DELTA

Human involvement in AI-driven digital pathology pathways

ethical and legal considerations

PHG FOUNDATION

making science work for health

Authors

Elizabeth Redrup Hill, Colin Mitchell, Tanya Brigden and Alison Hall

Disclaimer

The following report is intended to provide general information and understanding of the law. The report should not be considered legal advice, nor used as a substitute for seeking qualified legal advice.

URLs in this report were correct as of March 2023

Written and produced by PHG Foundation

2 Worts Causeway, Cambridge, CB1 8RN, UK +44 (0)1223 761900

Correspondence to:

intelligence@phgfoundation.org

How to reference this report:

Redrup Hill E, Mitchell C, Brigden T, Hall A. Human Involvement in AI-Driven Digital Pathology Pathways: Ethical and Legal Considerations. PHG Foundation. 2023.

This report can be downloaded from

www.phgfoundation.org

The PHG Foundation is a health policy think-tank and linked exempt charity of the University of Cambridge. We work to achieve better health through the responsible and evidence-based application of biomedical science. We are a registered company, no. 5823194.

©2023 PHG Foundation

Acknowledgements

We are extremely grateful to the valuable thoughts and insights of all stakeholders/participants involved in this research. We are particularly grateful to Professor Rebecca Fitzgerald, Dr Marcel Gehrung and Dr Maria O'Donovan for generously providing their time and expert input to our workshops and further research. We are also grateful to our other DELTA project collaborators for their feedback on this research as it progressed and to the funders for their support.

Funding

This work was supported by Cancer Research UK & Innovate UK [Grant number 41662]. This report has been drafted pursuant to Sub-Deliverable 3.7.3 of this Grant.

Acknowledgements

We are extremely grateful to the valuable thoughts and insights of all stakeholders/participants involved in this research. We are particularly grateful to Professor Rebecca Fitzgerald, Dr Marcel Gehrung and Dr Maria O'Donovan for generously providing their time and expert input to our workshops and further research. We are also grateful to our other DELTA project collaborators for their feedback on this research as it progressed and to the funders for their support.

A PHG Foundation workshop report supported by Cancer Research UK & Innovate UK as part of Project DELTA

Funded by

CANCER RESEARCH UK

UKRI Innovate UK

Contents

About Project DELTA

Interest is growing in how artificial intelligence (AI) or machine learning can be used safely and reliably at scale in routine health care.[1] AI and machine learning cover a wide breadth of activities and can simply be defined as a set of complex computational tools which allow data to be analysed to reveal patterns or associations that may not be obvious, even to the trained professional.[2] Used properly, AI promises to transform healthcare and medical research by supporting over-stretched health professionals in a variety of ways, and by unlocking patterns and associations that were previously unknown. This will consequently facilitate earlier, personalised, and more effective healthcare diagnosis, treatment and management.

One of the most promising areas for the development of AI is digital pathology, where computers are trained to recognise areas of potential concern and flag these for further investigation. In early October 2021, the PHG Foundation held a series of workshops that explored some of the ethical and legal factors that may impact the implementation of AI in digital pathology, drawing on the example of AI for detection of a condition called Barrett's oesophagus (BE) as part of Project DELTA. This report forms part of Project DELTA, funded by Innovate UK and Cancer Research UK which seeks to improve the diagnosis of oesophageal cancer. It is an output from Work Package 3 of the project which aims to develop and validate a deep learning framework to assess Cytosponge™ samples. By combining digital pathology, with a novel sampling device consisting of a swallowable and expanding sponge on a string (Cytosponge™) the Project aims to develop a scalable minimally invasive alternative to endoscopy.

Project Delta

AI could potentially add most value in pathways such as those for BE and oesophageal cancer where a risk factor is common in the population but there are barriers to early detection, inhibiting or preventing prompt management through a lack of access to accessible, accurate and affordable investigation.

BE is caused by the reflux of acid and bile from the stomach, often resulting in heartburn symptoms. In Western countries, around 10-20% of the adult population are affected by gastro-oesophageal reflux disease (GERD)[3] of whom between 1.8% - 7.5% may have Barrett's oesophagus.[4] BE may also occur in asymptomatic individuals.

The Cytosponge-TFF3 test is a minimally invasive test for BE where cells are collected via the Cytosponge™ device (a soluble capsule containing a compressed sponge on a string) which is swallowed and dissolved in the stomach of the patient, releasing the expanding sponge.[5]

Project DELTA[11] aimed to utilise novel methods (including AI) throughout the patient pathway to:

1. identify those who may be at increased risk through applying a novel risk-algorithm to electronic health records (Work Package One)
2. develop a novel, less invasive, sampling method (the Cytosponge™) to collect cells from the oesophagus (Work Package Two)
3. apply a novel stain (Trefoil factor 3 (TFF3)) to these cell samples to more accurately identify cellular changes that indicate disease (intestinal metaplasia) (Work Package Three)
4. develop AI tools to support, and in some cases, replace assessment and interpretation of these cellular samples by trained pathologists, to identify relevant cellular changes (e.g., from stratified squamous cells to columnar epithelium containing goblet cells) to guide future patient management (Work Package Three)

The sponge is then pulled back up via the attached string, collecting superficial epithelial cells from the top of the stomach, the oesophagus, and the oropharynx. This presents a sample of largely squamous, as well as gastric columnar epithelium and respiratory epithelium cells, and possibly intestinal metaplastic cells. Preservative solution is used to preserve the sample before processing.

