Independent report
Equity in medical devices: independent review – summary report

Foreword

The initial stimulus for this review was growing concern about a specific medical device – the pulse oximeter, which estimates the level of oxygen in the blood – in common use throughout the NHS. The COVID-19 pandemic highlighted that the pulse oximeter may not be as accurate for patients with darker skin tones as for those with lighter skin. An inaccurate reading could lead to harm if there was a delay in identifying dangerously low oxygen levels in patients with darker skin tones, which normally would have triggered referral for more intensive care.

When the Rt Hon Sajid Javid MP, the then-Secretary of State for Health and Social Care, asked me to carry out this independent review, the scope was extended to recognise the potential for bias in other medical devices, not just pulse oximeters, and beyond racial and ethnic bias to further unfair biases in performance, including by sex and socio-economic status.

My first step was to invite 4 wise professionals with a special commitment to equity in healthcare to join me: Professors Raghib Ali, Enitan Carrol, Chris Holmes and Frank Kee. Together, we formed the review panel and made the many decisions about the conduct of the review. This report reflects the conclusions of the whole panel. We were ably supported throughout by a dedicated secretariat of Dr Aleksandra Herbec, Maya Grimes and Jessica Scott.

To understand the latest evidence, we liaised with academics and reviewed relevant research, commissioning a series of focused literature reviews where necessary. To learn from experience in developing or using medical devices, we engaged with a wide range of stakeholders, holding individual and group sessions with:

- patient and public representatives
- national leaders from NHS agencies and health professions
- medical device regulators, developers and manufacturers
- independent health policy foundations

We also held a [public call for evidence](#). Finally, we tested our understanding in roundtables and follow-up events. We are immensely grateful to all who offered their advice and wisdom so generously.

In our review, we considered the evidence for differential performance of medical devices by socio-demographic groups that had the potential to lead to poorer healthcare for the population group disadvantaged by the bias. Crucially, we looked for evidence of the causes of the bias to inform our subsequent recommendations.

1. Optical devices

First, we focused on what we termed 'optical devices', where the initial differential performance stems from the physics of the hardware itself. These optical devices – pulse oximeters among them – send light waves of various frequencies through the patient's skin to make measurements of underlying physiology, but the light reacts differently with varying levels of melanin in the skin.

This differential performance by skin tone would not necessarily be a problem for healthcare if the variation in performance had been recognised and appropriate adjustments made to calibrate the devices according to skin tone. However, no such recognition appears to have happened either in the case of pulse oximeters or in an unknown number of other optical devices, which are not adjusted for differential performance by skin tone. The problem has been compounded in pulse oximeters by the practice of testing the devices on participants with light skin tones, so that these readings are taken as the norm.

There is some evidence – so far only from the US healthcare system – of adverse clinical impact of this racial bias in pulse oximeters on the healthcare received by Black patients compared with White patients. Our recommendations for optical devices, therefore, start with mitigating actions in relation to the pulse oximeters already in widespread use across the NHS and in homes all around the country.

We appreciate that pulse oximeters are valuable clinical tools, so we would definitely not advise curtailment of their use in patients with darker skin tones. Rather, we recommend immediate modifications to existing practice, including guidance for patients, health professionals, manufacturers and other relevant agencies. We commend the current intensive efforts by the Medicines and Healthcare products Regulatory Agency (MHRA) in the UK, the Food and Drug Administration (FDA) in the US and the EU to tighten the regulations and guidance on pulse oximeters in response to the evidence of inequities in health outcomes.

For optical devices overall, our recommendations aim at preventing adverse impacts arising in new devices by adding an equity lens to the whole device lifecycle.

2. Artificial intelligence

Second, we turned our attention to medical devices enabled by artificial intelligence (AI). The advance of AI brings with it not only great potential benefits to society but also possible harm through inherent bias against certain groups in the population, notably women, ethnic minority and disadvantaged socio-economic groups.

Few outside the health system may appreciate the extent to which AI has become incorporated into every aspect of healthcare – from prevention and screening to

diagnostics and clinical decision-making, such as when to increase intensity of care. Our review reveals how existing biases and discrimination in society can unwittingly be incorporated at every stage of the lifecycle of the devices, and then magnified in algorithm development and machine learning.

The evidence for adverse clinical impacts of these biases is currently patchy, though indicative. Seven of our recommendations are therefore focused on actions to enable the development of bias-free AI devices, with the voices of the public and patients incorporated throughout.

We were impressed with the initiatives on equity and AI that are underway by MHRA and international collaborations. Our recommendations are intended to strengthen and reinforce these ongoing efforts. In the final recommendation on AI, however, action at the highest levels is urgently needed to anticipate potential harm. We call for a government-appointed taskforce on large language models (LLM) – exemplified by ChatGPT – to assess the health equity impact of these potentially alarming digital technologies, together with the proper resourcing of the regulators to take on the challenges of assessment.

3. Polygenic risk scores

Third, we reviewed an emerging use of genomics – polygenic risk scores (PRS) – to consider what would be needed to future-proof their development in ways that would address equity concerns as they evolved.

