hosted on Amazon Web Services using the Amazon Elastic Compute Cloud.[30] The PainChek® website states that operators of the app must have the consent of the person they are assessing, provides a link to the privacy policy and lists potential uses of the anonymised data collected by the app including trend analysis, app usage, clinical reviews, content for educational programmes and/or content for publication.[31]

Funding and other issues

In 2019, the Australian Federal government gave $5m in grant funding under the Dementia and Aged Care Services program to PainChek®. The funds comprise $500,000 for training materials and an evaluation report, $4.4m for 100,000 PainChek® licenses for use with people living with dementia across residential aged care in Australia, and $100,000 for a

[29] Therapeutic Goods Administration, *op cit*. note 14.

[30] Attee et al. (2018b), *op. cit*. note 13, p. 4.

[31] *Painchek consent information*. (2018). Retrieved from http://painchek.com/data-consent-policy/ [Accessed 21 July, 2020].

report at the end of the contract term.[32] In late 2019, the non-executive chairman and the CEO both sold large quantities of shares in PainChek Ltd (ASX:PCK). One analyst interpreted this to indicate a possible lack of confidence in the company,[33] although sellers might have a variety of reasons other than lack of confidence for selling stock. After the sell off, the share price dropped from a high of around 0.370, reaching a low of 0.061 in March 2020.

IDx-DR

Aims of IDx-DR

IDx-DR is an autonomous AI system for the automatic detection of early signs of eye disease in diabetic patients.[34] IDx-Dr aims to bring specialist level diagnostics to primary care, thereby increasing access to, and decreasing the cost of, diabetic eye care based on the premise that improved access will facilitate better eye care, and ultimately fewer cases of blindness. Diabetic retinopathy (DR) is a known complication of diabetes and a leading cause of blindness worldwide. For early detection and optimal management of diabetic retinopathy, patients require regular eye examinations. Those with no or mild DR can be followed with annual screening, while those with more than mild DR and/or diabetic macular edema (DME) require specialist evaluation for management to avoid damage to vision. Despite the risk of becoming visually impaired, fewer than 50% of diabetic patients receive the recommended screening. Screening usually requires a separate appointment, potentially

[32] *PainChek ASX Announcements*. (n.d.). Retrieved from https://painchek.com/asx-announcements/ [Accessed 21 July, 2020].

[33] Anon., *op cit*. note 16.

[34] Abràmoff et al., *op cit*. note 17.

distant from the primary care provider, and may involve dilatation of the pupils, which can be uncomfortable and affect vision for an hour or longer. IDx-DR offers a specialist-level diagnostic service in primary care. It is currently in use in multiple locations around the United States.[35]

How does it work?

IDx-DR analyses retinal images to provide a diagnosis that classifies patients according to the presence or absence of more than mild DR and/or DME (referred to as mtmDR). Patients who are negative for mtmDR are recommended annual screening. Those with more than mtmDR are recommended for referral for specialist eye care. The diagnosis is delivered within one minute.

IDx-DR is deployed in primary care. Operators with no previous relevant experience require a single four-hour training session to operate IDx-DR. The operator uses a non-mydriatic retinal camera (i.e. patients do not require dilatation of pupils) to take two images of each retina. The images are immediately evaluated by the first of two algorithms for image quality. Images of adequate quality are analysed by the second algorithm. In the testing protocol, the operator could make two further attempts at imaging followed by three attempts with dilatation to obtain an adequate image.

The second, diagnostic algorithm is described as "clinically inspired" as it is trained to detect the characteristic lesions of DR that a clinician would look for.[36] The training of algorithm two is described in some detail. It has independent validated detectors to identify relevant lesions including microaneurysms, hemorrhages and lipoprotein exudates. Detectors have been implemented as multilayer convolutional neural networks (CNN). CNNs are a type of

[35] IDx-DR, *op. cit*. note 12.