During processing, the sample is embedded in paraffin and stained with hematoxylin and eosin (H&E), as well as immunohistochemically stained with TFF3 (trefoil factor 3). The H&E stains ease identification and quantification of cellular phenotypes for quality control.[6] However, TFF3 is the key diagnostic biomarker of BE because it enables the identification and quantification of goblet cells, which are indicative of intestinal metaplasia. Moreover, mucin-producing goblet cells are a key feature of BE and are over-expressed by TFF3 staining, making it a critical diagnostic biomarker for BE.[7]

AI triage-driven diagnosis of Barrett's oesophagus

The collaborators working on the surveillance and detection of Barrett's oesophagus in Project DELTA have developed a semi-automated triage system using deep learning (as reported in Nature Medicine).[8] This uses computational methods to sort patients into eight groups of varying priority, leaving pathologists to determine only the equivocal cases. This means that cases are either (A) fully automated (i.e., those of insufficient quality on routine staining (using hematoxylin and eosin) and/or those which yielded strongly positive or negative results using the novel diagnostic (TFF3) stain) or (B) semi-automated.

These partially automated cases (i.e., a minority (around 33.7 %) that have passed quality checks but were equivocal on TFF3 staining) require manual review from a pathologist.

Some advantages of this semi-automated approach include that it allows for sample stratification to be more interpretable and transparent than a fully automated approach whilst saving pathologist time. In the research cohort, which had a disease prevalence of 50%, this represented a 66% saving in pathologists' workload. In a representative real-world population, with a lower prevalence, the DELTA team estimate that this would equate to nearer a 57% reduction in workload, because of the more disparate distribution of Barrett's oesophagus patients within the triage classes.

The AI screening process is mapped out by Gehrung and colleagues in their recent paper.[9] They describe first mimicking the screening process of samples observed by experienced pathologists by replicating their decision-making scheme using standard convolutional neural networks (CNNs) architecture. They then used saliency mapping, which demonstrated strong agreement with manual landmarks placed by pathologists. Additionally, Grad-CAM (a machine learning technique used to identify parts of an input image that most impact the classification score) was used to provide further transparency for pathologists to reassure them that classifications are not based on spurious morphological characteristics caused by over-staining or stained spots without appropriate tissue content.

The developers have provided research that shows deep learning models achieve high performance for tile-level classifications, that saliency maps agree with pathologists' criteria for classification of tissue tiles, that a fully automated approach currently shows suboptimal performance compared to experienced pathologists, and that consequently a triage-driven approach is preferred to select patients for manual review. The AI and cell sampling processes are described in greater detail elsewhere.[10]

Research aim

Although automation has many potential advantages, it also raises ethical and legal challenges. These pervade many aspects of how AI tools are developed and implemented and how AI influences decisions about the nature and level of human oversight that might be needed. The workshops aimed to identify and explore these relevant ethical and legal considerations, how they may impact the development of AI systems, and identify safeguards or measures that may be taken to address potential negative impacts.

Background: Ethical considerations for AI in healthcare

Preliminary ethical considerations raised in the literature and in regulatory policy before the workshops in October 2021 included ensuring that the AI used is safe, reliable, and free from bias to mitigate the possibility of flawed findings and potential harm. Additionally, understanding how the tool has been developed such as the datasets that have been used, and the checks that have been made, were identified as crucial for trusting automated results. It was also identified as important that the datasets used in development are representative of the future clinical populations who will rely on these tests.

Another concern raised in the literature was that the tool might rely on spurious features to decide about a particular case which could lead to flawed outcomes. Using techniques such as saliency mapping which demonstrate potential areas of interest are often helpful for human assessors to understand the basis on which a decision has been made. Perhaps the most pressing ethical consideration raised was that users (both health professionals and patients) may feel differently about their result if it has been generated automatically, especially if this result has not been checked by a specialist. Moreover, how trust, confidence, transparency, and consent might differ if processes were fully automated or if they included a human as part of the decision-making process, were and remain important ethical questions for consideration.

Background: Legal considerations for AI in healthcare

Several different legal factors also potentially apply, ranging from the conditions and safeguards which apply to the data used to develop the AI tool, the obligations placed on developers for explanations to be provided, and the rights afforded to data subjects who have decisions made about them. For example, the EU and UK General Data Protection Regulations (GDPR)[12] mandate that additional safeguards are put in place when a legally significant decision is made that relies solely on automated methods.[13]

These include making provision for humans to be involved in the decision-making process, or for appealing the decision under Article 22 of the EU/UK GDPR. One of the aims of these workshops was to consider how these rules might apply in digital pathology and explore what their implications might be.

Many jurisdictions are considering bringing forward further legislation to create additional safeguards and protect the safety and fundamental rights of individuals in relation to AI processing. For example, the European Union is debating a Draft AI Regulation[14] which would establish an additional tier of regulation for 'high-risk' AI systems that pose significant risks to health and safety. These include a set of rules for human oversight under Article 14, which could have a bearing on the development of tools in digital pathology.

As with other areas of application, AI in digital pathology is being implemented within an ethical and legal landscape that is highly dynamic. Consequently, we formulated the following research question to guide our workshops:

What ethical and legal factors influence the nature and level of human involvement that is necessary or desirable in AI-driven systems for digital pathology and healthcare?

Within the workshops we aimed:

- to identify and explore reasons for automation choices along a spectrum, from fully automated to full human-in-the-loop
- to identify and evaluate ethical and legal factors which influence the level and nature of human involvement required in the implementation of AI in digital pathology
- to consider how these factors may impact design and implementation of AI in digital pathology choices
- to identify ways to minimise the challenges identified and maximise the potential benefits of AI in digital pathology

Scope and methods

Scope

Our research question explored the impacts on human stakeholders that arise in the context of digital pathology. Specifically, those that arise from using machine learning frameworks to automate pathologists' tasks in assessing oesophageal cancerous and pre-cancerous conditions. Our primary aim was to understand how the implementation of AI in pathology AI-driven automation in pathology pathways, might impact both patients and healthcare professionals. Therefore, whilst this report may touch on broader themes and debates that have been raised on the use and implementation of healthcare AI, we focused on ethical and legal considerations which influence choices about when, and in what form, humans should be kept in or taken out of the loop with Project DELTA providing an excellent exemplar for such analysis.