The data sources upon which PRS draw have a well-established bias against groups with non-European genetic ancestry, but, in addition, we were concerned by the potential for misinterpretation of results by the public and health professionals alike, especially in relation to genetic determinism, which may carry wider societal risks. We were impressed by the intensive efforts already underway nationally to tackle the genetic ancestry bias in major datasets. Our 3 recommendations, therefore, concentrate on action on the broader societal front.

4. Future work

Last, we identify emerging issues looming on the horizon that need urgent attention now. Our final call to action for future work is a review to be carried out of medical devices encountered during pregnancy and the neonatal period, as part of the wider investigations of health outcomes for ethnic minority and poorer women and their babies.

The panel and I believe wholeheartedly that these recommendations need to be implemented as a matter of priority with full government support.

The government is already signalling the need for urgent action on AI regulation by calling the Global Summit on AI Safety in November 2023. But nowhere is the need to ensure AI safety and equity more pressing than in medical care, where built-in biases in applications have the potential to harm already disadvantaged patients. Now is the time to seize the opportunity to incorporate action on equity in medical devices into the overarching global strategies on AI safety.

Professor Dame Margaret Whitehead
Chair, independent review on equity in medical devices

Panel reflections on the independent review recommendations

Over the 15-month period that the independent review panel convened, reviewed evidence and spoke to stakeholders, we made a number of reflections that may not be adequately conveyed in the body of the formal report.

1. Nature of bias

The first is on the nature of the bias we found. We found evidence of unfair bias in relation to medical devices but, from what we could discern, it was largely unintentional (for example, related to the physical properties of optical devices on darker skin tones or unrepresentative datasets in AI algorithms), compounded by testing and evaluation in predominantly White populations.

Some of the biases we found were even well-intentioned but misguided, such as the application of race correction factors based on erroneous assumptions of racial or ethnic differences, or attempts to devise 'fairness metrics' for AI devices that aim for equality rather than equity.

This 'unintentionality' – and the fact that many of the participants at our expert roundtables focused on or saw only part of the problem – speaks to us about the need for a whole-system or ecosystem view. Differentials in socio-economic conditions (including power, exposure to health hazards, employment and access to healthcare), together with systemic structural issues, amplify the bias further.

2. Socio-economic disadvantage

The second reflection is that our review, inevitably and rightly, mainly focused on the potential for ethnic and racial inequities, as these are more obviously likely to arise in the medical devices we covered, but ultimately the biggest driver of health inequity in the UK by far is socio-economic disadvantage (regardless of ethnic group).

Device manufacturers, regulatory bodies and the Department of Health and Social Care (DHSC) should always keep in mind their socio-economic responsibilities[footnote 1] to reduce the inequalities of outcome which result from socio-economic disadvantage and ensure new medical devices do not exacerbate these already wide inequities.

3. Regulation of AI

The third reflection is that the exponential increase in AI-driven applications in medical devices has far surpassed any increase in regulation of AI used to support clinical decision-making and AI-derived predictive analytics, including in genomics.

There is a real danger that innovations in medical device technology – whether in optical devices, AI or genomics – will not only outstrip the growth in our health professionals' AI literacy and skills, but will also exacerbate inequity, with potential to change the foundations of the doctor-patient relationship in unpredictable ways.

A government-appointed expert panel is required to oversee the inevitable disruption and potential unintended consequences that will arise from the AI revolution in healthcare. This is our final recommendation in the chapter on AI as a medical device.

4. Addressing inequities in access to medical devices

The fourth reflection is the subject of inequities in access to medical devices, which was out of scope for our review, but is a pressing issue for future action. The poorer access for more disadvantaged socio-economic groups to the new generation of AI and genomic innovations is yet another injustice that should not be tolerated in the NHS.

Inequities in access to medical devices form only one part of the wider problem of inequitable access to the health services that people need, which is largely due to structural biases in the broader health ecosystem. Addressing inequities in access is therefore an essential task for the government and leadership of the NHS.

5. Next steps

Finally, for our recommendations to be impactful, a renewed sense of urgency and commitment to address inequity is required at the highest levels of government. Additional resources are also required for MHRA and approved bodies for medical devices to ensure that equity assessments are conducted as part of the approvals process for new medical devices, and in post-marketing surveillance.

It has been a genuine privilege and honour to work on this independent review, and we hope that the implementation of our recommendations will go some way to turning the tide on inequity and unfair biases in medical devices in the NHS.

Margaret Whitehead
Raghib Ali
Enitan Carrol
Chris Holmes
Frank Kee

About the review

Scope

A core responsibility of the NHS is to maintain the highest standards of safety and effectiveness of medical devices currently available for all patients within its care. Evidence has emerged, however, about the potential for racial and ethnic bias in the performance of some medical devices commonly used in the NHS, and that some ethnic groups may receive suboptimal treatment as a result. Beyond racial and ethnic bias, there may be further unfair biases in performance, including by sex and socio-economic status.

This independent review was tasked by the Secretary of State for Health and Social Care to:

- establish the extent and impact of potential racial, ethnic and other factors leading to unfair biases in the design and use of medical devices
- make recommendations for improvements

Our recommendations were derived from a review of the scientific evidence and extensive engagement activities with both the developers and regulators of medical devices, on the one hand, and users and evaluators of the devices in the NHS, on the other, including the ultimate users: patients and the public.