[36] Abràmoff et al., *op. cit.* note 17, p. 4.

deep learning algorithm that can differentiate and assign importance to various aspects of an image. Each of the CNNs was independently trained and validated to detect its assigned lesions from a region of a retinal image, using a total of over 1 million lesion patches from retinal images from people with and without DR. The AI fuses the outputs of the detectors into a disease-level diagnosis. As the algorithm was trained to identify the visual indicators that human clinicians rely on, the authors describe it as transparent and interpretable. It is a locked algorithm so its diagnostic processes cannot change. The source codes for IDx-DR are copyrighted by the parent company IDx LLC and are not publicly available.

Evidence-base

The foundational evidence-base for IDx-DR is an observational study of 900 participants published in 2018.[37] Each participant underwent two examinations. The first, by the IDx-DR involved 4 retinal images analysed by the algorithm. The second involved pupil dilatation, retinal photos, images of the anterior chamber of the eye, and optical coherence tomography performed by independent photographers certified by the Wisconsin Fundus Photography Reading Centre (FPRC). Three experienced FPRC-validated readers then graded the latter images according to a protocol. The readers were blinded to the AI diagnoses. The diagnostic outputs of IDx-DR were compared with outputs from the human readers, which were used as the gold standard or ground truth. The AI system had sensitivity of 87.2% (i.e. 87.2% of the images classified by the human gold standard process as *having* mtmDR received the same classification from the AI) and specificity of 90.7% (i.e. 90.7% of the images classified by the human gold standard process as *not having* mtmDR received the same classification from the AI). These figures indicate that the AI had acceptable accuracy in

[37] Ibid.

identifying both those who had mtmDR and those who did not. The imageability rate was 96.1% (i.e. 4% of patients were not successfully imaged).

An independent contract research organization managed this study, and an algorithm integrity provider locked the system and blocked access by the sponsor to all results until the trial ended. Two further studies (of the IDx-DR-EU-2.1) in the Netherlands with over 3000 patients have validated and replicated these results.[38]

Regulation, approvals and data handling

IDx-DR received FDA approval as an autonomous medical device in April 2018, the first to be so approved.[39] IDx-DR was granted Breakthrough Device designation and approved through the FDA's De Novo premarket review pathway for some low- to moderate-risk devices that are novel and for which there is no prior legally marketed device. The FDA provided intensive guidance to ensure the project met regulatory requirements regarding statistical design and study endpoints. In addition, IDx-DR has clearance as a Class IIa Medical Device (other Class IIa devices include dental drills and ultrasound machines) for sale in the European Union and received the CE mark from Underwriter's Laboratory in 2013.[40] There is no information on the IDx-DR website about data storage and handling.

[38] Van Der Heijden et al., *op cit*. note 17; Verbraak et al., *op cit*. note 17.

[39] US Food and Drug Administration. (2018). *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [Accessed 21 July, 2020].

[40] IDx-DR, *op. cit*. note 12.

Funding and other issues

All three studies to date have been funded by IDx LLC. Several authors, including Abràmoff, have close connections with IDx LLC such as holding shares, holding patents or being an employee.

Section 3: Ethical analysis of the two cases

Our aim was to inductively identify ethical issues arising in the context of these AIs. We present these issues as they arise sequentially in the AI system lifecycle for each use-case, together with the relevant underlying ethical concepts (see Table 2 for a summary).

Promissory claims

AI 'hype' refers to promissory claims made in professional and public facing materials. These raise issues of veracity, transparency and trustworthiness, depending upon the extent to which they are backed by evidence. The front page of the Painchek® website has the tagline "Intelligent Pain Assessment" and claims "PainChek® uses AI, facial recognition and smartphone technology to intelligently automate the pain assessment process at the point of care".[41] However, the AI element of PainChek® is modest (one domain out of six; nine out of 42 items on the pain scale), and there is no information about how the pain score is generated to support the claim that it is "intelligently automated". Presenting PainChek® as a novel AI tool fosters what has been called the technological imperative in healthcare, where technical innovations such as AI are seen as superior practice, with the implication that not using the novel intervention signals dated and perhaps inferior care.[42] IDx-DR

[41] Painchek®, *op. cit*. note 12.

