Ethical issues arising from the identification of patients with undiagnosed rare diseases from electronic patient records in prospective scenarios

Executive summary

The systematic detection of undiagnosed rare and hard-to-diagnose diseases offers potential benefits to patients and to health systems. This report examines the ethical issues arising from the prospective analysis of electronic patient records located in NHS affiliated GP Practices using the MendelScan Tool. Mendelian's aim is to flag patients who meet diagnostic criteria in one of eight rare/undiagnosed diseases. This report focuses on seven rare inherited diseases (detailed in section 3 below). This report does not address familial hypercholesterolemia because this disease, although genetic, raises substantially different issues to the other diseases under review.

Classification of the test and verifying and evaluating its results

Any test carries with it the potential for benefits and harms. The potential benefits and harms (and uncertainties) of using this test in this population are not yet clear. This is due to a variety of factors:

1. Many of the diseases do not follow a consistent pattern of inheritance or clinical path: the genotype and phenotype associated with many of these diseases are heterogeneous. Some diseases have incomplete penetrance. Any tool used for mining and evaluation of patient data needs to demonstrate sufficient sensitivity and specificity to show that it works effectively in this patient population. The accuracy of the Tool is unknown, and thus the potential benefits and harms associated with its use, particularly in generating false positive or false negative results are unclear

2 The type of test is unclear: The status of the test – whether it is screening, case finding, service evaluation or part of clinical care – dictates the ethical landscape and the governance which should apply. We have set out some of the challenges associated with each designation

Regulators and policy makers are becoming increasingly concerned that algorithms should only be used in ethical ways: there has been a proliferation of guidance, but two principles, identifiability of data and transparency, have emerged as key. These concepts underpin the development of ethical principles surrounding the use of algorithmic tools as well as the legislation and governance supporting their use.

Processing patient data

One key question is the extent to which the data accessed by Mendelian constitutes personal data, and as such falls within the jurisdiction of the General Data Protection Regulation. If it does count as personal data, (a consideration which is complicated by the inclusion of rare genetic data as part of the data which will be under review by Mendelian), then various requirements apply. e.g. to be transparent about the purposes of the test, and to provide a meaningful explanation to users of how the Tool works, as well as a wider explanation of what it is trying to achieve.

mendelian

Another consideration is the fact that Mendelian is analysing de-identified patient EHRs to identify patients who are potentially at risk of developing disease. The secondary use of electronic patient records is potentially contentious and sensitive. This potential sensitivity is heightened by the fact that genetic and genomic information (but not raw genetic and genomic data) is included in the data evaluated by Mendelian: furthermore, Mendelian is a commercial company – which, according to the literature, is another potential source of concern. The report explores some of the ethical challenges associated with using patient EHRs for this purpose.

A key aspect of this proposed application is that the tool developed by Mendelian (MendelScan) will be used to flag patients at risk for disease from analysis of EHRs. This scanning function is relatively new, but some exemplars already exist. Two exemplars are noted in the report: the electronic frailty index (eFI) and case finding for familial hypercholesterolaemia (FH) but these are not addressed in detail because they fall beyond the scope of this report.

Implementing the test in the health system

We have been provided with an example of a Deployment Plan which sets out a potential pathway by which reports flagged by Mendelian are returned to patients' GPs and a set of potential actions that GPs might take on receipt of a flagged report. This includes recall of the patient for validation or specialist referral. The final section of this report explores the proposed pathway for Mendelian to return results to GPs, and then on to patients. It identifies a series of considerations which Mendelian might wish to take into account, for each part of this process. It also identifies three overarching recommendations for the project as a whole:

1. Clarifying whether the activity is research, screening or clinical care, and the implications that flow from this choice
2. Being transparent to patients and staff about what is proposed, using the website of a GP provider or other appropriate mechanism
3. Building formal patient participation into the delivery and evaluation stages, so that patient views are robustly represented throughout the lifespan of the project

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

1. Introduction

The expansion of genetic tests and development of novel 2nd and 3rd generation genomic testing has led to a wholesale change in how genetic and genomic tests are used within patient pathways. This has led to new opportunities for identifying patients at risk of developing disease which now extend beyond clinical genetics teams to other clinical specialities within secondary care, and even to primary care.

The purpose of this report is to analyse the potential application of a tool (MendelScan) to identify patients at risk of disease from evaluation of electronic patient records held by GPs. These records include details of consultations between patient and GPs but also letters detailing secondary and tertiary care. They may include genetic/genomic information in the form of test results and discharge summaries but not raw genetic/genomic data in the form of BAM or VCF files.

This report evaluates the ethical issues which might arise from the prospective use of electronic health records (EHRs) for identifying seven rare inherited diseases using MendelScan tool. It explores various aspects of the project, including project design, issues arising from the processing of data, utilisation of the MendelScan, and the interface between developers and health care professionals. Specifically it analyses how requirements for transparency and explanation might impact on communication and consent and might be influential in building trust and confidence of health care professionals and patients in what is being proposed.

2. The Mendelian Tool

Mendelian have developed a tool to systematically search a large set of patients' EHRs. An example of the deployment of this Tool is described in a project plan. Figure 1 of the plan describes data flow and systems integration through which Mendelian accesses pseudonymised patient data via a third party provider responsible for aggregating the clinical data within the Clinical Commissioning Groups (CCG) ecosystem.

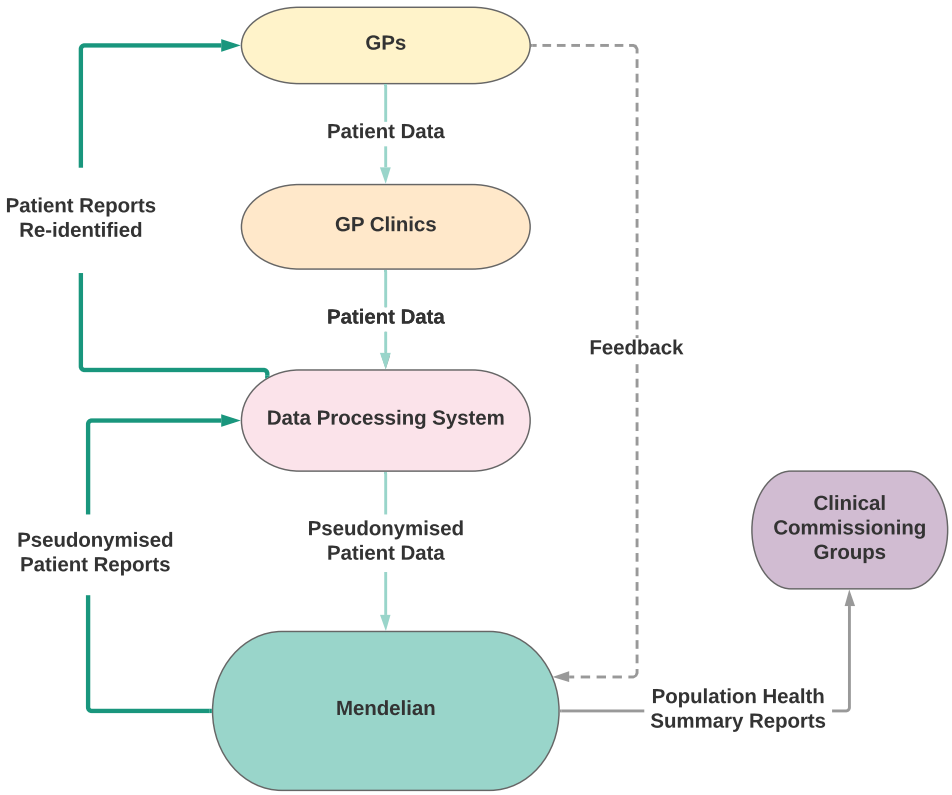

Figure 1: Data flow and systems integration

These records will constitute input data which will be used to identify pseudonymised patients who meet the diagnostic criteria associated with a number of specified diseases. The re-identification process is described in Figure 2 of the project plan.

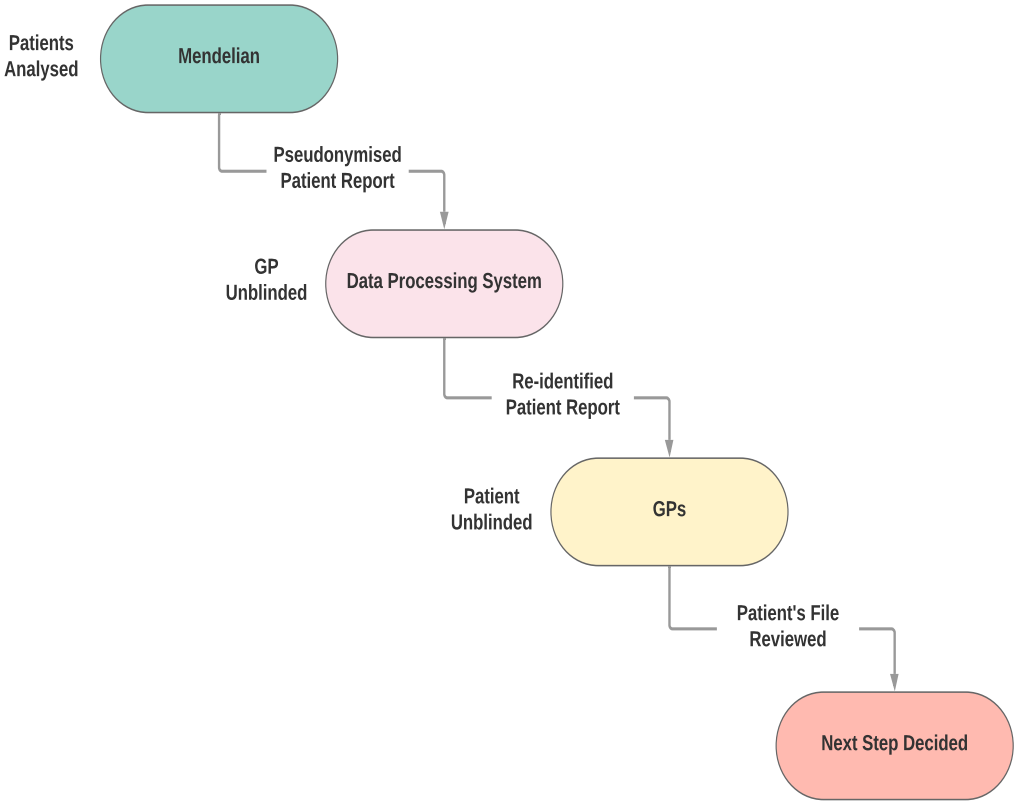

Figure 2: Re-identification procedure

Once patients are flagged, 'the system outputs a report outlining the next clinical steps in the pathway, sourced directly from established clinical care guidelines used to discover and validate the flagged condition.' It is Mendelian's intention that the MendelScan Tool will never be used in isolation. A clinician will always be involved in making a final decision about a patient's diagnosis and management.

3. Target diseases

This report focuses on the seven rare diseases selected for this project. These are Bardet-Biedel Syndrome, Alström Disease, Fabry Disease, Duchenne Muscular Dystrophy, TTR Amyloidoisis, Mucopolysaccharidosis Type 1, and Gaucher Disease. This report will not address screening/case finding in FH as this is a relatively common genetic condition (prevalence 1:250) and as many as 340 people may be undiagnosed with this condition in this GP federation.