Terminology

The terms 'race' and 'ethnicity' are often used interchangeably in the literature. We use the terms 'racial and ethnic inequities' and 'racial and ethnic bias' to describe concepts where medical devices do not work as well for some ethnic groups as a result of differences in biological characteristics, genetic predisposition, or under-representation in research.

In the context of polygenic risk scores (one of the categories of device that we reviewed), we use the term 'genetic ancestry' to describe the people that an individual is biologically descended from, including their genetic relationships. Knowledge of a person's ancestry can help determine frequencies of genetic risk variants, which may vary with ancestry.

We briefly explain the relationship between ethnicity and socio-economic status, as deprivation is a major risk factor for most health outcomes, and should also be taken into consideration when making comparisons between ethnic groups.

Applying equity principles to medical devices

There are many equity elements built into the NHS that combine to make it fair and equitable. Considering the relevant elements as equity principles in the context of this review, medical devices approved for use in the NHS should:

- be available to everyone in proportion to need
- support the selection of patients for treatment based on need and risk
- function to the same high standard and quality for all relevant population groups. If there are unavoidable differences in performance in relation to some groups, these need to be understood and mitigated, such as in how the device is calibrated

We reviewed the evidence on violations of these equity principles – in particular, evidence of the application of medical devices leading to biased selection of patients for treatment or differential performance that has the potential to lead to adverse clinical impacts on the health or healthcare of the patients concerned.

This review focused on 3 types of medical device that may be particularly prone to racial, ethnic or other unfair biases:

- optical medical devices
- artificial intelligence (AI)-enabled medical devices
- polygenic risk scores (PRS) in genomics

What we found

Optical medical devices

Section '6. Potential ethnic and other unfair biases in optical medical devices' of the full report focuses primarily on pulse oximeters, but also includes other optical medical devices that take measurements through a patient's skin and where results may vary by skin tone.

There is extensive evidence of poorer performance of pulse oximeters for patients with darker skin tones. Pulse oximeters overestimate true oxygen levels in people with darker skin tones, which is exacerbated in patients with low levels of oxygen saturation.

Evidence of harm stemming from this poorer performance has been found in the US healthcare system, where racial and ethnic bias in the performance of pulse oximeters has been linked to delayed recognition of disease, denied or delayed treatment and, worse, organ function and death in Black compared with White patients.

In these studies, the relationship between oxygenation overestimation and outcome cannot be said to be causative, but points to a strong association. We did not find any evidence from studies in the NHS of this differential performance affecting care but the potential for harm is clearly present.

Recommendations 1 to 3 are therefore focused specifically on pulse oximeters, and cover immediate mitigation measures to ensure existing pulse oximeters can perform to a high standard for all patient groups to avoid serious inequities in health outcomes – improvements in international standards for approval of new pulse oximeter models and the development, ultimately, of smarter devices for measuring blood oxygen saturation that are equally effective across a wide range of skin tones.

We also reviewed evidence for other optical devices where there were scientifically plausible mechanisms for results varying by skin tone. These include:

- near-infrared spectroscopy (NIRS)
- transcutaneous bilirubinometers
- dermoscopes

Evidence is mixed but is suggestive of a degree of bias in such optical devices.

For example, underestimation of tissue oxygenation was found in NIRS readings for participants with darker skin tones, with potentially unnecessary treatment given to patients to improve the oxygen values derived from the spectroscopy, when brain oxygenation is normal.

Overestimation of total serum bilirubin was found when transcutaneous bilirubinometers were used on infants with darker skin tones, which could lead to

needless follow-up blood tests on newborn babies, which are invasive, prolong hospital visits, increase parental stress and interrupt mother-infant bonding.

Some dermoscopes used to diagnose skin cancers and minimise unnecessary biopsies have machine learning algorithms that have been trained on datasets containing images of lesions predominantly from fair-skinned individuals. There are concerns that diagnosis may be delayed or negatively affected in patients with darker skin tones, though there is no evidence as yet of actual harm to patients from bias related to the AI algorithms.

Recommendations 4 to 7 are focused on:

- prevention of potential for harm through improved detection of bias in optical devices as a whole
- better research and testing tools
- more robust monitoring and auditing
- refreshed education of health professionals

AI-enabled medical devices

AI-enabled medical devices are entering the market at an unprecedented pace. Almost under the radar, their acceptance as 'routine' could obscure their potential to generate or exacerbate ethnic and socio-economic inequities in health.

Unfair bias can arise in AI device development and use in several different ways including:

- the way that health problems are selected and prioritised for AI-related development
- how data is selected for use in developing and testing a device
- how outcomes are defined and prioritised in the healthcare system
- how the underlying AI algorithms driving the device's functionality are developed and tested
- how the device's impacts are monitored once in use

The emerging evidence points to a critical need for:

- patients and clinicians to contribute to better articulation and prioritisation of the health 'problems' (for the device to solve)
- better AI and health equity literacy that will ultimately help us focus on the best data and outcomes that should count most in possible solutions to these biases

Solutions, whether through the use of more representative training data for the devices or better monitoring of their deployment to ensure fair outcomes, lie across the lifecycle and, while 'fairer' algorithms are being developed for such devices, they

are sometimes misguided and a whole-system approach will be necessary to mitigate the bias problem.