Methods

In October 2021, the PHG Foundation held a series of online workshops which explored the ethical, practical, and legal considerations arising from digital pathology. As described above these asked the question: *What ethical and legal factors influence the nature and level of human involvement that is necessary or desirable in AI-driven systems for digital pathology and healthcare?*

Due to COVID-19 restrictions our plans for a multidisciplinary, face-to-face workshop had to be revised. We chose to split the group by their profession and/or expertise into three subgroups which each adopted an online workshop format. Potential stakeholders were based on their expertise in a relevant area, based on their knowledge of Project DELTA, or as representatives of a particular stakeholder group (e.g., disease associations).[15] The stakeholders included, software developers (n = 4) and pathologists (n = 7) (Workshop One), professional body representatives and policy, legal and ethical experts (n = 11) (Workshop Two), and representatives from relevant patient groups or charities and frontline healthcare professionals (n = 9) (Workshop Three).[16] To check that our conclusions were not outlying or misrepresentative, we presented the findings from each of these three subgroups in a single plenary workshop with all participants the following week. This gave participants the time for reflection and further comment and the opportunity to inform our final conclusions.

We asked our workshop participants to consider differing levels of automation, from the automation of mundane tasks through to AI reaching clinically actionable decisions along the pathway mapped out below. Figure 1 also demonstrates some of the key legal and ethical considerations raised at relevant points along this pathway.

Workshop One consisted of the pathologists implementing the Cytosponge™ testing and Cyted software developers who are developing the deep learning framework to undertake Cytosponge sample analysis.

Workshop Two consisted of ethical, legal and policy experts familiar with the pertinent considerations of healthcare AI.

Workshop Three brought in wider considerations such as patient perspectives through discussions with frontline staff and patient representatives.

Introductory presentations were given by the lead pathologist and AI developer in Project DELTA to participants in Workshop One. These were recorded, and were played back to the ethical, legal and policy experts in Workshop Two and patient-facing stakeholders and representatives in Workshop Three, to provide a baseline for discussion and to kick start discussion.

We concluded our series of workshops with a plenary for participants to address any outstanding thoughts. These discussions were conducted under the Chatham House Rule, meaning that the information gathered from our workshops would be discussed in the Report but that participants would not be identified unless they provided their explicit consent.

Workshop one

We asked participants questions such as:

- how have choices about the level of automation at different stages along the pathway been made?
- what level of human involvement in AI-driven pathology is expected, desirable and acceptable?
- is there any part of the pathway where a human-in-the-loop is non-negotiable?

Workshop two

In Workshop two, we asked participants questions such as:

- what key ethical, legal and policy considerations influence human involvement in AI-driven digital pathology?
- what do key stakeholders need in order to accept and trust AI-driven automation aspects of digital pathology?
- how might the presence or absence of a human-in-the-loop alter the responsibility (and potential liability) of healthcare professionals, developers and patients?

Workshop three

Workshop three explored questions such as:

- what might patients' attitudes be towards the use of AI in digital pathology?
- what features and safeguards might be necessary for patients to be comfortable with the involvement of AI?
- how much information do patient-facing professionals and patients want (or need) about the nature and level of involvement of the AI?
- what are the main ethical/legal concerns around the use of AI in digital pathology and how might they affect patients?
- what level of human involvement is expected/desirable/acceptable?
- are there any instances or purposes where human involvement is non-negotiable?

Plenary

The plenary session compared key findings from each of the three workshops and discussed top level considerations for future policy development of the use of AI technologies in pathology and healthcare more generally. Attendees were also given the opportunity to raise further thoughts or concerns and provide feedback on workshop findings.

During and after each workshop, comments from stakeholders were collated and grouped into overarching themes. Those themes and comments were then compared to the comments from the stakeholders in the other workshops to understand where areas of agreement, divergence or points of additional nuance arose.

The overarching themes arising from the workshops and plenary session were:

- the risks and benefits of automation
- the impact on human specialists
- bias and inequity
- transparency and choice
- attributing liability and responsibility for error
- public engagement and trustworthy AI

These themes are illustrated using non-attributed quotes from the workshops to avoid individual identification. Where necessary we have paraphrased the original for clarity of meaning.

Results and key themes for discussion

There were many insights made in the workshop discussions. The transcript of the workshop was analysed and emergent themes identified by a primary researcher. These preliminary findings were independently checked and ratified by two additional researchers. Emergent themes were then ranked according to the extent to which they contributed to, or provided an overview of the topic, and also whether they illustrated interesting points for key stakeholders that were not represented elsewhere.

A. Risks and benefits of automating

> "If we have a health system that is over capacity because we are not using AI, that also won't build trust. Technology is not the issue, it's trust in the humans implementing and regulating it. There's lots of confidence and trust in the NHS, so if the NHS is using it, that should foster trust in and of itself. We need to look at trust as a system level, and not just AI. This also means HCPs' trust will be crucial to fostering trust in patients."
>
> - Workshop participant

> "I agree the statistics about the lack of human expertise in pathology are very sobering, the use of AI assisted screening seems to be necessary to plug this gap. To me, the issue seems to be not only whether this is an efficiency gain, but whether pathologists and patients are confident in the output of AI systems. My worry would be that [these systems] are being sold as more efficient, more objective and better, this consequently sets up a heap of expectations that could undermine the whole rollout of these systems."
>
> - Workshop participant

These comments articulate several of the key tensions at the heart of deciding what challenges and advantages AI offers healthcare. Issues such as the maintenance of public trust, responsibility for benchmarking and standard-setting, the impact on human practitioners and patients, and the costs of not automating in a resource-strapped national health service underpin much of the discussion on the benefits and risks of increasing automation.

The challenges raised under this overarching theme are evidently varied, ranging from the practical questions of implementation, the political questions of resource allocation, the legal questions of liability and the ethical challenges of responsibility.