Each of the seven diseases selected are diseases included in the NHS Highly Specialised Service commissioning route; have an NHS defined referral pathway; published patient finding criteria exist which are capable of detection using the MendelScan; and the disease is currently often mis-or under diagnosed. The diseases selected for initial trials are a subset of the diseases which MendelScan may eventually cover, although selected diseases are likely to share many of these characteristics.

3.1 What are the potential benefits of targeting these diseases?

The potential benefits of targeting these diseases is that detection enables appropriate referral and treatment of diseases that might otherwise remain underdiagnosed within the GP patient population. The diseases that have been targeted have a variety of different inheritance patterns: for example, Fabry disease is an X-linked enzyme deficiency disorder whilst most other lysosomal storage disorders are inherited as autosomal recessive traits.

3.2 What are benefits of early diagnosis?

Disease	Inheritance pattern	Presentation	Phenotype	Treatment/ intervention
Fabry disease[2]	X-linked	Neuropathic pain first sign in childhood	Serious cardiac, neurological and renal disease in adulthood	Enzyme replacement (ERT) is available
Gaucher disease	Autosomal recessive disorder	3 types all of which present in children	Variable presentation	Enzyme replacement therapy for types I and III. Substrate inhibiting drug Miglustat for type I. No treatment available for type II
Mucopolysac-charidosis Type 1 (Hurler syndrome)	Autosomal recessive disorder	Variable presentation depending on sub-type	Complex presentation across organ systems	Severe forms treated by hematopoietic stem cell transplantation (HSCT). Less severe forms by ERT e.g. Laronidase
Alström syndrome[3]	Autosomal recessive disorder	Typically blindness in childhood	Insulin-resistant diabetes, fibrosing cardiomyopathy, renal failure, deafness	Integrated clinical management in specialist clinics

Disease	Inheritance pattern	Presentation	Phenotype	Treatment/ intervention
Bardet-Biedl syndrome[4,5]	Autosomal recessive disorder	Obesity, blindness and renal failure	Also diabetes, global learning difficulties, Hirschprung disease and urological and neurological deficits	Prompt treatment of organ-specific problems e.g. renal failure slows disease progression. Comprehensive multidisciplinary clinics available for all ages
Duchenne Muscular Dystrophy	X-linked	Incomplete dystrophin production presents by age 3	Motor dysfunction with progressive GI tract and cardiac and respiratory decline	Ataluren can treat some forms of DMD resulting from nonsense mutations in the dystrophin gene in a sub-group of patients
TTR Amyloidosis	Autosomal dominant inheritance. Incomplete penetrance	Progressive disease presenting in adulthood	Pathogenic accumulation of amyloid protein	Various treatments including transplantation depending on organ systems affected

3.3 Potential ethical challenges associated with targeting these conditions

As seen above, the diseases that have been selected for case finding using the MendelScan Tool have variable inheritance patterns. Often they are heterogeneous in nature. For example, there are over 50 lysosomal storage diseases covered by the paediatric NHS England specification, but this notes that the 'LSD's … show a remarkably varied phenotype' with presentation ranging from newborn to late adulthood. Some sub-types of the diseases selected for analysis appear to have no available treatments available.

The justification for Mendelian selecting these conditions is that detection offers more benefits than harms. Given the diversity of these conditions and the likely presentation – how confident is Mendelian of the following issues:

- Sensitivity for each application: the ability of the MendelScan Tool to correctly identify those with the disease (true positive rate). What is the sensitivity of the MendelScan Tool for each of these conditions? Have likely phenotypic symptoms been included given the diversity of the phenotypic symptoms associated with each of these conditions? Will the MendelScan Tool detect new *de novo* variants that are mentioned in test results or health records? If so, how will these be dealt with?

- Specificity for each application: the ability of the MendelScan Tool to correctly identify those without the disease (true negative rate). The diseases are often heterogenous and have complex phenotypes. How robust are the confidence rates for null findings especially where differential diagnosis is finely balanced? Will individuals who are identified as being potentially at risk of disease be subsequently reviewed? How frequent will the diagnostic criteria be reviewed and the evidence base underpinning each of these criteria be re-analysed?

- Inheritance pattern: many of these diseases are inherited in an autosomal recessive fashion. Some, e.g. Bardet Biedl syndrome, are associated with communities where consanguineous marriage is common. Has account been taken of the socio-demographic characteristics of particular sub-groups in the catchment area of GP users?

- Autosomal recessive inheritance and misattributed paternity: four of the diseases under review are autosomal recessive conditions, meaning that a variant is inherited from both parents. How might the possibility of misattributed paternity be dealt with by GPs who see patients at their first consultation to discuss the flagged reports?

- Phenotypic development: the rationale for using the MendelScan Tool is that it might detect individuals who are mis- or underdiagnosed. Is it possible that it could be used to detect individuals at an earlier stage of disease progression, who do not meet existing diagnostic criteria? If this occurs, how will this group of potential patients will monitored? Given that most clinical genetic services do not have formal systems for instituting regular review, do the GP clinics have an infrastructure in place that might enable this, for those patients where referral to specialists are not warranted? Will phenotypic criteria be weighted according to the likely time of presentation (e.g. polydactyly) and the likelihood that they are associated with a particular condition?

- Eligibility for treatment: access to some of the treatment options for a number of the conditions under review is limited. Some of the treatments are very expensive and whilst funding arrangements have been reached, these may be for a limited period and restricted to certain subgroups of patients. For example, a drug has been developed to treat a sub-group of patients with a nonsense mutation, thought to represent around 10-15% of all DMD cases[6]. Patients in this group who are aged 5 and over and able to walk 10 steps unaided may be eligible for Ataluren through a NICE Managed Access Agreement (MAA)[7]. In return they must sign a Managed Access Patient Agreement through which the manufacturer has access to patient information to inform post-marketing surveillance and drug development. In this case, the committee concluded because of the clinical benefits in the relevant population in clinical practice, that Ataluren would represent value for money only when it was given through a MAA that reduced the costs to the NHS

    - To what extent will GPs/health professionals and patients be made aware of these limitations?
    - Does the fact that only a subgroup of patients is eligible for a treatment or intervention undermine the potential benefits of early detection?

- Case finding: the inherited nature of these diseases suggests that results from flagged patients could be used as a basis for identifying family members who might be at risk of disease, indeed the Mendelian Deployment plan describes this as 'post-hoc'. To what extent will the features identified on the screen be used by GPs for this purpose?

- Scaling up: patients who are flagged as being potentially at risk, have the opportunity for further investigation and potentially a diagnosis to be made. Flagging may also enable other actions to take place, such as surveillance of the patient's ongoing disease, improved symptomatic management, prompt recruitment to clinical trials and putting the patient in contact with appropriate patient advocate groups. Taken together, these actions can have a profound effect on patient lives

3.4 Conclusions

- The diseases that have been selected for evaluation of the MendelScan Tool are heterogeneous both phenotypically and genetically
- There are a number of actions that could be taken following the 'flagging' of a patient's EHR, including closer surveillance to detect early signs of disease occurrence or progression, improved symptom control; early recruitment to suitable clinical trials and more effective psychological support through patient advocate groups. These interventions offer considerable benefits to patients
- Although diagnostic criteria have been identified for each of these conditions, and services centrally commissioned, many caveats remain as to whether patients identified by the MendelScan Tool will have potential diagnoses confirmed by appropriate specialists; whether they will then be eligible for treatments (especially ERT and high-cost interventions); and the extent to which these findings can be relied on for cascading to other family members. This is likely to be relevant for cascade screening for familial hypercholesterolemia where lipid clinics manage this process[8]
- Our recommendation is that no findings are cascaded until findings have been confirmed and validated by an appropriately qualified specialist

4 Is the proposed activity screening, case finding or diagnosis?

One of the key determinants of the ethical issues that might arise is whether the activity is research, case finding or a diagnostic test embedded in clinical care. The boundaries between these three activities are often blurred and heavily dependent on context.

The following figure summarises the distinction between screening, case finding and diagnostic testing.

	Screening test	Case finding	Diagnostic test
Purpose	To detect potential disease indicators	Systematic target at-risk individuals	To establish presence/ absence of disease
Target population	Large nos of asymptomatic potentially at risk inds	Suspected at risk from particular disease	Symptomatic individuals or asymptomatic individuals with a positive screening test
Test method	Simple/acceptable	?	Potentially invasive/ expensive (justification to establish diagnosis)
Positive result threshold	High sensitivity not to miss potential disease	To guide further (potentially invasive) screening and treatment	High specificity (true negative)
Positive result	Indicate suspicion of disease	Indicates likely disease	Definitive diagnosis
Cost	Cheap	Intermediate cost	More expensive

The Mendelian Tool will be applied to the entire GP population which will constitute the sample population. Most of these individuals will be asymptomatic and healthy. These individuals will have no knowledge that their data is being used in this way. Indeed the majority will remain ignorant of this: the small minority of individuals who are approached by their GP following detection of symptoms, for further investigation, will be given some explanation of why they have been approached. It is not clear precisely what form this will take.

Individuals who potentially have some or all of the characteristics of the diseases under review will be flagged for further investigation by the Mendelian Tool. The Tool will utilise de-identified data which will have name and other identifiers removed, but will be tagged with a unique patient ID. This de-identification will be done by an intermediary company responsible for data processing (Figure 1). Flagged patients will be re-identified through a two stage process, first unblinding the GP to potential at-risk patients, which can only be done by the third party data processing system, and subsequently unblinding the patient by that GP to enable patient review and whether validation or referral is justified (Figure 3).

As described in section 3.3, it seems likely that the Tool will not perform perfectly from the outset, and is likely to generate some false positives and negatives. Furthermore, the extent of these false positives and negatives may be difficult to determine: in some cases, differential diagnoses may be made after referral to specialists. In other cases, the extent of true and false positive and negative cases may only be determined when (if) the disease phenotype has emerged.