To addresses these challenges, Recommendations 8 to 14 have the central aim of enabling the development of safe and equitable AI medical devices.

However, the healthcare and regulatory systems that we inhabit today will not look like the systems of tomorrow with the advent of LLM and foundation models (such as ChatGPT), which will disrupt our clinical and public health practice in unpredictable ways. It is imperative that we prepare now for that future. Regulatory bodies like MHRA will need to be adequately resourced to meet all these challenges.

Recommendation 15 therefore is a call for government action to initiate the thinking and planning that will be needed to face this disruption in relation to AI-enabled medical devices.

Polygenic risk scores (PRS) in genomics

Looking to the future, we reviewed devices in genomics utilising PRS, which are already available commercially (through direct-to-consumer tests), but have not yet been adopted by the NHS. PRS are used, among other factors, to assess risk of diseases that have multiple social, environmental and genetic causes.

There are 2 equity concerns. First, the major genetic datasets employed by PRS are drawn from populations that are overwhelmingly of European ancestry, which means that the results of PRS may not be applicable for people with other ancestries. This historical ethnic bias is well recognised and there are many important initiatives being taken at national level in the UK to improve the genetic datasets in the long term. We commend these initiatives and focus our recommendations on the second of our equity concerns – the societal challenges.

There are several societal challenges related to the possible introduction of PRS population-wide that have been relatively neglected so far. These include the:

- possible disruption that PRS may bring to long-standing efforts to tackle modifiable risk factors for disease
- vulnerability of PRS information to misinterpretation by the public – particularly, mistaken beliefs about genetic determinism

There is also the more immediate challenge for the NHS of dealing with patients' concerns about PRS tests that are coming into the UK through commercial, direct-to-consumer routes without any regulation or support for the people who receive this sort of information.

Recommendations 16 to 18 require action on these societal challenges.

Horizon scanning

Below we flag up 3 areas that, though not in scope for this review, cannot be ignored for the future in terms of equity in medical devices. These are the:

- transition of personal 'wearables' from wellbeing devices to medical devices
- wider inequities in access to medical devices that are developing with the advent of the digital device and genomic innovations
- special circumstances surrounding the medical devices encountered by women in pregnancy and the neonatal period

All these need attention now.

Our recommendations

We make 18 recommendations, detailed below, to address the unfair biases that we identified during the course of our review aimed at improving equity in medical devices. These improvements now need to be implemented as a matter of priority with full government support.

This review recommends:

Immediate mitigation for pulse oximeters

Recommendation 1:
Regulators, developers, manufacturers and healthcare professionals should take immediate mitigation actions to ensure existing pulse oximeter devices in the NHS can be used safely and equitably for all patient groups across the range of skin tones.

This requires action on several fronts as follows:

- MHRA should strengthen its guidance for patients and caregivers using oximeters at home, and for healthcare professionals, on the accuracy and performance of pulse oximeters. This should include guidance on taking and interpreting readings from patients with different skin tones. Renewed efforts should be made to promote this guidance to health professionals throughout the NHS, patients and the public

- health professionals should advise patients who have been provided with a pulse oximeter to use at home to look at changes in readings, rather than just a single reading, to identify when oxygen levels are going down and they need to call for assistance. Patients should also be advised to look out for other worrying symptoms such as shortness of breath, cold hands and feet, chest pain and fast heart rate
- clinical guideline developers and health technology assessment (HTA) agencies such as the National Institute for Health and Care Excellence (NICE) should produce guidance on the use of pulse oximeters, emphasising the variable accuracy of readings in patients with darker skin tones, and recommend the monitoring of trends rather than setting absolute thresholds for action
- Health Education England (part of NHS England) and the respective agencies in the devolved nations should educate clinicians about how the technology of pulse oximeters works, and advise that treatment should not be withheld or given on the basis of absolute thresholds alone. Clinicians should be trained to monitor trends rather than absolute thresholds for action
- manufacturers of pulse oximeters must update their instructions for use to inform patients and clinicians about whether the device is ISO compliant, the limitations of their model of pulse oximetry and any contra-indications, and its differential accuracy in patients with different skin pigmentation
- MHRA should issue updated guidance to developers and manufacturers on the need to make the performance of their device across subgroups with different skin tones transparent

Recommendation 2:

MHRA and approved bodies for medical devices should strengthen the standards for approval of new pulse oximeter devices to include sufficient clinical data to demonstrate accuracy overall and in groups with darker skin tones. Greater population representativeness in testing and calibration of devices should be stipulated.

The approach should include:

- MHRA and UK-approved bodies following the US FDA in requiring manufacturers to obtain validity data from a diverse subject pool with a:
    - large number of participants
    - diverse range of skin tones
    - clinically relevant range of oxygenation levels
- manufacturers and research-funding bodies commissioning studies that include the population upon which the device will be used, subjects with a diverse range of skin pigmentations and critically unwell subjects with poor perfusion. Validation of devices should be conducted in the intended use population and setting, such as at home or in an intensive care unit

- manufacturers of medical-grade pulse oximeters being required to comply with BS EN ISO 80601-2-61:2019 (medical electrical equipment – particular requirements for basic safety and performance of pulse oximeter equipment) to gain market approval
- healthcare equity impact assessments being essential requirements for developing or supplying pulse oximeters in the UK in order to identify whether mitigating actions are needed to ensure they are fit for purpose for all racial and ethnic groups and people of varying skin tones. Making these assessments an essential requirement is in line with technological progress and international best practice.