Whilst it can therefore be viewed as an all-encompassing theme there were several distinct considerations raised. In particular, concerns of over-anthropomorphising AI were discussed from the legal perspective of delineating liability and the ethical

challenges of ensuring that necessary human qualities, such as empathetic communication of results remained, particularly where results suggest serious or life-threatening consequences.

Additionally, the participants debated whether AI or humans should benchmark clinical standards; if AI were to become more efficient than humans, the question of accepting human fallibility becomes just as relevant as concerns that at least some AI is held to too high a regulatory standard. On the other hand, concerns of how AI can encapsulate the complexity and nuance that human health professionals do when reaching a clinical decision (by incorporating a patient's history, values and beliefs, comorbidities, resources, prognosis, physiological differences) are equally relevant.

These are not theoretical challenges: similarities can be drawn with the debates on the inefficiencies of evidence-based medicine where digitalisation and the use of randomised clinical trials have been lauded as the gold standard for clinical decision-making, despite the fact that such decision-making falls short in the real world where patients are a far cry from clinically controllable and predictable beings.[17] The discussions therefore demonstrated that deciding the extent of automation in a clinical pathway is a fine-balancing act and is often highly context-dependent.

B. The impact on healthcare professionals and patients

> "Screening the whole slide is fundamental practice in pathology. Consequently, pathologists may find it difficult to stop doing that and trust the AI only flagged areas. To do so would require a mental shift that goes against key training guidance."
>
> - Workshop participant

> "It's important to consider how this will impact pathologists. Will they spend more time training and maintaining AI and consequently will it negatively impact patient contact? Machine learning changes the relationship because of the increased need for oversight."
>
> - Workshop participant

Increasing the level of automation in healthcare pathways will inevitably impact the humans involved, including patients and healthcare professionals. Whilst the literature on healthcare AI tends to focus on how AI may benefit or harm patients (directly or indirectly), a key and diverging point of discussion alternatively focussed on how AI might change or impact the role of healthcare professionals.

It was noted that limiting their review to the areas of slides flagged by AI systems would require pathologists to make a significant mental shift from the current manual approach in pathology, where the whole slide must be examined. Such a profound shift could involve significant cultural change and may be slow to implement. It is therefore an important interdisciplinary consideration for software developers to be aware of when undertaking post-implementation surveillance of how the system is working in the real world.

Considerations such as the time and the resource saving impact of AI were also raised. It was emphasised that time-saving as a benefit of AI needed to be carefully communicated to the public because it was considered more appropriate to understand that time is reallocated, not saved within resource-strapped systems like the NHS. Consequently, understanding interdisciplinary challenges and mindsets is important for interpreting the impact of healthcare AI in both pre- and post-implementation impact assessments.

C. Bias and inequity

> **"It's about known-unknowns. Those who have Barrett's oesophagus are likely to be white, male and middle-aged. You maybe therefore didn't realise it, but there will already be inherent bias developers and healthcare practitioners will need to be aware of when using such AI."**
>
> **- Workshop participant**

> **"It's important to be aware that the demographics of the training cohort can also influence public perception and trust. Even if the system is not demographically biased, does the lack of a demographic's presence impact that demographic's trust in that system?"**
>
> **- Workshop participant**

Varied points were raised in the discussions on potential bias and inequity. Discussions were not limited to ethnicity and gender; age was also considered a potential point of inequity if healthcare becomes increasingly automated. Within the discussions on bias, it was considered important to be equally aware of the limits of current knowledge i.e., what are the known-unknowns, meaning awareness of known, and potentially foreseeable but unquantifiable, sources of bias in the literature. Such awareness means we recognise if potential bias occurs and how to mitigate or remove it entirely. The participants considered whether more guidance was needed on how to better detect potential bias for those developing, and the healthcare practitioners using, such health technology.

Emphasis was also placed on how bias and inequity impacts public trust. It was considered important for developers and healthcare specialists to be aware of how inequitable representation in training cohorts could mean that certain demographics distrust that AI, regardless of whether bias exists or not in that context. Therefore, how discussions around bias and inequity are presented, such as how training data sets are selected and utilised, any inherent limitations arising from these choices, and how these limitations are to be accounted for, are key to fostering and maintaining public trust and acceptance.

However, it was also acknowledged that there could be a danger in requiring tools and tests incorporating AI to meet regulatory standards which are disproportionate to the potential harm that it might cause, compared with non-AI medical devices. For example, the healthcare profession does not usually discuss the technical details of how tests work, their limitations (whether scientific or demographic) or provide options on different testing methods with their patients. Questions about whether this should be done with AI raise the prospect of adopting an exceptional approach to such tools. It was agreed that this should be considered carefully and not assumed to be a justifiable approach.

D. Transparency and choice

> "Patients would want to know more about how it impacts them and their care options and less about how the test is done. Transparency matters most when errors occur. Those are moments where transparency enables patients to seek redress, second opinions and question those decisions. Transparency doesn't mean, "tell them everything." It's about ensuring understanding and not causing further confusion i.e., what do patients and healthcare professionals need to understand about this technology?"
>
> - Workshop participant

> "Proper details about the tests and what they do or do not do, should be passed through the system (as an appendix to the report, perhaps) so that doctors interpreting the results have the best information for patients about their health risks and about opportunities for their condition to be treated or prevented, e.g., by radio frequency ablation. The risk of progression to cancer from Barrett's oesophagus is a fairly specialist subject that not all GPs may know much about."
>
> - Workshop participant

> "I feel uneasy about the choice being given to patients to decide between human or AI review. If the healthcare practitioner feels that something is not right, the same process of looking for a second opinion still follows. I can't see that there really is a choice for patients to opt-in or out. You do it in a way that is in the best interests of the patient (beneficence)... currently, healthcare practitioners don't tell them about other diagnostic support tools or other tests such as immunohistochemistry."
>
> - Workshop participant

Such considerations should be understood in light of the EU/UK General Data Protection Regulation (GDPR) which stipulates that data subjects must be given the option for human review where a decision has been fully automated.[18] The feeling of our workshop participants was that such an approach may be unhelpful in healthcare where patients are given the choice to refuse treatment but rarely or never a choice about how diagnostic tests are conducted.