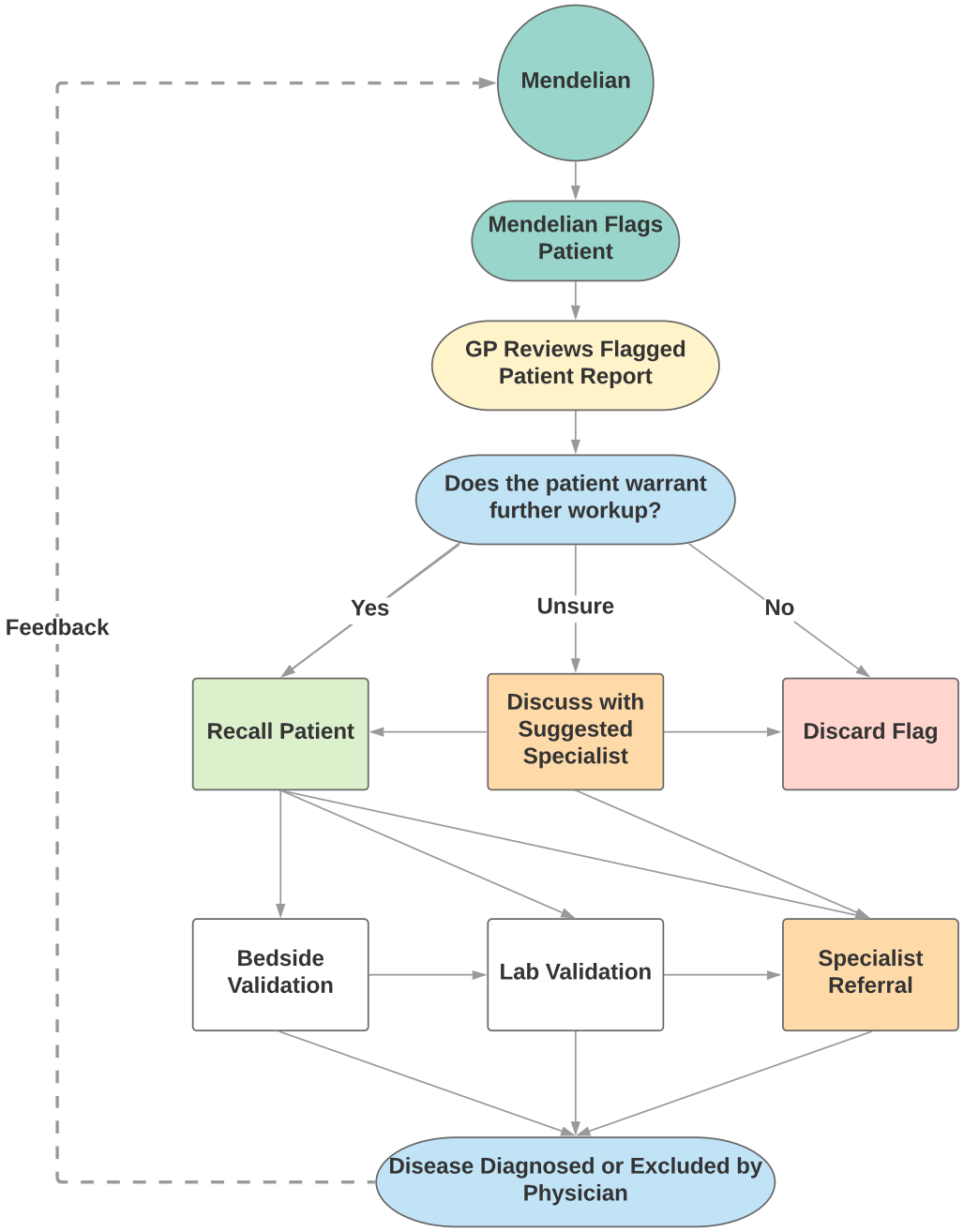

Figure 3: Potential next steps following a patient report

The ethical issues that arise depend in part on what the status of this activity - (screening/case finding/diagnostic test) - is deemed to be. This will also impact on the governance of the test.

4.1 Screening

Definition of screening

Screening) represents:

The systematic application of a test to detect early disease or risk factors for disease. Screening is defined as follows:

"Medical screening is the systematic application of a test or inquiry to identify individuals at sufficient risk of a specific disorder to benefit from further investigation or direct preventive action (these individuals not having sought medical attention on account of symptoms of that disorder)"

4.1.2 Issues for screening for very low population prevalence conditions

The National Screening Committee (NSC) is responsible for authorising and delivering screening testing in the UK. This Committee adheres to criteria for appraising the viability, effectiveness and appropriateness of a screening programme[9]. These criteria are based on a seminal paper by Wilson and Junger[10]. This paper uses the term 'selective screening' using the term 'for the screening of selected high-risk groups in the population which may still be large-scale, and can be considered as one form of population screening'. A WHO review of this work has highlighted the need for various additional criteria to be met including scientific evidence of effectiveness, quality assurance and a need to ensure "informed choice, confidentiality and respect for autonomy" [11.] Updated NSC criteria provide for 20 criteria to be met of which 3 are key in this context:

Important health problem as judged by its frequency and/or severity

Historically, screening for very rare inherited diseases have failed to meet these criteria, particularly that 'the condition should be an important health problem as judged by its frequency and/or severity' and 'the epidemiology, incidence, prevalence and natural history of a condition should be understood'. The criteria also require that there should be an effective intervention for patients identified through screening such that intervention at pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care.

Benefits should outweigh harms

In the past, the screening of very large numbers of individuals to identify a single case was felt to be ethically questionable, if the screening intervention involved some degree of harm or risk to the individual being screened. If there is to be routine use of the MendelScan Tool for screening and case ascertainment, there needs to be clarity about how the use of the tool might satisfy these conditions. In general, there should be evidence from high quality randomised controlled trials (RCTs) that the screening programme is effective in reducing mortality or morbidity.

Acceptability of the complete screening programme

To the extent that the use of the tool is regarded as screening, the complete programme should be 'clinically, socially and ethically acceptable to health professionals and the public'. Demonstrating acceptability may be problematic if there is a lack of transparency about how patient records are used.

4.1.3 Conclusions

- The proposed use of the MendelScan Tool to mine EHRs does not meet key criteria for a current population screening programme
- The evidence that might be gained from the proposed implementation of the MendelScan Tool might support a better understanding of the epidemiology of these conditions and demonstrate potential utility. If the use of the Tool cannot yet been regarded as 'screening' then other applicable definitions might apply

4.2 Case finding and opportunistic screening

"Case finding" has been used to describe the prospective targeting of at-risk groups. An example is targeted dementia screening of groups of patients thought to be at higher risk of dementia. However commentators like Margaret McCartney have been critical of the term. The term has its origins in the work by Wilson and Junger (see reference 10) which described case finding as "that form of screening of which the main object is to detect disease and bring patients to treatment, in contrast to epidemiological surveys" which as Margaret McCartney[12] argues – 'simply detect and do not treat risk factors or abnormalities'. There is no consensus definition of case finding. Margaret McCartney argues that in practice, case finding is the term used "when, to all intents and purposes, population screening is taking place but without evidence that would have enabled its approval as a screening programme".

In contrast, in the context of general practice, opportunistic screening has been used to describe when a chance encounter with health professionals has provided an opportunity for other enquiries or interventions which could impact on future care. It has a different meaning in the context of using genomic sequence results for opportunistic genome screening (see section 5)

4.3 Distinguishing between research and clinical care

Another key distinction is the extent to which the use of the MendelScan Tool constitutes research or part of clinical care. The rationale for the project is that use of the MendelScan Tool can offer clinical utility. It is on this basis that the seven rare conditions have been selected. The premise is that at risk individuals are flagged by the algorithm. The patient is de-identified and GP reviews the flagged patient report. The GP then decides whether to recall the patient, discuss with an expert once detected, or discard the flag. If a patient is recalled or an expert consulted, further validation of the patient's findings can occur through 'bedside validation', lab validation or specialist referral (Figure 1).

- On the evidence of deployment set out in the example Deployment Plan it is not easy to distinguish between research and clinical care
- Since the entire GP practice population is being reviewed (excluding some outliers for reasons that they might be easily identified) this activity seems closest to screening rather than to clinical care, but Mendelian should clarify with other participating organisations as to the status of the project. Is it more appropriate to describe it as research or a service deployment?
- This distinction between research and service deployment is critical and impacts on findings set out later in this report

5 Specific ethical issues related to use of genomic data for opportunistic screening

There are no plans for Mendelian to access raw genomic data (for example, in the form of BAM or VCF files), although it is possible that the MendelScan Tool may capture genomic information which forms part of genetic or genomic test reports incorporating diagnoses or phenotypic traits associated with the diseases under review. These data are likely to be available for those individuals who are known cases of the selected diseases, since they will be incorporated within lab reports and may also be mentioned in referral letters and discharge summaries.

6 Policy landscape for development and use of algorithms in health

6.1 The vision for digital, data and technology in health and care

The Government has strongly endorsed investment in digital technologies in order to build robust health and social care services in policy papers such as *The Future of healthcare: our vision for digital, data and technology in health and care*[14]. Alongside this document, NHS Digital published a draft *NHS digital, data and technology standards framework*[15] setting out expectations for use of data, interoperability, design, IT and commercial standards within the NHS.

6.2 Emerging landscape for regulating algorithms and machine learning

The policy landscape for regulating algorithms and machine learning is gaining pace and complexity following intensive scrutiny by statutory authorities. Multiple policy bodies are also showing an interest in how algorithms are regulated.

6.3 **Information Commissioner's Office:** Big data, AI, machine learning and data protection

The discussion paper: *Big data, AI, machine learning and data protection*[16] explains the Information Commisioner's Office (ICO)'s views on big data, AI and machine learning and provides an introduction to the GDPR. In particular, it provides an overview of the debates on algorithmic transparency, notes the flawed nature of 'notice and consent' models (such as those required by the GDPR) and suggests some strategies for achieving this in ways such as algorithmic auditing, utilising Natural Language Generation or interactive visualisation. It also highlights the importance of other strategies such as designated ethics boards and privacy impact assessments.

It makes six key recommendations:

1. Where possible reduce use of personal data through anonymisation
2. Enhance transparency through meaningful privacy notices
3. Embed a privacy impact assessment framework into big data processing activities (see a practical guide in Annex 1)
4. Adopt a privacy by design approach
5. Develop ethical principles to help reinforce key data protection principles
6. Develop auditable machine learning algorithms (see sections 189-197)

6.4 House of Lords Select Committee on AI

The Select Committee on AI published their report *AI in the UK: ready, willing and able?* [17] on 19 April 2018. Although the focus of this report was on artificial intelligence some of the policy developments that have been catalysed by this report have a potential impact on the implementation of algorithms like MendelScan, which are data-driven.

Concern was expressed about the 'current piecemeal approach' taken by NHS Trusts to datasharing, risking 'inadvertent underappreciation of data' (section 301). Building on recommendations from this report, NHS Digital and the National Data Guardian for Health and Care have created a draft public NHS Digital Data and Technology Standards Framework which sets out technical requirements for providers of digital devices to NHS providers. Some of these requirements may be relevant to Mendelian, especially requirements for contractual and data security standards. These standards are not yet in force.

6.5 Centre for Data Ethics and Innovation

Widely cited in the House of Lords Select Committee report, the Centre for Data Ethics and Innovation was created by the Government "to enable and ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies". This advisory body will work with Government, regulators and industry. Its key remit will be to identify policy and regulatory gaps, to create codes of practice and make recommendations to governments and regulators.

6.6 Conclusions

- The landscape for the regulation of algorithms is highly dynamic and increasingly crowded both at national and European level
- Addressing potential bias and the need for transparency is being seen as increasingly important
- In parallel there is a mounting call for the NHS to be able to 'recoup' the value of NHS data that is shared with algorithm developers, suggesting that mechanisms enabling data access are likely to become more bureaucratic, standardised and centralised

7. Guidance for developers

Over the last year, a plethora of guidance has been developed to support the obligations of developers in meeting the ethical and legal requirements for explanation.