Recommendation 3:

Innovators, researchers and manufacturers should co-operate with public and patient participants to design better, smarter oximeters using innovative technologies to produce devices that are not biased by skin tone.

This could include:

- developing enhanced algorithms for oximeter device software to address measurement bias
- exploring the use of multi-wavelength systems, which measure and correct for skin pigmentation, to replace conventional 2-wavelength oximeters

Equity of optical medical devices

Recommendation 4:

The professional practice bodies in the UK, such as the Royal Colleges, should convene a task group of clinicians from relevant disciplines – including medical physicists, public and patient participants, developers and evaluators – to carry out an equity audit of optical devices in common use in the NHS, starting with dermatological devices, to identify those at particular risk of racial bias with potential for harm, which should be given priority for further investigation and action.

Recommendation 5:

Renewed efforts should be made to:

- increase skin tone diversity in medical imaging databanks used for developing and testing optical devices for dermatology, including in clinical trials
- improve the tools for measuring skin tone incorporated into optical devices

This will require a concerted effort on several fronts, including:

- encouraging links between imaging databank compilers, professional bodies, optical device developers and clinicians to develop and improve accessibility of imaging data resources that reflect skin tone diversity within the population, such as in databanks for skin cancer diagnosis
- MHRA providing strengthened guidance to developers and manufacturers on improving skin tone diversity in testing and development of prioritised optical devices. MHRA is already working towards such guidance as part of its programme on pulse oximeters
- research funders supporting additional incentives and patient-centred approaches to address logistical, financial and cultural barriers that limit participation of minority ethnic groups in clinical studies of optical devices
- researchers and dermatologists developing more accurate methods for measuring and classifying skin tone, which are objective, reproducible, affordable and user-friendly. Current practice of using uncertain descriptors of ancestry, ethnicity or race to define patients with dark skin tones is ambiguous and problematic. In its discussions on updating standards, MHRA is examining which measures would be most appropriate, with the aim of agreeing a consensus. This work is to be commended

Recommendation 6:

Once in use, optical devices should be monitored and audited in real-world conditions to evaluate safety performance overall and by skin diversity. This will ensure any adverse outcomes in certain populations are identified early and mitigations implemented.

This requires a whole-system approach and should include:

- commitment from manufacturers at the pre-qualification stage to fund and facilitate the establishment of registries for collecting data across all population groups on patient demographic characteristics, use and patient outcomes, following deployment of the technology
- HTA agencies (such as NICE, the Scottish Health Technologies Group and Health Technology Wales) being provided with access to post-deployment monitoring and adverse effects data as part of their assessments of optical devices. This data should be considered alongside the wider evidence when determining the value of the optical device for NHS use
- NHS Supply Chain, National Services Scotland, NHS Wales Shared Services Partnership, Northern Ireland Procurement and Logistics Service and other contracting authorities including a minimum standard of device performance across subgroups of the target population, which will make transparent any equity impacts as part of the pre-qualification stage when establishing national framework agreements. Manufacturers need to declare whether they have considered minimum standards for equity
- DHSC and the devolved administrations updating the national pre-acquisition questionnaire used by NHS trust electrical biomedical engineering teams when buying medical equipment to include a minimum

designated standard for equity as part of the pre-purchase validation checks

- the approved body conducting regular surveillance audits of prioritised optical devices. The audits should include data submissions from the manufacturer and the Medical Device Safety Officers or Incidents and Alerts Safety Officers networks (representatives from NHS trusts in charge of reporting on safety), and should include data from the [MHRA Yellow Card](#) scheme for reporting adverse incidents and the [Learn from patient safety events service](#). These audits should include an evaluation of differential safety by ethnic group
- the continued strengthening of MHRA's vigilance role, as specified in the [Cumberlege report](#)'s recommendation 6, which called for substantial improvements in adverse event reporting and medical device regulation with an emphasis on patient engagement and outcomes
- better routine capturing of ethnicity data in electronic healthcare records, alongside better collection and collation of data on medical devices in use. This would enable MHRA to conduct more rapid studies to build the evidence when a hypothesis about potential inequity in an optical device is made

Recommendation 7:

A review should be conducted by the relevant academic bodies of how medical education and continuing professional development requirements for health professionals currently cover equity issues arising in the use of medical devices generally and skin diversity issues in particular, with appropriate training materials developed in response.

This should include:

- undergraduate and postgraduate medical and allied health professions training, including teaching clinicians about clinically relevant conditions where disease presentation differs between White and ethnic minority patients
- clinicians being made aware that, when using dermoscopy or other medical devices to examine skin lesions, clinical signs may differ according to skin tone, and their training should include images of skin lesions in all skin tones
- clinicians receiving training in identifying potential sources of bias in medical devices and how to report adverse events to MHRA
- where new devices are introduced into clinical practice, organisations and clinicians using the new devices, ensuring there is sufficient training to acquire skills and competencies before the device is used

Preventing bias in AI-assisted medical devices

Recommendation 8:

AI-enabled device developers and stakeholders, including the NHS organisations that deploy the devices, should engage with diverse groups of patients, patient organisations and the public, and ensure they are supported to contribute to a co-design process for AI-enabled devices that takes account of the goals of equity, fairness and transparency throughout the product's lifecycle.