The discussion of transparency in our workshops centred on whether using AI within a patient pathway required a distinct approach. Questions were raised regarding whether patients should be told that AI technologies are being used for review of their test and furthermore, whether patients should consequently be given a choice on whether to "opt-out".

It was felt that overly detailed discussions could be unhelpful because it could lead to over-explanation and excessive concerns around AI tools. For example, it was pointed out that current testing techniques are not discussed in depth and choice on how a test is undertaken or processed is not only never presented to patients, but further, that it is unlikely to be of real interest to them. Therefore, a careful balance needs to be struck on what information is helpful and what is harmful to provide to patients and their practitioners, on the use of healthcare AI.

E. Attributing or addressing liability and responsibility for error

"We should be wary of removing responsibility from humans and placing it on AI. Currently, if there was an error on an existing system for image analysis, the pathologist wouldn't be able to escape liability for error. This AI must be seen as a diagnostic support tool."

- Workshop participant

"Who will be responsible for errors and what will healthcare practitioners be responsible for? Will insurance companies insure … where results are partly reported by a machine?"

- Workshop participant

"What will happen if AI starts to pick up on human errors? How would this be tackled from the perspective of legal liability or moral responsibility? However, we need to be clear on what is meant by error. Error is a loaded word. "Discrepancy" is well described; it is known that pathologists' views are very subjective and agreement among them is commonly low … consequently, AI may have a role in homogenising standards. However, at what point does discrepancy have the potential to harm patients and become an 'error'?"

- Workshop participant

"Further clarity is needed on when the pathology report constitutes a 'decision' for the purposes of Article 22 UK GDPR. When is the decision being made and by who in this process? If the AI report does constitute a decision, is it practical to give patients a choice here? Alternatively, is it fair that their data is being processed by a system that they don't trust? Is it vital that a patient is given a choice in relation to how their results are processed and decisions reached?"

-Workshop participant

Within this report we distinguish between legal liability and moral responsibility. For example, legal liability for the purpose of this report should be understood as a form of responsibility that carries legal consequences, and that moral responsibility alternatively carries ethical consequences. There is nothing to suggest that in any given context one or more stakeholders could not simultaneously bear both forms of responsibility. However, where there is no explicit law establishing criminal or civil consequences for a breach, there may still be ethical or professional ethics consequences.

Queries regarding how liability and responsibility would be delineated among stakeholders in increasingly automated pathways were raised by all stakeholder groups.

The findings suggested that regulatory approaches lagged behind technological advances and that regulations are necessary but not sufficient for safe and effective implementation of healthcare AI. For example, it is currently under-explored how insurance companies might insure these professionals for relying on AI produced reports to treat patients.

Additionally, Article 22 EU/UK GDPR's approach to providing choice where decisions are fully automated is arguably out of step with how the NHS and medical regulation and law work, where decisions on testing and the appropriate standards for arriving at medical decisions are not regarded as choices for patients to make.

Of further concern to some stakeholders was the interpretation of 'error' and 'discrepancy', highlighting a possible interdisciplinary challenge. Discrepancy is commonplace within medical opinion, but concerns were raised about the point at which a discrepancy becomes an error giving rise to legal liability.

Whilst such questions are currently addressed in common law systems through tort law which sets out when a duty of care arises, there is currently little advice on what should happen when part or all of that medical decision was reached by AI.

These findings suggest that further regulatory clarity and/or professional guidance is needed to establish what should happen where AI is being used in the context of healthcare and diagnosis.

F. Public engagement and trustworthy AI

> "Managing public expectations will be key. Literacy initiatives are needed to ensure the public understand that AI is also fallible and does not produce high truth answers. There is a risk that a lack of public understanding of AI will result in a lack of confidence. Media depictions can spread distrust and fear of AI. On the other hand, the public currently hold clinicians to too high a standard and do not recognise that they are also fallible. Why should AI be held to a different account?"
>
> - Workshop participant