[42] Burger-Lux, M.J. & Heaney, R.P. (1986) For better and worse: the technological imperative in health care. *Social Science and Medicine*. 22(12),1313e20.

promotes the verifiable claim that it was the first autonomous AI to receive FDA approval. This claim may also invoke the technological imperative but as it is supported by evidence, does not raise questions of potential deception and lack of transparency raised by the PainChek® hype.

Value of stated goals

The goals of AI use-cases speak directly to questions of beneficence. Healthcare interventions are justified to the extent that they promote the patient's overall interests and contribute to decreasing avoidable morbidity and mortality. Both PainChek® and IDx-DR attempt to address this point. The PainChek® materials do not include evidence that diagnostic failures cause under or overtreament of pain in the relevant population, although the implication is prima facie plausible. IDx-DR materials include evidence of potentially avoidable morbidity from inadequate diabetic eye examinations and follow up. The IDx-DR authors explicitly state that a major motivation for developing the AI is to increase access and reduce costs, thereby appealing to social justice issues of access and equity as well as beneficence.

Information about the algorithm

Information about the algorithm and its development is ethically relevant for reasons including its understandability or interpretability, and hence the extent to which physicians might trust it. IDx-DR provides evidence to support its claim that it is clinically inspired. For PainChek®, this issue is hard to assess as there is no information about the contribution of the AI elements to the overall pain score, which is however based on clinical indicators used in other pain assessment tools. Technical details about the development of algorithms (such as the training sets used) have justice implications as they indicate the potential for unfair bias. The pivotal IDx-DR paper contains a description of the development of the diagnostic algorithm and addresses issues of potential bias with evidence of attention to gender and

racial diversity. There is no information about the training set used in PainChek and no discussion of potential bias.

Evidence of efficacy

Evidence of efficacy is essential for ensuring that interventions are beneficent: healthcare interventions provide benefit to patients to the extent that they achieve the relevant goals. High quality research is necessary to underpin claims of efficacy. This was lacking in the case of PainChek®. Its trials had low participant numbers and were at high risk of bias due to lack of blinding and lack of independent operators of PainChek® during the research. The research underpinning IDx-DR seemed rigorous and provided proof of efficacy. Neither PainChek® nor IDx-DR included patient-reported outcome measures (PROMs), which are increasingly used to ensure that healthcare interventions achieve outcomes that are valued by patients.[43]

Potential for harm

Providing information about the potential harms of healthcare interventions is essential to meet the ethical requirement of non-maleficence. Patients screened by IDx-DR may be harmed compared to screening by an ophthalmologist if they have as yet undetected conditions that IDx-DR is not trained to identify, such as glaucoma or macular degeneration. However, they may be better off overall if the comparator is no eye exam. Patients whose pain is managed by use of PainChek® may be benefited if it accurately identifies a need for more pain relief, or harmed if their subsequent pain relief is less adequate than that provided by normal care, or leads to overtreatment. This comparative information is not provided.

[43] Black, N. (2013). Patient reported outcome measures could help transform healthcare. *BMJ*. 346, f167.

Regulation, legal liability and managing interests

Regulatory approvals of new interventions indicate a level of independent assessment and can therefore foster both practitioner and patient trust. Both AI devices have regulatory approvals, but the degree of oversight varies between them. PainChek® is approved as a class 1 device which has very little oversight regarding safety or efficacy. IDx-DR is approved as a low to moderate risk De Novo device which does require a specific evidence base. As patients may not be aware of the different classes of regulatory approval and associated guarantees of safety and efficacy, clarifying the evidence required for different types of approval would enhance transparency.

Providing information on legal liability for harm ensuing from use of the AI demonstrates responsibility and is required by justice. This information was not available for either AI on 30 March when we closed data collection. Since then, the company responsible for IDx-DR has indicated that it is legally liable for the accuracy of the diagnosis.[44]

Conflicts of interest undermine patient care when they lead to bias about the efficacy of and need for interventions, with potential detriment to patients' interests and wellbeing. In both these case studies, companies with a financial interest in the success of the AIs funded relevant research. With PainChek®, most of the authors are commercially involved with the product and were directly involved in the research. No measures to guard against the effects of the financial conflicts were reported. Attempts were made to manage and minimise conflicts of interest related to company funding of IDx-DR research through the use of a contract research organisation and Algorithm Integrity Provider.