7.1 The Code of Conduct for data-driven health and care technology

One of the most important sources of guidance is the *Code of conduct for data-driven health and care technology*[20] which sets out 10 principles for developing AI applications in health. Updated in February 2019, it contains a number of requirements that are relevant to the development of the Mendelian Tool. We understand that NHSX is in the process of developing a Tool Kit to aid in the implementation of the Code which includes guidance for meeting the requirements of the Code which will be published in autumn 2019. The Code states an ambition to 'ensure that all those (developing, deploying and using data-driven technologies) abide by the ethical principles for data initiatives developed by the Nuffield Council on Bioethics, namely:

1. Respect for persons – 'which includes recognition of a person's profound moral interest in controlling others' access to and disclosure of information relating to them held in circumstances they regard as confidential'
2. Respect for human rights – including people's basic rights, such as the right to protection of private or family life, including limitations on powers of states to interfere with the privacy of individual citizens
3. Participation of people with morally relevant interests in determining how data is used (or reused) – this participation should involve 'giving and receiving a public account of the reasons for establishing, conducting and participating in a data initiative in a form that is accepted as reasonable by all'
4. Accounting for decisions – this includes both 'structures of accountability that invoke legitimate judicial and political authority, and social accountability'

In their report, *Biological and health data: ethical issues*[21], the Nuffield Council framed these principles in terms of determining the morally reasonable expectations about the governance and use of data, adding that compliance with the law cannot guarantee that a use of data is morally acceptable.

This report goes on to make the following recommendations to researchers:

- That researchers should operate demonstrably within a local governance framework able to maintain reasonable surveillance and thus identifying inappropriate use underpinned by sanctions for misuse (recommendation 12)
- The research community should develop robust and comprehensive, yet efficient privacy protecting rules, guidelines and measures through:
    - Providing greater clarity about the current and future uses of biomedical data together with an acknowledgement that there can be no absolute guarantee of privacy and confidentiality
    - Data controllers must publish information about their approach to data access, transparency and accountability (recommendation 17)

7.2 Is data use consistent with appropriate guidelines?

Principle 3 of the Code concerns the use of data and requires that it is in line with appropriate guidelines for the purpose for which it is being used.

- A crucial question is whether Mendelian is accessing personal data. Is the process of de-identification sufficient to render the data that Mendelian will handle outside the scope of the GDPR?

7.2.1 Is the data 'personal data' as defined by the General Data Protection Regulation?

Personal data is defined in Article 4(1) of the General Data Protection Regulation (GDPR)[22] as:

> 'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

- Is the process of de-identification sufficient to render all the data that Mendelian will handle outside the scope of the GDPR?

7.2.2 Is there a sufficient legal basis for processing under the GDPR/Data Protection Act 2018?

If the data used by Mendelian is insufficiently de-identified and it counts as personal data, those processing data have to demonstrate that they are doing this under an appropriate legal basis. Article 6 sets out a range of different legal bases. NHS organisations (and those involved in delivery of health and social care) will usually rely on public interest (that the processing is necessary for the performance of a task carried out in the public interest (Article 6(1)). However, the full extent of what constitutes public interest has not been fully clarified in the healthcare setting. There is some uncertainty where services are provided by a consortium of public/private providers for example.

Consent is an alternative legal basis but the GDPR defines this in terms which are more onerous i.e. freely given, specific, informed and unambiguous indication of the wishes of the data subject signified by clear affirmative action.

- Using consent as a legal ground for data processing is unlikely to be applicable in the proposed context. However consent may still be important for other reasons, including for disclosure of personal confidential information (see section 9)

7.2.3 Does the anonymisation of data itself require an appropriate legal basis for processing?

One of the enduring questions within the existing regulatory framework is whether consent from the data subject or another legal basis is required to support the anonymisation process (i.e. the removal of identifiers). Debate about this point was the justification for setting up the Health and Social Care Information Centre (now NHS Digital), with the idea that it could act as the hub within the NHS for collecting disparate sources personal patient data, and then de-identifying that data in accordance with the law. The ICO is planning some further guidance on anonymisation, but it is unclear when this will be forthcoming.

7.2.4 Are safeguards for processing 'special category' data in place?

The GDPR provides for an extra layer of protection for the processing of some categories of data. These data include genetic or biometric data used for the purpose of uniquely identifying a natural person. Although the MendelScan Tool will not be accessing raw genetic/genomic data, genetic/genomic information such as a diagnosis or disease name could satisfy this definition. Processing of these data is prohibited unless an exception applies including:

- The data subject has given explicit consent (Article 9(2)(a))
- The processing is necessary for the purposes of …. medical diagnosis, the provision of health or social care or treatment (Article 9(2)(h))
- The processing is necessary for scientific research purposes in accordance with Article 89(1)… (Article 9(2)(j))

The UK has established further safeguards for different types of processing in the Data Protection Act 2018[23].

- A key question for Mendelian will be whether there is any data remaining in pseudonymised patient records which has been insufficiently de-identified to the extent that it remains personal data and thus subject to the requirements of the GPDR and DPA?
- Mendelian will need to seek assurances from the 3rd party provider that no personal data remains and that processing is consistent with industry standards
- Mendelian should see reassurances that the 3rd party provider has taken account of the fact that data might be potentially identifiable depending on context, and the type and extent of other data collected

7.2.5 Have the requirements for providing information, transparency and the right to an explanation in the GDPR been met?

One of the aims of the GDPR is to ensure that data subjects are better informed about how their data is used.

General obligations for transparency (GDPR)

Article 13 (where personal data are collected from the data subject) and Article 14 (where the personal data has not been obtained from the data subject) set out a list of information which must be provided to data subjects, including, in the case of automated decision-making, 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'. There has been considerable debate about what the requirement to provide meaningful information might mean in cases of algorithmic use.

- Does the requirement for provision of meaningful information necessitate disclosing the algorithms or codes supporting automated processing? This might limit how developers can protect their work, and potentially the future sustainability of the product?
- There is not yet consensus about what these obligations mean in practice. In particular, there are debates about the complexity of the explanations that are given to data subjects, and the timing of these. Opinion is split as to whether a general explanation is sufficient, or whether an explanation of a particular decision or outcome needs to be provided

Data processing for research (GDPR)

There is recognition in the GDPR of the key importance of research. Thus Article 89 provides that personal data can be processed in the public interest, scientific or historical research or statistical purposes provided that appropriate safeguards are in place to protect the rights and freedoms of the data subject, such as ensuring that technical and organisational measures are in place (such as data minimisation). Certain rights can be put aside if the rights 'are likely to render impossible or seriously impair the achievement of the (research)'.

- Where research can be done without identifying data subjects (such as using pseudonymised data) the research 'shall be fulfilled in that manner'. (Art. 89(1))
- Whether de-identified (pseudonymised) data falls outside the Regulation, depends very heavily on context. Updated guidance from the Information Commissioner's Office (ICO) on pseudonymised data is awaited.

7.2.6 Data Protection Impact Assessment (DPIA)

The GDPR establishes a mechanism for a data protection impact assessment to be carried out for processing which is likely to result in a high risk to individuals including specified types of processing (section 35 GDPR). The supervisory authority (the Information Commissioner's Office) has created a checklist of types of data processing for which a DPIA is required, where a type of processing is "likely to result in a high risk to the rights and freedoms of natural persons".

The ICO guidance cites European guidance from the Working Party 29 working group on DPIA's[24]. One of the relevant determinants is 'scale' of the processing. This guidance compares a hospital information system which processes patient's genetic and health data (for which a DPIA would be required) and the processing of personal data by an individual physician for which it would not. Examples of whether a DPIA is likely to be required include a hospital processing its patients' genetic and health data. See also ICO Data Protection Impact Assessment checklist[25].

- Is the processing by Mendelian of a sufficient scale to warrant a DPIA being done, should the data count as 'personal data'? Recital 91 provides some guidance including the number of data subjects, volume/range of data items, the duration of data processing activity and its geographical extent. This question should be considered in terms of current plans but also for future roll-out
- If this deployment is seen as 'novel' by the ICO, then this also might trigger the requirement for a DPIA
- There is no obligation to publish a DPIA but controllers might consider publishing parts (e.g. summary or conclusion)

7.3 Be fair, transparent and accountable about what data is being used (Principle 4 - Privacy by design)

Principle 4 of the Code includes a requirement to utilise data protection by design principles, with appropriate data sharing agreements, data flow maps and data protection impact assessments.

ICO guidance[26] sets out requirements for data protection by design and default (Article 25(1) and (2) GDPR respectively) to 'bake in' the requirements of the GDPR into every stage of the data processing pathway through implementation of appropriate technical and organisational measures; and to integrate necessary safeguards.

- This obligation falls onto data controllers and a key question is whether Mendelian is viewed as a data controller or a data processor. Depending on the details of the contracts to use data, they could be viewed as either a data controller or processer. Third party companies used for data de-identification (data processor) can assist with security measures
- 'Appropriate measures' include adopting a privacy-first approach which provides a meaningful choice to individuals over personal data that is processed and providing individuals with a choice as to whether to make personal data publicly available. Guidance from other countries including Canada[27] and Norway[28] may be helpful
- Issues that should be considered include:
    - Minimising the processing of personal data and prompt pseudonymisation
    - Ensuring transparency in respect of the functions and processing of personal data
    - Enabling individuals to monitor the processing
- A key question is whether these requirements apply only to personal data, or whether any commitment to building wider trust and confidence, implies any obligations in respect of de-identified data

7.4 Additional requirements for transparency

The Code puts great emphasis on the need for transparency (principles 6 and 7).

7.4.1 "Be transparent about the limitations of the data used and the algorithms deployed"

Principle 6 requires an assessment of the limitations of the data, its quality and scope. This Principle is highly relevant when thinking about the materials provided to GPs and their patients which describe how the MendelScan Tool works.

- Are the limitations of the data discussed? Have all relevant populations been included in training data? Does limiting the source data to the GP records systematically exclude a type of data that might be relevant in interpreting the data (e.g. imaging data)? What uncertainties might there be in using the MendelScan Tool?

7.4.2 "Show algorithmic type, data use, performance validation and integration"

Principle 7 is a key principle for developers. This has four elements:

I. Show what type of algorithm is being developed or deployed
II. The ethical examination of how the data is used
III. How its performance will be validated
IV. How the algorithm will be integrated into health and care provision

The guidance elaborates this requirement to suggest the following actions:

- Characteristics of the algorithm itself: functionality, learning methodology
- Performance: what are the strengths and limitations of its use?
- Context: show how the decision has been made on the acceptable use of the algorithm in the context it is being used (i.e. by committee, evidence or equivalent)
- Wider system implications: is the algorithm 'ready for deployment or in training'? What are the resource implications?