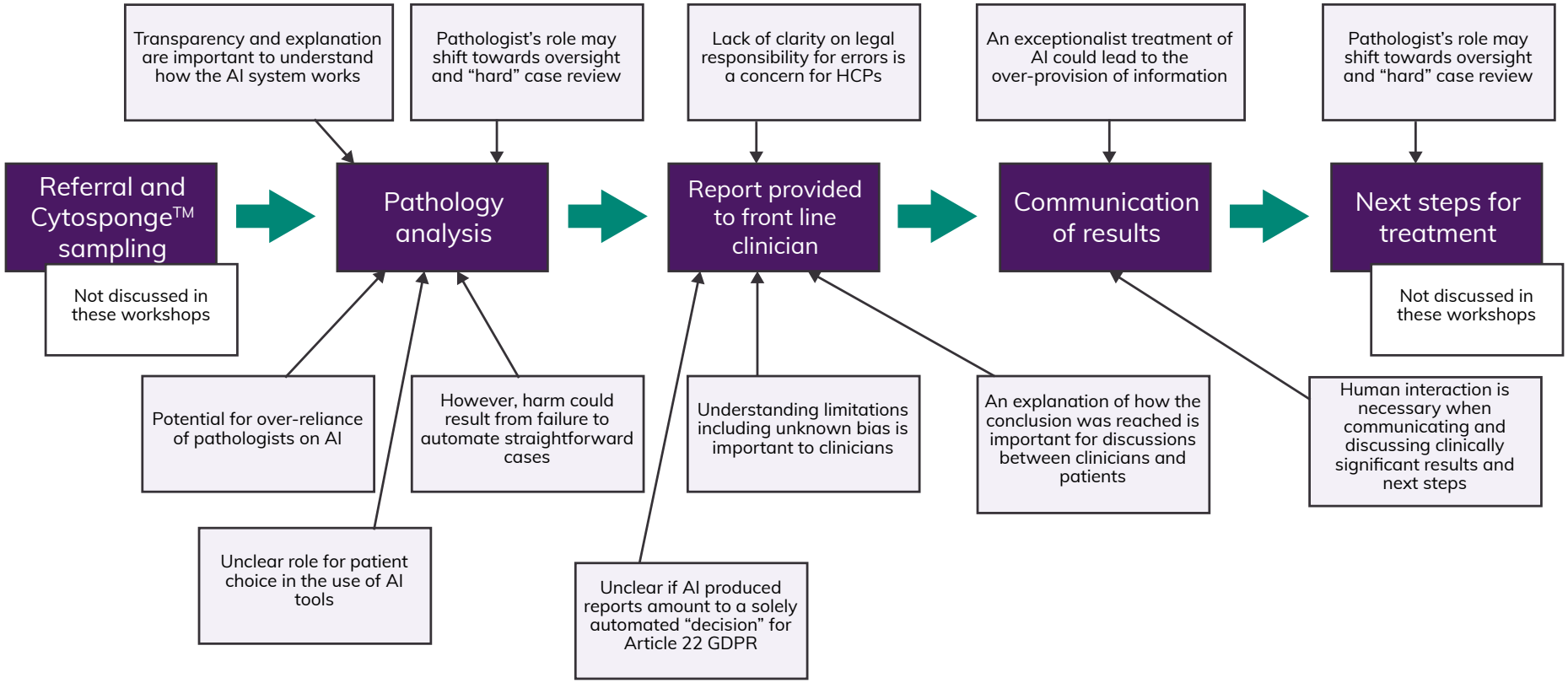
> "Key opinion leaders are needed to foster trust. Those that are patient-facing must be convinced that AI is safe and effective and that it is as good as, or perhaps better, than current methods. Public discourse should in practice fall into line with that but the public needs to be involved."
>
> - Workshop participant

> "It is also important to understand at the beginning of the design process what goal we are trying to achieve for this system. What values would the patient prioritise here? What values are driving the production of the system? Such values and priorities are important to define at the outset for clear public communication of what the benefits and risks are of new healthcare innovations."
>
> - Workshop participant

Public engagement and trust were also key themes for discussion. It was considered important at the design stage of any healthcare AI innovation, that funders and developers should be clear about their objectives for implementing innovative technologies and the values and priorities that will be adopted to achieve these objectives. Such discussions are important for subsequent public engagement on the impact it will have on humans in that healthcare pathway.

Public literacy initiatives were considered important to ensure realistic understanding of potential harms and benefits from AI, and to address inaccurate media coverage. Likewise, having patient-facing experts to champion novel AI technologies could increase uptake in the system through strengthening the chain of trust within the NHS service, flowing from regulatory boards and tertiary care experts all the way through to patients and publics.

Figure 1: A diagram of the discussed pathway and some key ethical and legal considerations[19]



Discussion

The key overarching themes of each workshop

Workshop one : Developers and pathologists

The key considerations raised in workshop one were:

- the need for appropriate regulation
- the ethical and **practical** considerations of increasing automation
- risks of not increasing automation
- impact on HCPs and patients
- bias and inequality
- liability and responsibility arising from error
- trust (what further knowledge is needed to espouse trust in this AI)

Of these, the most commonly discussed themes were:

- the ethical and **practical** considerations of increasing automation
- liability and responsibility arising from error
- impact on HCPs and patients
- trust (what further knowledge is needed to espouse in this AI)

Workshop two: Lawyers, policy experts and ethicists

As previous workshop, but with slightly different emphasis. *Key nuances **added** were:*

- for increased trust, **public engagement** initiatives are vital
- the considerations of **harm** that results **from failing automate** is also a valid consideration
- trust (what further knowledge is needed to espouse trust in this AI)

Workshop three: Patient-facing HCPs & representatives

As previous workshop, but with slightly different emphasis. *Key nuances **added** were:*

- trust and confidence is linked to the **provision of knowledge** and the possibility of **choice** e.g., what do HCPs and/or patients need to know? Should patients be given a choice on how such decisions are reached?
- liability and responsibility for error were discussed through the lens of **information and choice** and how that might be exacerbated if the balance is off

Plenary

All stakeholder groups

The plenary discussed all of the overarching considerations raised in the workshops. However, some were more commonly raised due to their felt importance for answering the question of whether automation should be increased or not. These were

Bias and inequality

Appropriate regulation

Trust and public engagement

Whilst similar themes were raised in each of the stakeholder groups, each placed differing emphasis and nuance on the topics raised. The table above outlines the overarching themes raised within each workshop for further analysis below. The three groups were largely in agreement about the overarching themes that were important, and these topics were replicated in the plenary discussion. These included the need for appropriate regulation; the practical and ethical considerations of increased automation; the harm that could result from both not automating and from automating pathways; the sociological impact on humans; bias and inequality; liability and responsibility for error, and public engagement.

The first overarching theme where there were differences in emphasis and nuance was trust. Developers and pathologists considered trust from the perspective of practical implementation. They considered the challenge of how to foster professional uptake, and consequently, trust in healthcare AI. They therefore considered what information might be required by pathologists to trust in the use of this technology and methods of redress if error occurs. However, in Workshop Two, the lawyers, policy experts and ethicists looked at trust from the perspective of public uptake and consequently, expressed the importance of public engagement. They also tied the idea of trust to information provision but from the perspective of increased public literacy to help manage expectations of AI (both its advantages and shortcomings). Providing information to foster trust was also discussed from the perspective of addressing gaps in medical guidance which do not currently specifically address challenges arising from AI-human hybrid pathways. For Workshop Three's patient-facing healthcare professionals and patient representatives, trust and information provision were considered from the perspective of choice: How much information is helpful? What would patients want to know? And should patients have a choice in how their results are processed and medical decisions reached? These nuanced views demonstrate that trust is a complex issue and getting information right for the aim of fostering trust is a delicate balancing act. Moreover, it suggests that without engagement from all key stakeholders any new initiative will likely miss the mark.