[44] Abràmoff, M.D., Tobey, D., & Char, D.S. (2020). Lessons Learned About Autonomous AI: Finding a Safe, Efficacious, and Ethical Path Through the Development Process. *American Journal of Ophthalmology*. *214*, 134–142.

Management of information and data

AI applications in healthcare rely on patient data, such as images, test results or biometric data. Patient autonomy can be supported by requiring informed consent for data collection and management, and controlling access to collected and stored patient information in ways that protect patients' privacy. PainChek® uses facial images and pain scores which are highly personal, making any data breach a potentially serious breach of privacy. It is unclear how consent was obtained for participants in the PainChek® trials, given that many had cognitive impairments and may not have been able to provide consent themselves. The PainChek® authors claim that data are held securely but their data are hosted on the Amazon cloud which has been subject to security breaches.[45] The IDx-Dr materials indicate that an independent Algorithm Identity Provider managed the research data, implying data security, but no further details are provided. Patients gave informed consent for participation in the research. There is no information about where the data are stored or potential secondary uses.

Justice and care

The impact of these AI tools on the delivery of care raises questions about the relationships in healthcare and about social justice and equity issues. Neither of the cases provide information about the impact on delivery of care, such as the amount of time it takes to use PainChek® compared to the alternative paper-based tool, or the impact of providing IDx-DR in a primary care setting. In both cases, use of the AI changes the carer and patient relationship. With IDx-DR, the ophthalmologist is eliminated from the examination process if

[45] See e.g. Scroxton, A. (2020). Exposed AWS buckets again implicated in multiple data leaks. Retrieved from https://www.computerweekly.com/news/252476870/Exposed-AWS-buckets-again-implicated-in-multiple-data-leaks [Accessed 10 February 2021).

the screening is negative with the risk of missing incidental findings. In both the specialist and AI situation, a human takes the retinal images, but there is no information on patients' experiences of and views about being screened by IDx-DR. PainChek® does not eliminate a human from the care process. The impact on the relationship between carer and patient of the use of a smart phone compared to a paper-based tool is not reported and may be variable. Looking at a video of the resident's face may lead to the carer paying more attention to them, but on the other hand, the carer may be inclined to leave the decision to the app and administer pain relief based on the pain score generated by PainChek® rather than anything the resident says or does, reflecting automation bias.[46]

Healthcare interventions have social justice implications as their nature and delivery may exacerbate or mitigate health inequities. Although a stated goal of IDx-DR is to improve access to diabetic eye care and reduce costs, increased access and lower cost have not been demonstrated in practice.[47] There is no information about, or discussion of, the impact of PainChek® on equity of access to healthcare or reducing health inequalities. As mentioned above, the development and training of the algorithm may introduce, exacerbate or minimise racial and gender biases.

Costs and opportunities

Finally, ethical questions about cost and opportunity arise because new interventions tend to be more expensive than existing ones and should therefore be required to demonstrate

[46] Gretton, C. (2017). The dangers of AI in health care: risk homeostasis and automation bias. Retrieved from https://towardsdatascience.com/the-dangers-of-ai-in-health-care-risk-homeostasis-and-automation-bias-148477a9080f [Accessed 10 February 2021].

[47] Savoy, M. (2020). Diagnostic Tests: What Physicians Need to Know IDx-DR for Diabetic Retinopathy Screening. *American Family Physician*. 101(5), 307-308.

improved health outcomes and/or greater efficiencies compared with existing care. There is no information about the cost of using PainChek® compared to usual care or, if outcomes are improved, the incremental cost of these improvements. PainChek® is currently offering free twelve-month licenses, funded by the Australian Government, for all Australian aged care residential facilities. The cost of the license is not revealed on its website but it represents an additional cost to usual care unless use of the app significantly reduces carer time, or the cost is warranted by improved outcomes for residents regarding their pain management. There is no information about the cost of using IDx-DR compared to usual care, or the incremental cost of any improvements in health outcomes.