■ Is Mendelian in a position to answer these questions about the performance of the MendelScan Tool? If not, this suggests that what is being done is research

7.5 National Institute for Health and Care Excellence: Evidence Standards and User Guide for Digital Health Technologies (March 2019)[29, 30]

These evidence standards were developed to support Principle 8 in the Code, (generate evidence of effectiveness for the intended use and value for money) but are also applicable to other principles. They are applicable to digital health technologies that use fixed algorithms but not adaptive algorithms. Applying these standards to the MendelScan Tool suggests the following evidence standards apply:

■ Tools that diagnose a specified condition or guide diagnosis are classed as evidence tier 3B (Figure 1) However, Mendelian have advised that the MendelScan Tool will always be used in conjunction with a health professional. Applying the contextual questions in Table 2 i.e. that the Tool will be used with support from a health care professional suggests that this could be downgraded as the Tool could be 'considered to have lower risk than digital health technologies (DHTs) intended to be used by the patient on their own'

7.6 The Data Ethics Framework: processing data in the public sector

This Data Ethics Framework[31] applies to data processed directly or indirectly in the public sector. The Framework is expressed to build on core values of the Civil Service Code including integrity, honesty, objectivity and impartiality. The Framework consists of three parts – a set of principles, additional guidance for each principle in the framework and a workbook. Many of the principles mirror those included in other guidance. They include:

- Clarity about user need and public benefit (Principle 1)
- Being aware of relevant legislation and codes of practice
- Making work transparent and accountable (Principle 6). The workbook contains some practical advice for how algorithm developers can gain public trust in sharing data and algorithmic models as much as possible and using peer review to get feedback on what has been developed

- Depending on the precise levels of oversight, the proposed data processing could count as indirect data processing in the public sector, and this Framework might apply

8 The impact of the UK Data Protection Act 2018

It is not enough just to consider how the GPDR impacts on data processing. The Data Protection Act 2018 (DPA) creates a comprehensive framework for general data processing by filling the legal spaces allowed by the GDPR standards, and establishing UK specific safeguards and exemptions where the GDPR allows. It also provides for continued application of GDPR standards after the UK leaves the EU. Until then the GDPR applies directly. The ICO provide a good overview of the Data Protection Act 2018[32].

8.1 Key provisions providing exceptions and safeguards to processing

The DPA runs to over 350 pages, but a number of provisions are directly relevant to the proposed use of the MendelScan Tool.

8.1.1 Public interest

- The Act provides more detailed consideration of when special categories processing might apply. For example processing for health and social care, public health or for research must meet a condition in Part 1 of Schedule 1 DPA
- Processing for health or social care purposes includes medical diagnosis, provision of health or social care or treatment (Schedule 1 Part 1 (2)(2)). Processing for research is covered in a separate section (Schedule 1 Part 1 (4))

8.1.2 Health test

- A data controller may disclose information relating to another individual without consent if the health data test is met – i.e. the information is held in a health record and the other individual is a health professional who has contributed to the health record or is involved in diagnosis, care or treatment of the data subject. This allows sharing of health data between professionals to facilitate care (Schedule 2, Part 3, para 17(2)). It would cover the discussions between GPs and specialists as described in the figure 3 of the Deployment Plan

■ Since Mendelian is not a health professional, this would not cover communication between GPs and Mendelian concerning the operation of the Tool or understanding the findings that have been generated

8.1.3 Research

- Sections 15-16 and 19 of the DPA provide exemptions from data subjects rights where processing is necessary for scientific research. However, certain conditions apply: that the processing does not support measures or decisions affecting particular individuals unless it is for the purposes of approved medical research (section 19); that the data is not processed in ways that will (or will be likely to cause substantial damage or distress to a data subject (section 19) and results of research are not made available in a form which identifies a data subject (Schedule 2, Part 6 para. 27(3)(b))
- 'Approved medical research' is defined in section 19(4) – as research which has approval from a research ethics committee; or another designated body including a relevant NHS body as defined in section 19(4). These safeguards reflect existing mechanisms for health research

■ On this basis, use of the MendelScan Tool would not seem to constitute 'approved medical research' because it has not received approval from a research ethics committee

8.1.4 The offence of re-identification of de-identified personal data

- The Data Protection Act creates a number of specific offences which are relevant to Mendelian's activities. Section 170 creates an offence for a person 'knowingly or recklessly to obtain or disclose personal data without the consent of the controller'. A defence is available if the use can be justified as being in the public interest, or that the person had a reasonable belief that he had a legal right to obtain, disclose, procure or obtain the data
- It is also an offence for a person to knowingly or recklessly re-identify information that is de-identified personal data without the consent of the data controller (section 171). Here – personal data is 'de-identified' if it has been processed in such a manner that it can no longer be attributed, to a specific data subject (section 171(2)(a))

■ If Mendelian's processes enabled identification without adequate consent from the data controller an offence could be committed unless Mendelian could show that this re-identification was in the public interest

8.2 Summary assessment of the impact of the DPA

■ Mendelian should consider the impact of the DPA as part of its future planning. As considerable regulation making power has been reserved to the Secretary of State, Mendelian will need to be mindful of further changes

9 Ethical issues relating to the use of EHRs

Increasing digitisation offers opportunities for more systematic data capture, surveillance, and research to support high quality care and more effective and efficient health services. The systematic use of EHRs is comparatively new, and provides novel ways of improving patient management, increasing diagnostic yield and targeting interventions more effectively. Actionability following diagnosis of a rare disease allows more active and effective surveillance of disease progress, targeted clinical trial recruitment and also improvements in support available to individuals through patient advocacy groups.

The literature relating to the use of EHRs has been dominated by discussions around data access, quality and governance[33]. Currently there is not an extensive literature on the ethics of data mining and case finding, especially the extent to which the use of algorithmic tools might lead to unintended consequences with patient safety implications, although these have been described hypothetically (See for example AOMRC Artificial Intelligence in Healthcare[18]).

In line with the UK's Industrial Strategy, which lists AI as one its four Grand Challenges[34], policy makers are tending to develop practical guides to implementing algorithms in different sectors, rather than systematically reflecting on the ethical challenges involved.

9.1 Precedents for systematic analysis of EHRs in primary care

Although the systematic analysis of primary health care records is quite new, there are two existing precedents which are used by GPs. The first precedent is the electronic Frailty Index (eFI). This index supports GPs in meeting their requirement in the GP Contract to identify all patients aged 65 and over who may be living with moderate or severe frailty, and to undertake specific interventions for those with moderate or severe frailty.

The use of the eFI has targeted evidence based interventions. For those with severe frailty (around 3% of over 65s) requirements are an annual medicines review, a falls risk assessment if clinically appropriate, and promotion of an enriched Summary Care Record (SCR). For those with moderate frailty (around 12% of over 65s), the requirement for GPs under the GP Contract is to consider undertaking a medicines review, a falls risk assessment and promotion of an enriched SCR. The eFI is an automated risk prediction tool which is used to identify those populations of people likely to be living with varying degrees of frailty based on 'deficits' including clinical signs, symptoms, diseases, disabilities and abnormal test values. It is made up of 36 deficits comprising around 2000 Read Codes.

The NHS webpage confirms that the score is strongly predictive of adverse outcomes and has been validated in around 900,000 patient records.

The eFI is not a clinical diagnostic tool but a population risk stratification tool which identifies sub-populations of people who are likely to be living with specified degrees of frailty. Direct clinical assessment and judgement is therefore needed from a member of the clinical team to confirm identification of moderate to severe frailty, particularly as the eFI has relatively high sensitivity and low specificity so tends to over-identify people living with frailty. Ultimately the expectation is that the use of the eFI together with targeted clinical assessment will reduce the incidence of falls and avoid adverse effects of medication which often result in hospital admission.

A second precedent which is relevant to the MendelScan Tool is the precedent of case finding for familial hypercholesterolaemia (FH). NICE published an updated guideline in November 2017 to reflect new evidence concerning the identification of FH in health settings including primary care, cost-effectiveness of genetic-based cascade screening and treatment with high-intensity statins. Those at highest risk of FH typically are missed by routine search tools such as QRISK because affected individuals only have one of the standard cardiovascular disease (CVD) risk factors (i.e. raised cholesterol). As a result, use of these tools focus on those with the highest levels of cholesterol (at the 99.5th percentile in the Health Survey for England 2003-2013)[38].

- Issues such as completeness/continuity of electronic patient record, previous coding history and demographic characteristics may impact on the sensitivity and specificity of the index. In the same way, these issues may impact on the implementation of the MendelScan Tool and should be taken into account by GP providers when piloting, validating and implementing the Tool

9.2 Is the use of EHR without specific consent breaching patient/doctor confidentiality?

A requirement to adhere to lawful data processing means that Mendelian has not only to consider the GPDR, and UK DPA, but also the impact of the common law. The requirement for doctors to respect patient confidences and keep them confidential is a fundamental ethical obligation. This requirement is enshrined in the Hippocratic Oath and in professional guidance across a variety of professions. Confidentiality is also protected as a human right through the European Convention on Human Rights Article 8 which provides for a right to respect for (an individual's) private and family life (now incorporated into English law through the Human Rights Act 1998)[36].

9.2.1 Confidentiality and the common law

There is a common law (i.e. judge made) duty of confidentiality. In a series of recent cases, the court has capitulated on the extent and nature of this duty. It is now generally recognised that an obligation of confidence will arise where the person in question 'had a reasonable expectation of privacy' [37] that is, where the nature of the information itself is private. This will include most information regarding a patient's health kept by their clinicians. Although some legal debate continues, most legal academics believe that the common law duty of confidence is integrated into the GDPR and DPA through the first data principle, namely that 'personal data shall be.. processed lawfully, fairly and in a transparent manner in relation to the data subject' (GDPR Article 5(1)(a)).

9.2.2 There is not an absolute duty to keep data confidential

Confidentiality is an important legal and ethical obligation but it is not absolute. Confidentiality guidance[39] from the General Medical Council (updated on 25 May 2018 to take account of the GDPR and DPA) provides guidance on lawful disclosure. For example, patient information can be disclosed in circumstances where it is not appropriate or practicable to ask for explicit consent (such as where disclosure is mandated by law or would put others at risk of serious harm, or information can be disclosed in the public interest).

9.2.3 Does disclosure of anonymised data breach confidentiality?

There is continued debate about whether disclosures of anonymised patient data are breaches of confidence. In the past the assumption has been that such disclosures are not breaches of confidence provided that the patient is not identified or cannot reasonably be identified from the information i.e. not identifiable. The case of *R v Department of Health, ex parte Source Informatics Ltd (1999) 52 BMLR 65(CA)* confirmed that the disclosure of anonymised information is not a breach of confidence. However debate continues as to whether the unauthorised uses of the information constitute a 'privacy' issue. Legal academics such as Jeffrey Skopek[40] have highlighted the importance of distinguishing between privacy (where we know the subject, but not a fact about a person – the predicate), and anonymity (where we know the predicate but not the identity of the subject). Applying this distinction suggests that there are not potential privacy concerns associated with using the Tool, but there may be concerns about whether the anonymisation process is sufficiently robust.

- There is a lack of clarity about the effectiveness of the anonymisation proposed and whether this complies with both the GDPR and the common law of confidentiality. Mendelian need to be mindful that there are sensitivities and some legal uncertainties in this area

9.3 The NHS Constitution

The NHS Constitution[41] embeds the requirement for patients, families and carers to be 'involved in and consulted on all decisions about their care and treatment'. This includes rights for patients to access their own health records and to be informed about how their information is used. It also enshrines the obligation for NHS staff to keep confidential information safe and secure, to inform patients about their use of confidential information and to record their objections, consent and dissent. Patients also have a right to request that their confidential information is not used beyond their own care and treatment and to have any objections considered.

9.4 Concluding thoughts about confidentiality

- The idea that the reasonable expectations of patients include sharing personal confidential data to provide services is already somewhat entrenched
- This principle has been extended to include the idea that personal confidential data can be disclosed in order to do research which can, in turn, support an evidence based health service. This idea is gaining traction on both academic[42] and policy circles[43]
- This points to a need for Mendelian to engage more explicitly with patients arounds the proposed uses of their data. Do your GP providers have a patient participation group that Mendelian could consult?