A further overarching theme where difference and nuance were prevalent was harm. Pathologists and developers considered harm from the perspective of legal liability and moral responsibility for erroneously interpreted and/or missed cases. However, lawyers, policy experts and ethicists considered harm from the perspective of not just harm arising from further automation but also from failing to automate, such as the ethical consequences of failing to do something that (if implemented appropriately) would be in the interest of public and patient welfare. Patient representatives and patient-facing healthcare professionals alternatively considered harm from the perspective of

redress. They considered how patients would feel if they subsequently found out AI had processed their results or made the clinically actionable decision that had resulted in harm. Concern was expressed that a lack of redress may exacerbate any harm suffered by patients because failing to fully inform them would rob patients of choice.

The impact on humans in the loop was also considered from slightly differing viewpoints. Pathologists and developers in Workshop One envisaged a future role for pathologists as specialists overseeing reports and decisions with less time examining every slide in detail. They considered that retraining might be needed to increase uptake and accuracy as only examining flagged parts of a slide would contradict existing pathology practice and would necessitate changes to training and practice. However, the lawyers, policy experts and ethicists (Workshop Two) considered potential social harms which could result from anthropomorphising such technology and which could complicate how liability and responsibility should be shared by key stakeholders. They also expressed concern that the purported advantage of timesaving through automation may not always lower the burden on health professionals, as such time savings are likely to be reallocated elsewhere. The patient-facing healthcare practitioners and patient representatives in Workshop Three discussed the challenge of explainability and redress if humans are increasingly removed from the loop. In tandem they discussed the complexity of medical decision-making and the need for human communication for serious disease and illness, suggesting it would never be advisable to entirely remove healthcare practitioners from such pathways.

The most surprising areas of nuance and divergence were in relation to harms that could result from failing to automate and the discussion on the prevalence of bias and inequality in pathology. The lawyers, policy experts and ethicists raised an interesting point that diverged from the common line of thinking in the literature that predominantly focuses on harms from automating and rarely considers the harm that would result if there were a failure to at least automate the straightforward tasks in our healthcare system. Additionally, the consideration of unknown bias or inequality that could be represented at a cellular level caused some participants to express more caution in assuming that an individual could not be identifiable at a cellular level, or at least that key demographic information could not be revealed. This raised the question of whether it is correct to assume that the information held in digital pathology slides do not amount to personal information or have the potential to disclose information that could result in biased treatment just because the personal identifiers have been removed or anonymised from the report the pathologist sees.

Conclusions

Many of the overarching considerations raised in these four workshops could be categorised into challenges of harm or trust. Both require significant consideration when deciding whether to automate, or to what extent to automate, an existing healthcare pathway. For example, we have found that issues of trust can be broken down into sub-categorisations such as professional confidence, and public and patient confidence, which can be split into further considerations depending on demographics and other factors. Likewise, harm cannot only be considered from the perspective of harm that occurs due to automation: the harm that might occur if we fail to automate is also a relevant policy driver which should not be ignored.

Further professional guidance is needed to bridge the gap for challenges that specifically arise in AI-human hybrid pathways, such as guidance on appropriate information provision for health care professionals, patients and publics; grappling with unknown bias and the extent to which patients should have a choice or not. Moreover, legal and regulatory clarity will be needed as Article 22 EU/UK GDPR's potential requirement for data subjects (in this case patients) to be offered a choice on whether an automated tool should be used or not sits ill-at-ease with medical practice and common law rulings on appropriate decision-making and discharging healthcare professionals' duty of care. Discussion of Article 22 and patient choice also highlights possible exceptionalist treatment of this technology, where other support tools not utilising machine learning but conferring equivalent harms and benefits are not regulated so closely. Getting the balance right will depend on the level of automation and extent to which clinically actionable decisions are reliant on AI processing.

These workshops demonstrated the importance of multistakeholder and multidisciplinary discussions. For example, the terminology of discrepancy and error might be used synonymously outside of pathology, but they suggest differing forms or degrees of accountability within the profession itself. Such discussions are therefore key for clarifying appropriate terminology which will be needed to effectively regulate and undertake surveillance on such pathways.

The extent to which a pathway should be automated or not is also highly context specific. It will depend on available resources, the likelihood of finding abnormal results and degree of complexity involved, and how informed the public and healthcare professionals are on the advantages and limitations of AI within that context. Managing expectations will be vital for increasing uptake by both health professionals and patients and also for effective post-implementation surveillance.

Policy considerations

The findings of this workshop provide evidence that more work is needed to understand the perspectives of key stakeholder groups on the implementation of AI. Such discussions are highly valuable and will help to mitigate future implementation challenges as these technologies are rolled out into practice. Our findings suggest the following considerations for policy makers, regulators, developers, and healthcare bodies:

1. healthcare providers and NHS England should ensure that multistakeholder perspectives, particularly the views of patients, have been gathered and addressed for the development of AI-human hybrid pathways. Considerations should include how to grapple with unknown bias and inequity in practice; increasing awareness of the limitations of AI and explainability of AI-human medical decisions (including lessons to be learnt from previous implementation strategies); as well as guidance on what amounts to appropriate information provision in such pathways and when, if at all, patient choice is appropriate.
2. regulators will need to consider how liability is to work in AI-human hybrid pathways to foster trustworthiness from health professionals and patients. This includes addressing the potential clash in approach between the current Article 22 EU/UK GDPR and common law rulings on how to discharge a duty of care as automation in health increases. In order to improve oversight of these technologies, medical device regulators will need to develop guidance that reflects interdisciplinary discussions so that requirements for post-implementation surveillance can be properly interpreted and implemented.
3. AI developers should consider the mitigations that could be adopted to cover adverse events, for example, ensuring that AI decisions are supported by measures that increase transparency and potentially the need for appropriate insurance and/or compensation where necessary.