Engagement frameworks from organisations such as NHS England can help hold developers and healthcare teams to account for ensuring that existing health inequities affecting racial, ethnic and socio-economic subgroups are mitigated in the care pathways in which the devices are used.

Recommendation 9:

The government should commission an online and offline academy to improve the understanding among all stakeholders of equity in AI-assisted medical devices.

This academy could be established through the appropriate NHS agencies, and should develop material for lay and professional stakeholders to promote better ways for developers and users of AI devices to address equity issues, including:

- ensuring undergraduate and postgraduate health professional training includes the potential for AI to undermine heath equity, and how to identify and mitigate or remove unfair biases
- producing materials to help train computer scientists, AI experts and design specialists involved in developing medical devices about equity, and systemic and social determinants of racism and discrimination in health
- ensuring that clinical guideline bodies identify how health professionals can collaborate with other stakeholders to identify and mitigate unfair biases that may arise in the development and deployment of AI-assisted devices
- encompassing an appreciation of AI within a whole-system and lifecycle perspective, and understanding the end-to-end deployment and potential for inequity

Recommendation 10:

Researchers, developers and those deploying AI devices should ensure they are transparent about the diversity, completeness and accuracy of data through all stages of research and development. This includes the sociodemographic, racial and ethnic characteristics of the people participating in development, validation and monitoring of product performance.

This should include:

- the government resourcing MHRA to provide guidance on the assessment of biases that may have an impact on health equity in its evaluation of AI-assisted devices, and the appropriate level of population detail needed to ensure adequate performance across subgroups

- encouraging the custodians of datasets to build trust with minoritised groups and take steps with them to make their demographic data as complete and accurate as possible, subject to confidentiality and privacy
- developers, research funders, regulators and users of AI devices recognising the limitations of many commonly used datasets, and seeking ones that are more diverse and complete. This may require a concerted effort to recruit and sample underrepresented individuals. We commend initiatives internationally and in the UK (such as the National Institute for Health and Care Research-led INCLUDE guidance) to encourage the development and use of more inclusive datasets. Data collection by public bodies must be properly resourced so that datasets are accurate and inclusive
- dataset curators, developers and regulators using consensus-driven tools, such as those by STANDING Together, to describe the datasets that are used in developing, testing and monitoring
- regulators requiring manufacturers to report the diversity of data used to train algorithms
- regulators providing guidance that helps manufacturers enhance the curation and labelling of datasets by assessing bias, being transparent about limitations of the data, the device and the device evaluation, and how to mitigate or avoid performance biases
- regulators enforcing requirements for manufacturers to document and publicise differential limitations of device performance and, where necessary, place reasonable restrictions on intended use
- making sure that the Health Research Authority and medical ethics committees approving AI-enabled device research do not impose data minimisation constraints that could undermine dataset diversity or the evaluation of equity in the outcomes of research

Recommendation 11:

Stakeholders across the device lifecycle should work together to ensure that best practice guidance, assurance and governance processes are co-ordinated and followed in support of a clear focus on reducing bias, with end-to-end accountability.

This should include:

- MHRA adjusting its risk assessment of AI-assisted devices so that all but the simplest and lowest-risk technologies are categorised under Class IIa or higher, including a requirement for their algorithms to be suitable for independent evaluation, the use of a test of overall patient benefit that covers the risks of biased performance, and a requirement for manufacturers to publish performance audits with appropriate regularity that include an assessment of bias
- supporting health professionals' involvement early in the development and deployment of AI devices. We commend the use of ethical design checklists, which may assist in the quality assurance of these processes

- manufacturers adopting MHRA's [Good Machine Learning Practice for Medical Device Development: Guiding Principles](#)
- all stakeholders supporting MHRA's [Software and AI as a Medical Device Change Programme Roadmap](#), such as promoting the development of methodologies for the identification and elimination of bias, and testing the robustness of algorithms to changing clinical inputs, populations and conditions
- placing a duty on developers and manufacturers to participate in auditing of AI model performance to identify specific harms. These should be examined across subgroups of the population, monitoring for equity impacts rather than just unequal performance

Recommendation 12:

UK regulatory bodies should be provided with the long-term resources to develop agile and evolving guidance, including governance and assurance mechanisms, to assist innovators, businesses and data scientists to collaboratively integrate processes in the medical device lifecycle that reduce unfair biases and their detection, without being cumbersome or blocking progress.

Recommendation 13:

The NHS should lead by example, drawing on its equity principles, influence and purchasing power, to influence the deployment of equitable AI-enabled medical devices in the health service.

This should include:

- NHS England and the NHS in the devolved administrations including a minimum standard for equity as part of the pre-qualification stage when establishing national framework agreements for digital technology
- NHS England updating the digital technology assessment criteria used by health and social care teams when buying digital technology to recommend equity as part of the pre-purchase validation checks
- working with manufacturers and regulators to promote joint responsibility for safety monitoring and algorithm audits to ensure outcome fairness in the deployment of AI-assisted devices. This will require support for the creation of the right data infrastructure and governance

Recommendation 14:

Research commissioners should prioritise diversity and inclusion. The pursuit of equity should be a key driver of investment decisions and project prioritisation. This should incorporate the access of underrepresented groups to research funding and support, and inclusion of underrepresented groups in all stages of research development and appraisal.