9.5 Public trust and confidence

Empirical work on public attitudes to data sharing suggests that the public prefer data to be anonymised but also are generally supportive of data sharing if it supports better, safer care. There is a big policy push (through the Chief Medical Officer's Generation Genome Report[44] and Industrial Strategy[45]) on releasing NHS data to fuel innovation and jobs within the UK, and indirectly to improve patient care and population health. However, a substantial minority remain concerned particularly about the use of their health data for secondary uses. They are particularly sensitive about the use of some types of data (such as genomic data) by certain providers (such as commercial companies) for specific applications (such as insurance).

The One Way Mirror report[46] commissioned by Wellcome Trust/IPSOS Mori explored the factors influencing attitudes toward commercial organisations accessing health, biomedical and genetic data. In sixteen qualitative workshops, the researchers also explored the trade-offs which participants were willing to make between different types of uses. The qualitative work was followed up by a quantitative survey. This research highlighted some conclusions which are relevant to the use of the MendelScan Tool:

- Public views show low understanding of data and healthcare, leading to a general wariness about the idea of commercial access to healthcare data. Some believed that the default option should be that data sharing should be consented
- In particular, individuals believed that they owned their personal individual-level data, even if anonymised
- Genetic data was not known about or understood
- Just 16% were aware that commercial organisations use NHS data, and the commercial sector tended to be mistrusted

- Although public engagement in other contexts suggests that there is broad support for uses of patient data to directly support care, some types of uses (in connection with novel technologies or by commercial companies) are viewed in a potentially more hostile way

9.6 National Patient Data-Opt-Out

In parallel with the implementation of the GDPR, the National Data Guardian for Health and Social Care has recommended that patients be able to opt-out of their personal data being used for secondary purposes (e.g. for research, management). The basis for this recommendation[47] was to increase transparency about how personal data is being used within health and social care, and through this, to improve trust and confidence in health systems and services. Public confidence has been shaken by a series of high-profile data breaches; by security concerns; and the failure of NHS organisations and third parties such as NHS Digital and NHS England to command public trust and confidence. The cumulative effect of incidents such as the failure of the care.data programme and the Cambridge Analytica scandal in which users' profiles were accessed from Facebook without consent, is that use of patient data by commercial companies is seen as being potentially sensitive.

9.7 Model consent and opt-out guidance from NHS Digital

The national data opt-out became operational in England from 25 May 2018, and provides a facility for individuals to opt-out from the use of their data for research or planning purposes. By March 2020 all health and care organisations will be required to have applied these preferences 'in all research and planning situations in which confidential patient information is used'. This data opt-out replaces a previous 'type-2' opt-out which prevented patient confidential patient information to be used for purposes beyond 'direct care' by limited its transfer to third parties by NHS Digital. The default will be that citizens allow their data to be used for planning and research. In order to opt-out, citizens will be asked to give a yes/no response to the statement – "I allow my confidential patient information to be used for research and planning".

NHS Digital have produced a series of leaflets to support the national data opt-out choice. Updated versions were due to be published in April 2019 but have not yet been published.

- NHS Factsheet – IA Data use and patient choice: this explains the context for sharing personal data for health and care to support care and for research. It provides examples of research uses
- National Data Opt-out: Factsheet 2 – When it applies

9.7.1 Some caveats concerning the scope of the opt-out

Although the opt-out has been in force for almost a year, a number of caveats apply.

The opt-out only applies to confidential patient information

The national patient data opt-out will only apply to confidential patient data that is used for purposes beyond an individual's care and treatment. This is information that meets all these three requirements:

1. Identifiable or likely identifiable (e.g. from other data in possession of recipient)
2. Given in circumstances where the individual is owed a duty of confidence
3. Conveys some information about the physical or mental health or condition of an individual. (Factsheet 3)[48]

9.7.2 What constitutes sufficient anonymisation?

The opt-out will not prevent anonymised data from being used for purposes beyond individual care, where those data are anonymised in line with the ICO code of practice on anonymisation[49]. Since the ICO Code does not cover genetic/genomic data in detail, the approach of the ICO is not entirely clear. Other organisations have published further guidance on what is meant by anonymisation including Understanding Patient Data[50] – a multidisciplinary initiative aiming to improve patient and public understanding of data uses for medical purposes – and the PHG Foundation (*Identification and Genomic Data*). The current test for satisfactory anonymisation is that data controllers mitigate the risk of re-identification until it is remote, but this obligation must take account of the capabilities relevant technologies (i.e. by all means reasonably likely – recital 26 of the GDPR).

- Only around 2.8% of patients have registered an opt-out, so the impact of the opt-out on secondary uses of patient data is difficult to predict. Where NHS organisations operate the opt-out, it will apply at the point at which the proposed use of data changes from individual care to research (Opt-out Operational Guidance[51])
- Where a national data opt-out needs to be applied this means that the entire record associated with that individual must be fully removed from the dataset used for this purpose. Removing identifiers is not sufficient
- What plans do potential GP providers have to implement the opt-out? Will this impact on Mendelian's proposed use of the MendelScan Tool?

9.8 Building trust and confidence

Research on using EHRs for secondary purposes has highlighted the concern that there is potential to create new risks and exacerbate others. Stockdale et al in their review "*Giving something back*" [52] highlight that potential sources of distrust arise both from a party's competence to ensure data security (including the ability of the NHS to guarantee security), but also distrust of the motivations underlying requests for access. In some contexts, support for providing commercial access to patient data is finely balanced.

9.8.1 Data sharing agreements

Where personal data is provided by one organisation to another, a data sharing agreement should be put in place to confirm who maintains responsibility and control over the data (i.e. who is defined as data controller and has responsibility for data protection compliance). The ICO has published generic guidance – Data sharing code of practice[53] – including guidance on data sharing agreements and their contents (see section 1) but this has not yet been updated since the DPA became law.

9.8.2 Complying with relevant guidance

Other policy initiatives which could assist in building good practice include use of the NHS Digital Data Access Request Service (DARS) which requires compliance with strict information governance protocols and adherence to a data sharing agreement.

There are various sources of professional guidance which might apply to the use of GP electronic patient records. The *Good Practice Guidelines for GP electronic patient records* were updated in 2011[54]. These Guidelines suggest using a series of techniques to effectively de-identify patient records used for research – including removal of identifiers; pseudonymisation; aggregation; using derivation or banding and shuffling (creating synthetic data). They also recommend that the practice should have the opportunity to review the data before it leaves the practice to allow checks for:

- Incomplete or inaccurate data
- To check compliance with requests from patients for withholding patient information
- To check inclusion of information that is confidential to another person (e.g. a family member)

- To what extent do GPs within the practices have an opportunity to vet uploaded data for incomplete/inaccurate data; national data opt-out compliance; the potential implications for other people (e.g. family members) in accordance with GP EHR Guidelines?

The GMC's Confidentiality Guidance[39] reiterates that patients have a right to expect their doctors will hold information about them in confidence. The GMC requires doctors to make information available to patients about disclosures of personal information for their own care. Consent to this is often implied in the absence of objection. Other disclosures need to be within the reasonable expectations of patients. As we have seen, doctors must obtain patients' express consent to the disclosure of identifiable information for purposes other than provision of care unless the disclosure is required by law or justified in the public interest.

- Are GPs compliant with relevant aspects of this guidance?

9.9 Professional guidance

Adopting strategies such as these, together with strictly controlling data access and minimising data release may help to mitigate the chances of re-identification but are not full proof. This is because the definition of data as anonymised is contingent on the extent of integration with other datasets.

- The legal status of de-identified (pseudonymised) data is unclear under the GDPR and in other policy guidance
- Mendelian therefore needs to keep the status of these pseudonymised datasets under constant review, and take all reasonable steps to ensure that the risks of re-identification are low, and that the processes have the trust and confidence of all stakeholders

10 Clinical risk management and harmonisation of standards

Specialised requirements apply to the application of clinical risk management by health organisations responsible for deployment, use, maintenance or decommissioning of health IT systems. The new virtual department NHSX is also set to mandate the use of internationally-recognised technology and data standards across the NHS to ensure a consistent language of clinical terms to help health care professionals share information, and patient records, seamlessly across health and care settings improving patient safety[55]. These are not explored in detail in this report[56].

11 Ethical issues associated with sites

The proposed use of the MendelScan Tool raises a number of questions about how Mendelian will engage with current and prospective test sites.

11.1 Engagement with HCPs at GP practices

Requirements for good practice for doctors and other health professionals impose obligations for doctors to understand the risks, benefits and uncertainties related to what is being proposed. Questions that might arise include:

- To what extent has Mendelian engaged with doctors and other health care professionals in each of the practices concerned? Do the staff understand the key elements of the pathway and the likely benefits and harms to patients at each step?
- These might include the following:
    - The risks associated with the de-identification process which involves third party contractors having access to personal data, and then passing de-identified data to Mendelian, and the reverse re-identification process, could patients be identified by staff? This might be mitigated by having contractual terms enforcing sanctions for deliberate re-identification and/or disclosure
    - The selection of the conditions screened for? Has the justification for selection of these conditions been communicated? Have the potential restrictions on access to treatment or management through applicable NICE guidance been communicated (e.g. that treatment is contingent on meeting phenotypic requirements – such as the 60 metre walk test which is used as a diagnostic measure in Duchenne Muscular Dystrophy?)
    - The fact that algorithms will be used and the issues associated with these techniques? This could include the confidence intervals that are used in setting criteria for patient records that are selected
    - The training and validation methods that have been used internally in developing the tool? For example, is the tool likely to have foreseeable biases? Has it been trained on a population similar to the potential trial patient population (in ethnic origin for example)?
    - The likelihood that a condition will be found within that population of patients (taking into account particular risk factors within those groups)
- Health professionals will need to understand the significance of a patient being flagged. If a patient is flagged, what are the confidence intervals that have been used? Do all the diagnostic criteria have to be present for a patient to be flagged? Are the criteria matched appropriately for patient's age and sex to take account of criteria which may not yet have developed?
- The implications of a potential diagnosis in one family member for other family members (depending on the inheritance of the disease)
- Have meetings between health care professionals and Mendelian been held? Is there written proof of what was communicated (such as information sheets and presentation packs)? Are these available for other staff members who were not available for a formal session? If not, how can staff at the practices get more information about what is being proposed?

- How is consent from the practices to working with Mendelian evidenced? Have GPs at each of the practices been assigned as proxies for staff in each of the practices? Are there internal mechanisms within the practices to allow this delegation of responsibility?
- Do the doctors have sufficient knowledge to be able to engage with patients in the practice? Has a particular point of contact been identified should queries or questions arise about what is being proposed?

11.2 Wider engagement of GP Providers with patients

As we have seen from the sections related to data protection, to confidentiality and to meeting developing principles for algorithmic applications within public sector organisations, increasing onus is being put on a requirement for transparency. This might include a requirement for potential providers to be explicit that they are working with Mendelian to test the performance of this tool; that electronical patient records will be accessed by Mendelian (in de-identified form) and that initial findings will be fed back to GPs and may be discussed with patients.