Table 2 around here

Section 4: Discussion

Attending to the context of both development and implementation of AI use-cases in healthcare has led us to identify a range of ethical issues, some of which have had little emphasis to date in general AI ethics frameworks. Here we focus on three of these: veracity and deception, public and patient involvement (PPI) and healthcare relationships. Veracity and deception arise in the context of AI hype. Honest presentation of AI in healthcare matters because both respect for autonomy and acting in patients' best interests require a commitment to honesty, which is a fundamental value in the practitioner-patient relationship. Healthcare algorithms however, are often developed in the context of competitive venture capitalism, the values of which differ from, and may be incompatible with, the values of healthcare. This observation suggests the need to critically evaluate new healthcare AI technologies in their social, legal and economic contexts as well as in the clinic. While veracity and deception relate to the broader concepts of transparency and trustworthiness, both of which appear in the AI ethics literature, the particular issue of hype has not previously been emphasised in AI ethics frameworks.

Any lack of involvement of patients and/or the public in developing AI for healthcare raises concerns.[48] Given the potential impact of AI-assisted healthcare on patients' experiences and outcomes, there is a strong ethical mandate for PPI in the commissioning, design, deployment and evaluation of healthcare AIs. The absence of PPI risks the development of AI to address problems that are amenable to AI solutions and/or likely to be profitable, rather than addressing issues that are important for patients. We found no evidence of PPI in the development and deployment of IDx-DR or PainChek®.

Clinical AI applications have consequences for healthcare relationships. Potential adverse effects include impaired communication, loss of trust, or conflicted decision making, especially if there are discrepancies between advice from the AI and from the relevant clinician. Potential benefits for patients include freeing clinicians from routine and largely administrative tasks to focus more on humanistic aspects of care including communication and healing.[49] Addressing this point requires attention to detail as the potential impact varies with the type of AI. There was no information about patient experiences of and responses to care provided using IDx-DR or PainChek®. Including patient reported outcomes measures (PROMs) in the research base for new healthcare AIs is critical to address this point.

Other issues that we identify have been flagged in existing AI-ethics frameworks. However, the process of identifying issues at each stage of development and deployment ensures that no issues are neglected. Additonally, this approach has practical utility as it can inform attribution of responsibility for attending to the issues where they arise in the AI life-cycle. In

[48] Ho, *op. cit*. note 10, p. 36; Future Advocacy, *op. cit.* note 10, p. 36.

[49] Topol, E. (2019). *Deep Medicine: how Artificial Intelligence can make Medicine Human again*. New York: Basic Books; Ho, *op. cit*. note 10, p. 37; NHS England, *op. cit*. note 10.

contrast, top-down principles such as transparency or trustworthiness have little clarity or effect unless they are explicitly linked to particular features of the AI. Our findings support the claim that there is wide scope for interpretation of ethical concepts, how and where they apply, and actions required to support ethical practice.[50]

Finally, our analysis points to the close connection between ethical evaluation for healthcare and technical reporting. Details about the training set, the way the algorithm is constructed, sensitivity and specificity, data storage and so forth are essential for making ethical evaluations. For example, while data privacy is a well-recognised concern in the use of AI interventions, the level of concern relates to the nature of the data. PainChek® collects and stores facial images which reveal who the person is, their dementia diagnosis, and that they are being assessed for pain management. It is not possible to anonymise data like facial images. Retinal images are also potentially identifying and revelatory of health information,[51] but without access to IT resources it is considerably harder to recognise an individual from their retinal scan than their facial image. This point suggests that ethicists must contribute to multidisciplinary evaluations of healthcare AIs, to ensure accuracy in interpreting the ethical implications of technical specifications.