This should include:

- requiring that AI-related research proposals demonstrate consideration of equity in all aspects of the research cycle
- ensuring that independent research ethics committees consider social, economic and health equity impacts of AI-related research

Recommendation 15:

Regulators should be properly resourced by the government to prepare and plan for the disruption that foundation models and generative AI will bring to medical devices, and the potential impact on equity.

A government-appointed expert panel should be convened – made up of clinical, technology and healthcare leaders, patient and public involvement representatives, industry, third sector, scientists and researchers who collectively understand the technical details of emerging AI and the context of medical devices – with the aim of assessing and monitoring the potential impact on AI quality and equity of LLM and foundation models.

Future proofing: polygenic risk scores

Recommendation 16:

The focus of PRS studies should be widened beyond genetic diversity to include:

- the contribution of the social determinants of health – including lifestyle, living and working conditions, and environmental factors such as air pollution – to overall disease risk
- how these affect the predictive potential of PRS among different ethnicities and socio-economic groups

Developments with this wider research focus should aid the refinement of overall risk assessments so that they better reflect the role that PRS play alongside non-genetic risk factors.

Recommendation 17:

National research funders should commission a broad programme of research and consultation with the public, patients and health professionals to fill the gaps in knowledge and understanding concerning PRS. The programme should cover:

- the public's understanding of the nature of genetic risk and the meaning of the PRS they are presented with
- explorations of how health professionals interpret these risks, and can best communicate and support people in understanding the results of their PRS

The research programme should cover impacts on diverse population subgroups, and be informed by extensive engagement with the public and patients to gain their perspectives.

Results from this research programme, together with actions on recommendation 16, should feed into the development of clinical applications for PRS medical devices covered in recommendation 18.

Recommendation 18:

UK professional bodies – such as the Royal Colleges and health education bodies across the UK – should develop guidance for healthcare professionals on the equity and ethical challenges and limitations of applying PRS testing in patient care and population health programmes.

The guidance should:

- include the interpretation of risk scores, communicating risk to patients and the public, and counselling and support
- be informed by extensive public and patient engagement

Horizon scanning and next steps

During our review, we have noted several issues regarding equity in medical devices looming on the horizon. Some are topics that were out of scope for this review, but nevertheless raise important equity issues for the future.

We highlight here the continuing growth of wearables that are crossing over from personal wellbeing improvement devices to medical devices, the continuing challenge of inequities in access to medical devices – made all the more critical with the growth of the digital health divide – and a pressing need to improve equity in medical devices used routinely in pregnancy and the neonatal period.

Wearables

The first is the advent of wearables – electronic devices such as smartwatches and fitness trackers designed to be worn on the user's body. Many devices collect health-related data such as heart rate, blood pressure, sleep patterns and physical activity using biosensors on the wearer's skin.

Currently, they are largely marketed to consumers as 'wellbeing devices' and so are not subject to medical device regulations. As such, they were outside the scope of

our current review. During the course of our review, however, we discovered that there are many clinical applications under development, including for cardiovascular management and mental health monitoring[footnote 2][footnote 3], that would eventually bring them into the category of 'medical device'.

The equity issues

They have the potential to suffer from the same kind of biases inherent in the medical devices in our review. For example, if they use optical techniques through the skin to track physiological change, they may not be as accurate, or may not work at all, in people with darker skin tones. The ability of smartwatches to track heart rates in people with dark skin is already being questioned.[footnote 4] Testing of the devices on mainly White participants is a familiar underlying problem.

The large datasets on which the algorithms that are driving these devices draw are likely to be unrepresentative of ethnic minority and more disadvantaged socio-economic groups because of the well-known bias in recruitment to such studies. This is compounded in digital studies by a common requirement for participants to have their own expensive piece of equipment, such as a smartwatch. One such study of the accuracy of an arrhythmia detection algorithm requiring ownership of an Apple product was found to be biased towards a young, wealthy and technology-savvy population.[footnote 5] This may render the algorithms only applicable to more affluent groups reflecting the study composition.

Then there are potential equity issues with the internal algorithms and computational models being developed for applications in psychiatry, whether for mental illness detection, prediction or individualisation of treatment.[footnote 3] In many instances, digital psychiatry is moving to online interactive platforms driven by AI algorithms that attempt to interpret the responses and language being used by 'patients' to describe their symptoms. But what is emerging from the literature is that current natural language processing algorithms being used on these platforms can be biased against certain ethnic groups because of the different ways people from different cultures and ethnicities express themselves.[footnote 6]

It is clear that the equity implications of wearables used as medical devices will need to be assessed as a next step in preparation for their increasing adoption in clinical practice.

Inequities in access to medical devices

Second is the issue of inequitable access to medical devices or to the services they support. Again, this issue was judged out of scope of our review, though equitable access is noted as an important component of equity in the NHS system as a whole in section '3. What does 'striving for equity in medical devices' mean?' of the full report.