The following issues may be relevant:

- Has this group of practices experiences of developing learning healthcare system approaches that might impact directly on the care of patients?
- Are there any distinctive factors about this group of practices (in terms of risk factors for each of the diseases in question) that might influence the prevalence and the incidence of the diseases being looked for? This could include ethnicity; demographic profile; known cases of any of the diseases where there is a possibility of cascading out results from family members); socio-economic factors etc
- What other ways do GP practices routinely communicate with their patients about the collaborations and projects that they are involved with? Does this include newsletters to patients (electronic or on paper), reports in local community websites or community newsletters?
- Are there posters in waiting rooms, informing patients that their electronic notes will be shared with collaborators, giving information about relevant projects?
- Are any of these communications targeted at known vulnerable groups who are likely to be affected? (e.g. those people who do not have access to digital resources, or are in lower socio-economic groups)

12 Reporting results to GPs

The Deployment Plan sets out the process by which flagged patient reports are re-identified by a third party contractor providing a data processing system and sent to the patient's GP (see figure 2 above). Figure 3 of the Deployment Plan also sets out the potential next steps following a patient report.

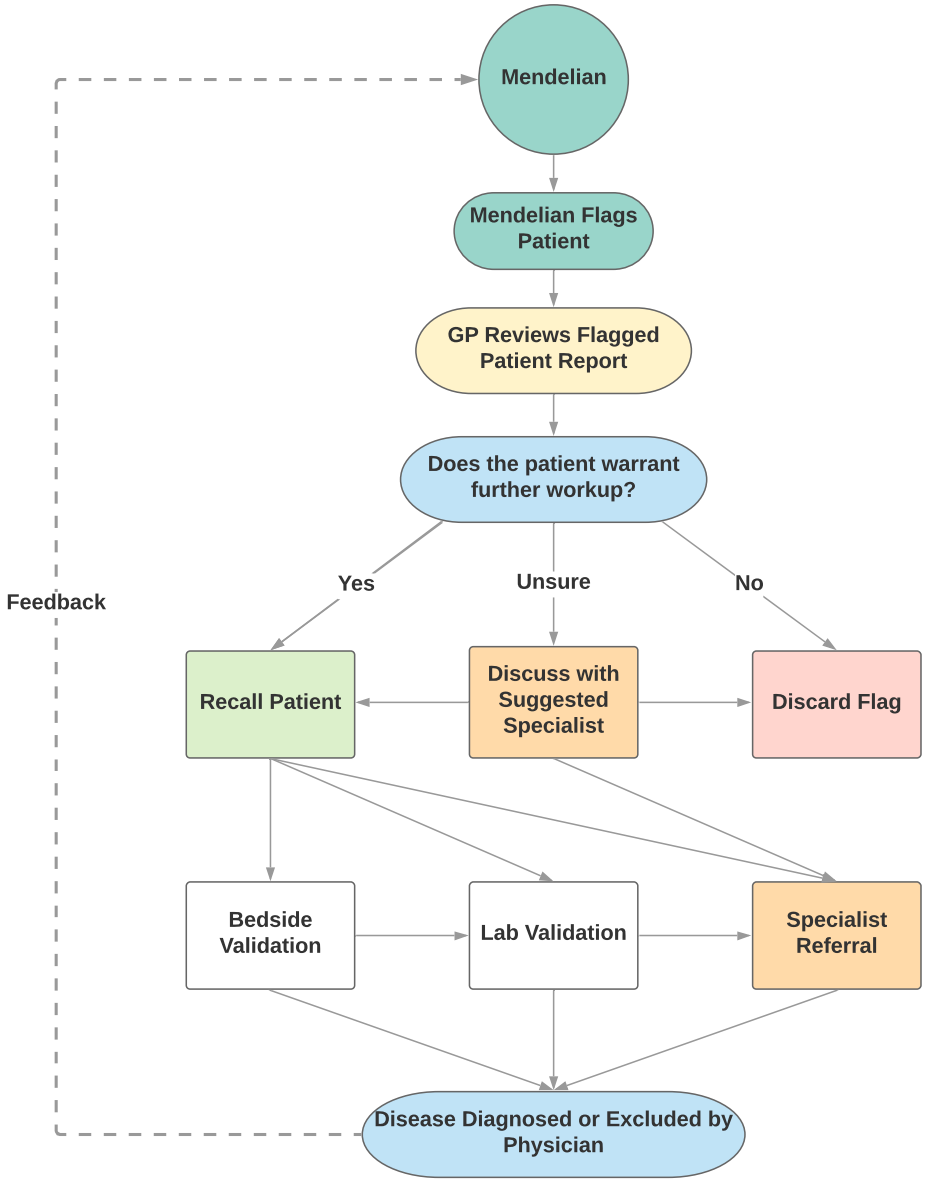

Figure 3: Potential next steps following a patient report

Although GPs are increasingly being expected to understand the clinical criteria for ordering a genetic test, and the implications for the patient when a result is returned, many GPs have not had much training on genetics and genomics.

12.1 Expertise of GPs in genetics and genomics

Expertise typically resides in a few GPs who act as clinical champions. GPs should be expected to have the following skills:

- To understand risk factors
- To understand when a genetic or genetic test might be appropriate for a specific patient
- To understand the significance of a test result as part of a wider clinical picture
- To understand the implications for ongoing treatment of the patient
- To understand significance for other family members
- To understand the implications for future care and treatment

- Will the GPs who receive the de-identified patient reports from Mendelian (via third party contractors to de-identify data):
  - Have experience or training in clinical genetics and genomics?
  - Have access to a clinical champion to discuss best practice?
  - Have the ability to consider differential diagnosis or to take account of phenotypes that are not fully developed in deciding whether to refer the patient to specialist services?
- Might GPs who are unfamiliar with genomic data under or overestimate the severity of the conditions and misunderstand the magnitude of the risks?

12.2 Expertise of GPs in understanding algorithms

Lack of knowledge about genetics and genomics could be compounded by a doctor's ignorance about how the algorithm works.

- How will doctors explain the reasons and processes underlying the MendelScan Tool to patients?

12.3 The format of the report provided to GPs

Mainstreaming genetic and genomic tests to non-genetics professionals has highlighted many difficulties in communicating the results of a test. Even the terminology associated with a test (i.e. 'positive' and 'negative') may be open to misinterpretation. The PHG Foundation held a workshop to explore these challenges associated with the mainstreaming of ordering and receiving genetic and genomic tests, which involved multiple clinical professionals[57]. These aspects are also relevant when considering the communication of test results to patients, and indeed it is usual practice for reports from other clinical teams to be copied directly to the patient.

Various groups have analysed how the form of the report influences how it is understood by health professionals and by patients. Important factors include the terminology used; the significance of the findings for the patient's current health and in the future; the need for additional phenotypic tests to confirm or refute a purported diagnosis and the need for additional tests (such as Sanger sequencing) to validate the genetic/genomic findings. Groups such as the Winton Centre for Risk and Evidence Communication are working to improve communication best practice with the Sanger Institute, the NHS and the Association for Clinical Genomic Science[58].

Work such as this has demonstrated the need for findings to be communicated very clearly to health care professionals and to patients, clearly indicating the significance, risks and uncertainties relating to a particular result. This could include highlighting areas of particular uncertainty related to a particular finding and including technical content on a clearly headed technical section of the report.

- What form will the report to GPs take?
- Are the findings clearly expressed? Are any technical terms defined?
- Is the report divided into technical and non-technical elements?
- Does the report indicate that AI algorithms have been used for generating the report and the uncertainties attached to this?
- Depending on whether this is considered to be research, screening or case finding – (and these are not mutually exclusive) - consider an appropriate clarification of the status of the report and its significance (e.g. that the report has been generated as part of a research project).

13 Reporting results to patients

The planned deployment provides for considerable discretion for the GP to review the flagged patient report and make a choice about what action to take (namely, recall the patient, discuss the case with a suggested specialist before recall or discard the flag and take no further action) – Figure 3[59]. There are considerable ethical challenges associated with communicating potentially clinically significant results to patients, particularly if there is a lack of transparency about the fact that (de-identified) patient records are being used in this way.

If there is no communication by GP Providers to patients that this project is underway (irrespective of whether it is research, screening, clinical audit or clinical care), then such use might not be within the reasonable expectation of patients. Empirical research relating to secondary uses of patient data suggests that there are particular sensitivities when data is used for commercial purposes (see section 9.4). As we have seen in section 9.1, there are precedents population risk stratification using EHRs, which could be regarded as similar to the use of the MendelScan Tool.

- Will reporting results be done by the clinician or other health professional?
- What will be said?
- How will issues arising from unknown sensitivity and specificity of the test (i.e. the possibility of false negative and false positive results) be addressed?
- How best might doctors need to explain that the reasons and processes underlying the MendelScan Tool, and the important role of physicians judgement in making a decision about the next steps that should be taken?
- Will patients be given other sources of information/support?
- What materials/discussions are needed to support consent?

14 Materials to support GPs and HCPs in discussions with patients

A variety of materials are being developed to support discussions with patients prior to referral for further investigation and genetic/genomic testing. The Joint Committee on Genomics in Medicine has published guidance on Consent and confidentiality in clinical genetics practice which includes information on the elements which should be included in a consent discussion and a draft form.

Additional materials are available from the Genomics Education Programme[60].

- **Returning results from the 100,000 Genomes Project**

15 Wider policy implications for use

Use of the MendelScan Tool has great potential for earlier case detection, and prompt identification of at-risk patients. However, Mendelian might also wish to consider the longer term implications if the Tool significantly under or over performs in this project.

- Are Mendelian confident that the sensitivity and specificity selected for the MendelScan Tool will avoid GP budgets being put under strain by increasing numbers of referrals to specialists or increasing requirements for ongoing care/treatment of patients within their practices?
- Could GPs feel burdened by the potential scale of undiagnosed patients in their practices?
- Is there sufficient capacity to deal with the patients that might be diagnosed through use of the MendelScan Tool?

16 Wider policy implications for scaling up

In developing their business model, Mendelian may be considering whether the MendelScan Tool can be scaled up, either to include other diseases at the same site, or to be rolled out across a wider geographical area. The following considerations may be relevant:

- Do all the diseases under review have robust guidance associated with case finding, so that health care professionals are clear about future management decisions and patient pathways that might apply?
- For health systems generally are there wider implications of identifying considerable numbers of patients with undiagnosed conditions?
- Existing service specifications for each of the targeted diseases are predicated on specified prevalence and incidence. Is there capacity within the system for increased demand? Are there sufficient staff/clinics to meet increased demand?
- What are the wider consequences of identifying large numbers of currently diagnosed individuals if existing management/affordability is predicated on a certain number of patients being treated and identified, (both from cases and those cases being cascaded out through family members)?
- Does increased use of algorithmic search tools put managed access schemes (such as those for Gaucher disease and for Duchenne Muscular Dystrophy) under increased strain?
- Is this patient population representative of other practices within the NHS? Are there specific groups which might be unrepresented through them being less likely to be registered with a GP, e.g. certain vulnerable groups such as the travellers community, other ethnic groups or young adults?