Conclusion

Our detailed analysis of use-cases illustrates the value of fine-grained examination of specific AI applications in identifying and addressing relevant ethical issues. Further detailed use-

[50] Jobin et al., op cit. note 1: p. 391.

[51] Poplin, R., Varadarajan, A., Blumer, K., Liu, Y., McConnell, M., Corrado, G., … Webster, D. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*. *2*(3), 158–164.

cases are required to develop an inventory of ethical issues in practice, supplemented with empirical research to ascertain impacts of specific AIs on patients, publics, healthcare providers and other stakeholders. We note that addressing ethical issues arising from healthcare AIs requires engagement with the values of healthcare throughout the AI development process in order to meet patient and practitioner expectations. General ethical frameworks for AI may not adequately address healthcare-specific expectations such as beneficence.

Without further work to specify ethical standards, the force of ethical frameworks in domains such as healthcare is unclear. For example, to meet the injunction for healthcare AIs to be beneficent, the meeting of certain outputs or thresholds should be specified (e.g. in terms of reduced morbidity/mortality or increased equity). How these should be developed is an open question but points to the need for PPI involvement as well as studies of AI applications in healthcare to be reported in ways that facilitate specialist healthcare ethics review as well as technical analysis. Proposals for critical appraisal frameworks for healthcare AIs offer opportunities for, and could be strengthened by, including ethical benchmarks,[52] but as we have shown, ethical reflexivity and attention to particulars will also be necessary to make a full evaluation of any application.

There is a need for stronger communication between all stakeholders involved in developing and implementing AIs in healthcare, to gain a shared understanding of both technical limits

[52] Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J.P.A., & Shah, N.H. (2020). MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *Journal of the American Medical Informatics Association. 27(12),* 2011-2015.

and context-specific ethical obligations, and work towards solutions. While acting in the patients' best interests and avoiding harm are fundamental to healthcare, these principles may seem out of scope to developers who consequently disregard some relevant benefits and harms in developing their AIs. Inclusion of PROMs is a critical part of this process. Our paper provides a potential blueprint for further use-case analyses. It is a contribution towards developing robust ethical evaluation of healthcare AI that can be integrated with other appraisal tools. However, much work remains to support ethically robust AI-assisted healthcare.

Table 1. Principles for ethical AI summarised from Jobin et al.

Concept	**Scope of applicability**	**Ethical underpinnings**	**Actions required to uphold principle**
Transparency	Refers to: explainability or interpretability; nature and scope of communication; disclosures	To minimise or avoid harm, improve AI, foster trust, enable engagement/debate	Disclosure of information by those developing or deploying AI (such as AI uses, source code, data use, evidence base, limitations etc.); provision of non-technical explanations; public and stakeholder interactions.
Justice, fairness and equity	Refers to: fairness, minimizing/preventing bias and discrimination, supporting diversity/equity; also procedural issues for appeal, and fair access to AI/data and fair benefits	To ensure equity in access to AI, data and the benefits of AI; minimise harms; provide fair access to redress and remedy; support social justice	Technical solutions (standards, explicit normative coding); transparency; testing and audit; ensuring laws/regulations that are fit for purpose; systemic changes for greater inclusivity
Non-maleficence	Refers to: safety and security; avoiding foreseeable or unintentional harms including to social	To avoid or minimise harms (understood as discrimination, privacy violations, bodily harms); to	Technical solutions and governance including in-built data quality evaluations, in-built security, privacy by design, appropriate

	wellbeing, infrastructure, and psychological, emotional and economic aspects	prevent loss of trust, skills, to prevent other negative impacts	oversight processes and practices, independent audits
Responsibility and accountability	Refers to: acting with integrity, transparent and clear chains of responsibility and legal liability; focusing on how harms may arise; promoting diversity and openness	To minimise or avoid harm; to promote diversity; to attribute responsibility and support justice in seeking redress for harms	Identify relevant actors (AI developers, designers, institutions, industry); clarify degree to which responsibility and liability can be attributed to AI or whether human actors are always responsible
Privacy	Refers to upholding privacy and protecting right to privacy with regard to data protection and security	To protect right to privacy; to respect autonomy (freedom); to build trust	Technical solutions (e.g. privacy by design); more research; improved awareness and regulatory approaches