References and notes

1. We use artificial intelligence (AI) and machine learning interchangeably. We recognise that machine learning can be represented on a scale from input-output algorithms all the way to independently learning algorithms that adapt and improve as they learn, and possible cognisance. We have intentionally left this phrase ambiguous as we wanted to capture stakeholder thoughts on algorithmic capability across this wide spectrum.
2. For more on this, see our previous report on black box algorithms and transparency challenges: Ordish J, Mitchell C, Murfet H, Brigden T, Hall A. Black box medicine and transparency. PHG Foundation. 2020. Available from: https://www.phgfoundation.org/report/black-box-medicine-and-transparency
3. El-Serag H B, Sweet S, Winchester C C, Dent J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. BMJ Gut. 2014;63:871-880. Doi: http://dx.doi.org/10.1136/gutjnl-2012-304269
4. Fan X, Snyder N. Prevalence of Barrett's Esophagus in Patients with or without GERD Symptoms: Role of Race, Age, and Gender. Digestive Diseases and Sciences. 2009; 54:572-577. doi: https://doi.org/10.1007/s10620-008-0395-7; Rex D K, Cummings O W, Shaw M et al. Screening for Barrett's esophagus in colonoscopy patients with and without heartburn. Gastroenterology. 2003;125(6):1670-1677. doi: 10.1053/j.gastro.2003.09.030; Herrera Elizondo J L, Monreal Robles R, Garcia Compean D et al. Prevalence of Barrett's esophagus: An observational study from a gastroenterology clinic. Revista de gastroenterologia de Mexico. 2017;82(4): 296-300. doi: 10.1016/j.rgmx.2017.01.006.
5. Gehrung M, Crispin-Ortuzar M, Berman A G, O'Donovan M, Fitzgerald R C, Markowetz F. Triage-driven diagnosis of Barrett's oesophagus for early detection of esophageal adenocarcinoma using deep learning. Nature Medicine. 2021;27: 833-841. Doi:10.1038/s41591-021-01287-9. See also Paterson A L, Gehrung M, Fitzgerald R C, O'Donovan M. Role of TFF3 as an adjunct in the diagnosis of Barrett's Esophagus using a minimally invasive esophageal sampling device- The Cytosponge TM. 2020;48(3): 253-264. doi: 10.1002/dc.24354.
6. ibid.
7. ibid.
8. Gehrung M, Crispin-Ortuzar M, Berman A G et al. Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. Nature Medicine. 2021;27: 833-841. doi: doi.org/10.1038/s41591-021-01287-9.
9. ibid.
10. ibid.
11. Project DELTA. Available at: https://www.deltaproject.org/
12. General Data Protection Regulation (EU) 2016/679 (GDPR). Available from: https://gdpr-info.eu/ [Accessed 13 March 2023]. Note the UK GDPR is not contained in a single instrument. The UK GDPR is the retained EU law version of the General Data Protection Regulation ((EU) 2016/679) (EU GDPR) as it forms part of the law of England and Wales, Scotland and Northern Ireland by virtue of section 3 of the European Union (Withdrawal) Act 2018.
13. Information Commissioner's Office. Guide to Data Protection: Rights related to automated decision-making including profiling. (n.d). Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/
14. EU Draft AI Regulation. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. This Regulation has now had multiple amendments, in December 2022 the Council of Europe approved a compromise version available from: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf and the EU Parliament is expected to pass it by April 2023, for adoption later in 2023. For more information on the Regulation, see: European Commission. Regulatory framework proposal on artificial intelligence. European Commission. n.d. Available from: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

15. Redrup Hill E, Mitchell C, Hall A, Brigden T. Ethical and legal considerations influencing human involvement in the implementation of AI in a clinical pathway: A multi stakeholder perspective. Frontiers in Digital Health. 2023;5:1139210. doi: 10.3389/fdgth.2023.1139210.

16. ibid.

17. Greenhalgh T, Howick J, Maskrey N. Evidence Based Medicine: A Movement in Crisis? British Medical Journal. 2014;348:g3725; Wenger D R. Limitations of Evidence-Based Medicine: The Role of Experience and Expert Opinion. Journal of Paediatric Orthopaedics. 2012;32(2):S187; Charon R. Where does narrative medicine come from? Drives, diseases, attention and the body. In Rudnytsky P L, Charon R (eds). Psychoanalysis and narrative Medicine. State University of New York Press. 2008; Pope C. Resisting Evidence: The Study of Evidence-Based Medicine as a Contemporary Social Movement. Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine. 2003;7(3):267. Hampton J R. Evidence-Based Medicine, Opinion-Based Medicine, and Real-World Medicine. Perspectives in Biology and Medicine. 2002;45(4):549; Sackett D L and others. Evidence Based Medicine: What it is and What it isn't. British Medical Journal. 1996;312:71.

18. General Data Protection Regulation (EU) 2016/679 (GDPR), Article 22 (Automated individual decision-making including profiling). Available from: https://gdpr-info.eu/art-22-gdpr/ [Accessed 13 March 2023]. Note the UK GDPR is not contained in a single instrument. The UK GDPR is the retained EU law version of the General Data Protection Regulation ((EU) 2016/679) (EU GDPR) as it forms part of the law of England and Wales, Scotland and Northern Ireland by virtue of section 3 of the European Union (Withdrawal) Act 2018.

19. This figure has been reproduced from Redrup Hill E et al at endnote 15.

The PHG Foundation is a non-profit think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and citizens.

intelligence@phgfoundation.org

UNIVERSITY OF CAMBRIDGE

PHG FOUNDATION

making science work for health