Essentially, we reviewed evidence related to medical devices causing biased selection of patients or exhibiting biased performance against one or more groups in the population. The separate equity issue of whether all population groups can gain access to effective medical devices on the basis of need was outside our remit and, indeed, draws on completely different evidence. Nevertheless, many access issues were brought to our attention during the review upon which we have been reflecting.

The equity issues

With the advent of digital health technologies and recent genomic innovations, new manifestations of the inverse care law are emerging all the time. This 'law' ("the availability of medical care tends to vary inversely with the need of the population served"[footnote 7]) can be seen in the tendency for digital health innovations to be available and taken up more readily in more affluent groups and areas with better health profiles in the first place.

The concept of 'digital poverty' or 'digital exclusion' has been invoked to capture the experience of groups in society who do not have full access to the online world when they need it and so are excluded from the benefits of advances in digital services that are on offer. This exclusion could be because of:

- cost
- language barriers
- technological proficiency

In 2022, for example, a US study found that wearables and other digital devices were not used as widely in minority and low-income groups, with cost and education affecting use.[footnote 8] Online services offered in primary care and telemedicine may benefit the staff but exclude the elderly, those with low educational attainment or poorer patients with the greatest need.

This concern about inequities in access to medical devices is also growing in relation to genomics. Once an effective but expensive pharmacogenomic treatment becomes available, for instance, questions arise about who gets it. But, with the under-representation of certain ethnic groups in pharmacogenetic research[footnote 9], there will be far more uncertainty around the cost-effectiveness of such tailored therapy in ethnic minority groups, casting doubt on the equitable allocation of scarce resources.

The issue of inequities in access to medical devices in the NHS has become even more pressing with the new technologies on the horizon. Addressing this is an essential task for the government and NHS leadership.

Equity in medical devices during pregnancy and the neonatal period

Third is the special circumstances surrounding pregnancy and the neonatal period, when all women and their babies under the care of the NHS encounter a variety of medical devices in routine screening tests, some of which will have the potential for ethnic or socio-economic bias. This is a critical situation because of the marked ethnic and socio-economic inequities in pregnancy outcomes, which the NHS should be striving to reduce rather than exacerbate.

During our review, examples from the pregnancy and neonatal period came up in all 3 types of medical device we studied. But studying potential bias in individual medical devices could not give a complete picture of the cumulative effect that exposure to a variety of devices might have if encountered over a 9-month period.

An alternative approach would be to start from the perspective of patients rather than the devices. This approach would be to follow women's experiences with the various tests and devices throughout pregnancy, and see whether subsequent pregnancy outcomes differed by ethnic or socio-economic group.

The equity issues

As a first step towards this approach, we commissioned a rapid review of the evidence taking such a perspective, which found evidence of the potential for ethnic bias in 3 of the routine tests classed as medical devices in pregnancy and 3 for newborn babies.[footnote 10] There was also evidence of adjustments that reduced or eliminated the bias in the devices in some instances.

It was clear, however, that there were substantial scientific debates in this field about the best course of action to tackle the identified ethnic bias in the devices and, indeed, whether the ethnic disadvantage observed in specific health conditions in pregnancy was attributable to the effects of socio-economic disadvantage, rather than to distinct ethnic differences.

The task of building a consensus as a basis for recommendations to improve equity in medical devices used along the pregnancy pathways is therefore a substantial undertaking in its own right, and one that needs to be carried forward with some urgency.

Our final call for action as a next step, therefore, is that a review should be carried out of equity in the medical devices encountered during pregnancy and the neonatal period, as part of the wider investigations of health outcomes for ethnic minority women and their babies.[footnote 11]

1. As defined by the Equality Trust. Socio-economic duty (accessed 1 June 2023). ↩

2. Zinzuwadia A and Singh J. 'Wearable devices - addressing bias and inequity.' The Lancet Digital Health 2022: volume 4, issue 12. ↩

3. Hauser T and others. 'The promise of a model-based psychiatry: building computational models of mental health.' The Lancet Digital Health 2022: volume 4, issue 11. ↩ ↩2

4. Colvonen P and others. 'Limiting racial disparities and bias for wearable devices in health science research.' Sleep 2020: volume 43, issue 10. ↵

5. Perez M and others. 'Large-scale assessment of a smartwatch to identify atrial fibrillation.' New England Journal of Medicine 2019: volume 381, issue 20. ↵

6. Straw I and Callison-Burch C. 'Artificial Intelligence in mental health and the biases of language based models.' PLoS One 2020: volume 15, issue 12. ↵

7. Tudor Hart J. 'The inverse care law.' The Lancet1971: volume 297, issue 7696. ↵

8. Holko M and others. 'Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices.' npj Digital Medicine 2022: volume 5, issue 53. ↵

9. Asiimwe I and Pirmohamed M. 'Ethnic diversity and warfarin pharmacogenomics.' Frontiers in Pharmacology 2022: volume 13. ↵

10. McHale P and others. 'Equity in medical devices: a rapid review of potential biases for informing patient pathways in pregnancy.' Commissioned report for the independent review of equity in medical devices. University of Liverpool. 2022. ↵

11. Department of Health and Social Care. Maternity Disparities Taskforce: terms of reference (accessed 1 June 2023). ↵