Table 2. Summary of ethical issues raised by PainChek® and IDx-DR

Issue	**PainChek®**	**IDx-DR**	**Ethical underpinnings**
AI 'hype'	AI contributes only a small portion of the pain score generated.	Accurate claim that this was the first autonomous AI to receive FDA approval.	Veracity/avoiding deception Avoiding unwarranted uptake of new technologies (beneficence/non-maleficence) Building trust
Goal of AI	To improve pain management for patient cohort. No evidence provided that improved diagnosis will lead to better pain management.	To decrease avoidable visual impairment from diabetic retinopathy, to bring specialty level diagnostics to primary care, thereby increasing access and lowering cost.	Beneficence/non-maleficence Equity Social justice
Information about algorithm	Role of the algorithm in generating pain score is not explained. The training set used for the algorithm is not described.	Two algorithms and their roles are described. Training set used for the algorithm is described. Diagnostic algorithm is based on well-established 'racially invariant' biomarkers.	Veracity/avoiding deception Justice/avoiding discrimination Building trust
Evidence-base	Claims of efficacy not well supported by the published research which is at high risk	Claims of efficacy supported by the pivotal trial, and validated by later studies.	Beneficence Veracity/avoiding deception

	of bias. No information about the representativeness of the study participants. No reporting of patient-oriented outcome measures. No data on effectiveness of PainChek® in improving pain management.	Study participants are representative of the US diabetes population. No reporting of patient-oriented outcome measures. No data on effectiveness in reducing visual loss, or of cost-saving or improved access.	Fostering public and patient involvement
Potential for harm	No reports of adverse effects from using PainChek®. No information on adequacy of pain relief using PainChek® compared to usual care.	Patients with concurrent eye disease may feel falsely reassured and/or have delayed diagnosis of incidental conditions that IDx-DR does not identify (noted by authors).	Non-maleficence Building trust
Regulatory approval	Approved in Australia as a Class 1 device.	Approved by the FDA as a low to moderate risk De Novo device, and by CE as a Class II device.	Beneficence/non-maleficence Fostering trust
Legal liability for harm	No information on legal liability for harm ensuing from use of PainChek®.	IDx LLC is legally liable for accuracy of diagnosis, with some caveats.	Attributing responsibility Justice
Conflicts of interest	Most of the authors are commercially invested in the	Some of the authors have financial ties to the relevant	Beneficence/non-maleficence

	product. No evidence of strategies to minimise effects of commercial interests.	company IDx LLC which funded the research. Third parties engaged to minimise effects of commercial interests.	Building trust Impact on relationships of care Veracity/avoiding deception
Data uses	Little information provided about the consent process. Participants have dementia so may be unable to give informed consent. Data protection measures are described as 'secure' but data are hosted on Amazon cloud with potential privacy and security implications.	Participants gave informed consent for the study. No information provided about consent to future data uses. An Algorithm Integrity Provider managed the research data. No information about post-study data storage or any secondary uses other than that data will be available upon 'reasonable request'.	Consent and autonomy Privacy Building trust Veracity/avoiding deception
Impact on delivery of care	Not described. Operators had to be familiar with the patients to use PainChek® but the effect on care relationships is not described.	Not described. Impact on relationships is unclear.	Social justice Equity Impact on relationships of care
Impact on social justice	No information about impact on reducing health inequities.	No information about impact on equity of access to diabetic eye care or reducing health	Social justice Equity

	No evidence of transferability of the app to other racial or ethnic groups.	inequalities. Algorithm is effective in all racial groups.	
Cost and opportunity cost	No information about the cost of using PainChek® compared to usual care, or the incremental cost of any improvements. Overall opportunity cost is not stated.	No information about the cost of using IDx-DR compared to usual care, or the incremental cost of any improvements. Overall opportunity cost is not stated.	Beneficence Resource allocation