17 Recommendations for further actions

17.1 Be clear about whether the activity is research, screening or clinical care

The governance of the use of the MendelScan Tool will depend heavily on context, and whether the application is classed as research, screening or part of clinical care. It seems premature to regard this as clinical care in the absence of validated studies showing that this application is safe, effective and has clinical utility for the diseases that have been selected. Similarly, although the objectives of using the Tool are to detect people who are at higher risk, this will be done without their explicit knowledge or consent and has similarities with screening although an evidence base supporting its use is currently lacking. There are clear similarities of the MendelScan Tool with case finding algorithms such as the elderly frailty index which is used for population risk stratification in that both tools compare patient profiles with a suite of phenotypic features. With the eFI, the percentage of the population that is judged as severely frail is estimated to be around 3%. The rarity of the diseases being searched for in the MendelScan case suggests that this activity is closer to case finding than population risk stratification, but this will be heavily dependent on having a more detailed understanding of how the MendelScan Tool works.

17.1.1 Matters for clarification

- It is not clear what thresholds will be used for flagging patients. Will these err towards greater sensitivity or specificity?
- Given that the purpose of this deployment is to test whether cases generated by the MendelScan Tool match existing true positive cases as well as identify new undiagnosed cases, it seems likely that the use of the Tool in early deployments is research and not clinical care
- Once the thresholds for the use of the Tool have been validated in this population, and can be used reliably for case finding, then it seems more likely that deployment could be viewed as clinical care
- Given this conclusion, potential patient populations should be informed that their data will be used in de-identified format for screening/case finding

17.2 Be transparent about what is being proposed

This might include a requirement that GP Providers be explicit with their patients that they are working with Mendelian to test the performance of this Tool; that electronical patient records will be accessed by Mendelian (in de-identified form) and that initial findings will be fed back to GPs and may be discussed with patients.

This engagement could be done in a variety of ways, either directly via the provider websites, or via patient newsletters. Some practices also engage with local patients through village newsletters at low cost. Mendelian should consider with GP Providers and other stakeholders as to whether these methods are an effective way of engaging with vulnerable groups e.g. those who are particularly disadvantaged by ethnicity, socio-economic status or other characteristic.

17.3 Consider formalised patient participation

The Nuffield Council report suggests involving those with morally relevant interests to be able to arrive a public set of expectations about how data will be used and in continuing governance and review. (page xxvi Biological and health data: ethical issues, full report). We recommend that Mendelian discuss with GP Providers and other relevant stakeholders how they might engage in more formalised patient participation. The example Deployment Plan envisages this, but details of the proposed engagement are not clear.

18 Conclusion

Use of the MendelScan Tool has great potential for more effective and systematic case finding and diagnosis of rare genetic diseases. Although individually rare, these diseases represent a significant burden on those who are potentially affected. By using a novel algorithm to detect potential patterns of phenotypic and genotypic features, this innovative project represents a key opportunity to identify at-risk individuals at an earlier stage.

This report summarises the potential ethical issues that might arise from the prospective use of GP EHRs to detect at-risk individuals. It addresses the objectives of the project, key design features (such as the diseases being sought, and the nature of the MendelScan Tool) and then goes on to consider ethical challenges relating to relevant legislation (GDPR and Data Protection Act); relevant guidance and codes of conduct especially relating to transparency and explanation; and requirements for confidentiality.

The specific ethical challenges relating to the release of reports to GPs and onward transmission of these findings to patients are explored, and the report finishes with some considerations for Mendelian as they take this work forward.

Endnotes

1. Carter D et al Mendelian Deployment at Hertfordshire CCG Project Plan Version 0.1 (April 2019)
2. 2013/14 NHS Standard Contract for Lysosomal Storage Disorders Service (Children) NHS England E06?S(HSS)/c
3. 2013/4 NHS Standard Contract for Alström syndrome service (all ages) A17/S(HSS)e
4. 2013/14 NHS Standard Contract for Bardet-Biedl syndrome service (all ages) A17/S(HSS)f
5. Forsythe E and Beales P Bardet-Biedl syndrome. European Journal of Human Genetics (2013) 21 8-13. Doi:10.1038/ejhg.2012.115
6. Muscular Dystrophy UK information sheet. http://www.musculardystrophyuk.org/wp-content/uploads/2016/04/Translarna-for-Duchenne1.pdf
7. NICE Managed Access Agreement for nonsense mutation Duchenne muscular dystrophy https://www.nice.org.uk/guidance/hst3/resources/managed-access-agreement-july-2016-pdf-2553024061
8. Knowles J et al (2017) Cascade Screening for Familial Hypercholesterolemia and the Use of Genetic Testing JAMA. 2017;318(4):381-382. Doi:10.1001/jama.2017.8543 https://jamanetwork.com/journals/jama/fullarticle/2645088
9. Public Health England Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme
10. Principles and practice of screening for disease. Wilson JMG, Jungner G. World Health Organization, 1968. http://whqlibdoc.who.int/php/WHO_PHP_34.pdf
11. Andermann A, Ingeborg Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years Bulletin of the World Health Organization2008;86:241-320
12. McCartney M (2014) "Case finding" in dementia is simply screening with no evidence of benefit. BMJ 2014; 349 doi: https://doi.org/10.1136/bmj.g4791 (Published 24 July 2014) https://www.bmj.com/content/349/bmj.g4791
13. Wright C (2013) Policy challenges of clinical genome sequencing. BMJ 2013;347:16845
14. UK Government (2018) The future of healthcare: our vision for digital, data and technology in health and care. https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care
15. NHS Digital (2018) Public NHS Digital Data and Technology Standards Framework. https://hscic.kahootz.com/connect.ti/PubNHSDDTSF/view?objectId=10508852

16. Information Commissioners Office (2017) Big data, artificial intelligence, machine learning and data protection
https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

17. House of Lords. Select Committee on Artificial Intelligence. AI in the UK: ready, willing and able?
https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

18. Academy of Medical Royal Colleges (2019) Artificial intelligence in Healthcare.
https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf

19. Nuffield Council on Bioethics. Artificial intelligence (AI) in healthcare and research.
http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf

20. Department of Health and Social Care (2019) Code of conduct for data-driven health and care technology. Updated 19th February 2019
https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

21. Nuffield Council on Bioethics (2015) The collection, linking and use of data in biomedical research and health care: ethical issues. A guide to the report:
http://nuffieldbioethics.org/wp-content/uploads/Biodata-a-guide-to-the-report-PDF.pdf
Full report: http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf

22. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). See ICO Guide to the General Data Protection Regulation
https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation

23. UK Data Protection Act 2018.
https://www.legislation.gov.uk/ukpga/2018/12/contents

24. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 (wp248rev.01).
https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

25. ICO Data protection impact assessments
https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/

26. ICO Data protection by design and default
https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/

27. Information & Privacy Commissioner of Ontario. Privacy by Design.
https://www.ipc.on.ca/wp-content/uploads/2013/09/pbd-primer.pdf

28. Datatilsynet Software development with Data Protection by Design and Default https://www.datatilsynet.no/en/regulations-and-tools/guidelines/data-protection-by-design-and-by-default/

29. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies (2019) https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf

30. NICE Evidence standards framework for digital health technologies: User guide: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/user-guide.pdf

31. Data Ethics Framework https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

32. DCMS Data Protection Act 2018 Factsheet. General Data Processing (Sections 4-28). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/711181/2018-05-23_Factsheet_2_-_General_processing.pdf

33. McBride, S., Tietze, M., Robichaux, C., Stokes, L., Weber, E., (January 31, 2018) "Identifying and Addressing Ethical Issues with Use of Electronic Health Records" OJIN: The Online Journal of Issues in Nursing Vol. 23, No. 1, Manuscript 5.DOI: 10.3912/OJIN.Vol23No01Man05

34. Department for Business, Energy and Industrial Strategy. Policy Paper. The Grand Challenges. https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges#artificial-intelligence-and-data

35. PHE Screening (14 March 2019). New guidance for AI in screening. https://phescreening.blog.gov.uk/2019/03/14/new-guidance-for-ai-in-screening/

36. The European Court of Human Rights in the case Z v Finland said: "Respecting the confidentiality of health data is a vital principle in the legal systems of all the Contracting Parties to the Convention. It is crucial not only to respect the sense of privacy of a patient but also to preserve his or her confidence in the medical profession and in health services in general".

37. Campbell v Mirror Group (2004) UKHL 22

38. Qureshi S, Quereshi N (2018) Familial hypercholesterolaemia: case finding, assessment and management. Guidelines in practice Cardiovascular.

39. General Medical Council. Good practice in handling patient information (2018). https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/confidentiality

40. Skopek J.M. (2015) Reasonable expectations of anonymity. Va L Rev. 2015 May; 101(3);691-762.

41. Department of Health & Social Care (2015) The NHS Constitution for England. https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england

42. Taylor M and Wilson J (2019) Reasonable Expectations of Privacy and Disclosure of Health Data Medical Law Review. https://academic.oup.com/medlaw/advance-article/doi/10.1093/medlaw/fwz009/5479980

43. Department of Health & Social Care (2018).
https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care

44. Chief Medical Officer. Annual Report of the Chief Medical Officer 2016. Generation Genome
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631043/CMO_annual_report_generation_genome.pdf

45. H.M. Government. Industrial Strategy. Building a Britain fit for the future
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf

46. Wellcome Trust (2016) The One-Way Mirror: Public attitudes to commercial access to health data.
https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

47. National Data Guardian for Health and Care. (2016) Review of Data Security, Consent and Opt-Outs.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

48. NHS Factsheet 3 What data and organisations it applies to.
https://digital.nhs.uk/services/national-data-opt-out-programme/guidance-for-health-and-care-staff

49. ICO Anonymisation: managing data protection risk code of practice
https://ico.org.uk/media/1061/anonymisation-code.pdf

50. Understanding Patient Data. What does 'anonymised' mean?
https://understandingpatientdata.org.uk/what-does-anonymised-mean

51. National data opt-out operational policy guidance

52. Stockdale J et al (2019) "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland Wellcome Open Research
https://wellcomeopenresearch.org/articles/3-6/v2

53. ICO Data sharing code of practice.
https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf

54. The Good Practice Guidelines for GP electronic patient records v.4 (2011)
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215680/dh_125350.pdf

55. http://www.thehtn.co.uk/2019/04/04/nhsx-to-mandate-health-tech-standards/

56. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0160-clinical-risk-management-its-application-in-the-deployment-and-use-of-health-it-systems

57. PHG Foundation (2018) Genomics in mainstream clinical pathways. http://www.phgfoundation.org/report/genomics-mainstream-clinical-pathways

58. https://wintoncentre.maths.cam.ac.uk/what-we-do

59. Mendelian Deployment Report, op. cit. Figure 3 (amended image provided by Mendelian).

60. JCMG Consent and Confidentiality in Clinical Genetic Practice (2nd Edition 2011) https://www.bsgm.org.uk/healthcare-professionals/confidentiality-and-genetic-information/

URLs correct as at 29 May 2019

PHG Foundation was commissioned by Mendelian to produce this report

About the PHG Foundation

PHG Foundation is a non-profit health policy think tank and a linked exempt charity of the University of Cambridge. The core mission of the PHG Foundation is to 'make science work for health'. Our work has a special focus on how genomics, digital and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and society. Our multidisciplinary team has expertise in biomedical science, law, ethics, regulation, public health and medicine.

By monitoring, analysing and synthesising research findings we help drive the translation of biomedical advances into practical benefits, which has earned us an international reputation for providing comprehensive policy analysis and advice across the public and private sectors.

ww.phgfoundation.org

Author: Alison Hall

Contact: alison.hall@phgfoundation.org

phg foundation
making science work for health