phg
foundation
making science
work for health

Polygenic scores
and clinical utility

UNIVERSITY OF CAMBRIDGE

Authors

Sowmiya Moorthie, Alison Hall, Joanna Janus, Tanya Brigden,
Chantal Babb de Villiers, Laura Blackburn, Emma Johnson, Mark Kroese

Published by PHG Foundation

intelligence@phgfoundation.org

January 2021

URLs were correct as of January 2021

This report can be downloaded from
www.phgfoundation.org

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

UNIVERSITY OF
CAMBRIDGE

Executive summary

It is widely recognised that many diseases have a genetic component, and knowledge gained from testing and analysis of rare high penetrance variants is often used within clinical and public health practice to inform prevention and treatment. As common low penetrance variants associated with diseases have been uncovered, attempts continue to be made to harness this knowledge for improving healthcare. Polygenic scores are the mechanism by which knowledge of common variants can be used within healthcare. They serve as a biomarker to provide an estimate of the genetic liability (risk) for a particular disease, by capturing the aggregated impact of multiple variants associated with disease.

Discussion continues as to whether polygenic scores are a useful biomarker and the readiness of polygenic score analysis for incorporation into clinical and public health practice. This report informs this discussion by providing a background on key topic areas that we consider are relevant to considerations of the clinical utility of polygenic scores as well as providing a balanced exploration of the discussions around the potential utility of polygenic score analysis for healthcare. An understanding of the key issues in these discussions is informative for those developing specific applications as well as those who are considering implementation.

Key findings

Demonstrating value or utility is linked to the intended objectives of a test and is a key part of clinical evaluation. In the context of healthcare associated tests, two main considerations that influence decisions about usefulness are value judgements about the information derived from a test and how this information influences further action or considerations. Utility is ultimately a subjective and summative assessment influenced by the context of use and the perspective of the assessor.

There is broad agreement that polygenic scores as a biomarker could have potential in informing clinical and public health practice or personal decision making. The information obtained from polygenic score analysis can be used to inform different healthcare scenarios. However, a judgement on clinical utility requires a clear understanding of what this information means, the context of its use, and how this fits with regulatory requirements.

Polygenic scores provide a range of probabilistic risks, similar to other biomarkers such as cholesterol and blood pressure. Risk is not strongly linked with the presence of particular variants and will be significantly modulated by environmental influences. This differentiates them from the information gathered from rare high penetrance variants that have a greater impact on risk. Consequently, the capacity in which polygenic score information is likely to be used will differ substantially for different conditions, depending on the underlying genetic architecture of the disease as well as the degree to which this information can aid clinical and public health practice. This means determining whether polygenic scores will be useful requires an understanding of how they can inform specific clinical and public health pathways.

Many different applications have been proposed based on the analysis of polygenic scores. However, there has been relatively little discussion around the nature of the test that will be offered and its intended objectives. Obtaining a polygenic score is reliant on standardised processes for obtaining genotype data followed by application of a polygenic score model to that data to obtain a score. These scores can be interpreted by themselves or form an input for an existing or novel risk prediction model to provide a combined risk estimate along with other risk factors.

Thus polygenic score models can be considered as assays that can form the basis of different tests. Furthermore, depending on how these models are intended to be used, the different elements of genotyping, polygenic-score models and combined risk models could be viewed as generating a discrete test. Alternatively, these elements could form parts of a single test system. Further clarity is needed with respect to these elements and in defining the test and its intended purpose for enabling clinical evaluation and determining the appropriate level of regulatory oversight that is needed.

Clinical utility can be viewed from different perspectives: public health, clinical, personal or social and is intricately linked with the purpose and context of testing. Personal utility has been put forward by many as a separate concept, in the context of genomics and is often differentiated from clinical utility. Our conclusion is that these are interlinked concepts. Decision makers may place differing emphasis or weighting on various aspects of clinical utility (e.g. clinical and personal outcomes, cost-effectiveness, feasibility and test delivery) which can lead to contrasting decisions on clinical utility when provided with the same evidence for a particular test. Furthermore, decisions with respect to clinical implementation are likely to be influenced by a broader set of factors including other competing priorities and availability of resources. Similar to other tests, the context of use of polygenic score-based tests will influence the required evidence base and the thresholds that need to be met.

Our analysis indicates that thus far much of the debate with respect to polygenic scores has not openly acknowledged the presence of these differing perspectives nor the subjective nature of conclusions on clinical utility. Lack of clarity about the intended purpose of polygenic scores analysis for specific traits and the expected outcomes is also adding to unstructured discussions of their clinical utility.

The development of polygenic-score based tests is ongoing, with certain applications holding promise for use within healthcare. However, much of the effort in this field has been in evaluating the performance of polygenic score models. Evaluation is now needed of tests or test systems developed on the basis of such models as well as research into how they might function as part of specific care pathways. This requires clear articulation of proposed applications, specifying the clinical context of use and target population. Such efforts are important in developing evidence towards demonstrating the clinical utility of polygenic-score based tests.

Contents

1   Introduction

1.1   Polygenic scores and clinical utility

Genetic factors influence disease development and therefore can be an informative biomarker in healthcare. This is exemplified by the use of genetic testing to inform care and management of individuals with diseases that have a strong genetic component, such as Mendelian diseases. In such contexts genetic testing, analysis and interpretation is usually focussed on rare variants that have a large impact on disease. Many studies have been attempting to use knowledge of common variants in the form of a polygenic score as a biomarker for risk prediction, stratification and prognosis. Recent advances in the field have fuelled debate on the use of polygenic score analysis and the evidence that is needed for their implementation within clinical and public health practice. Specifically, evidence of utility or usefulness in terms of 'clinical utility' and 'personal utility', along with differing views of the evidence base in relation to these terms are a key feature of discussions about the value of polygenic score analysis.

Critical appraisal of emerging research and evaluation of the evidence base with respect to the use of polygenic score analysis for different healthcare applications requires an understanding of these concepts and the discussions surrounding them. This report examines the key concepts of clinical and personal utility with a view to informing ongoing discussions on the application, implementation and use of polygenic scores as a biomarker in healthcare.

The aim of this report is to provide a balanced exploration of the debate around the potential utility of polygenic score analysis for healthcare. An understanding of the key issues in these discussions is informative for those developing specific applications as well as those who are considering implementation.

1.2   Report structure

This report is structured in two parts: the first part (chapters 2-5), provides a background on key topic areas that we consider are relevant to considerations of the utility of polygenic scores. This includes an overview of the concept of tests, the terms personal and clinical utility and how test evaluation and regulatory processes approach these concepts. The second part (chapter 6), of this report provides a synthesis of our findings with respect to discussions on clinical utility of polygenic scores elicited from reviewing the academic literature, grey literature and from interviews with key individuals in the field of polygenic score research.

1.3   Methodology

Our approach in carrying out this project was to harness our in-house expertise on the topics of genetic test evaluation and regulation and supplement this with a desk-based literature review and discussion with external experts. Through this process we identified key features relevant to considerations of the utility of polygenic scores.

To synthesise the debate on the utility of polygenic scores, we examined published and grey literature on the topics of clinical and personal utility of polygenic scores analysis. We identified and interviewed key individuals using a semi-structured process to further ascertain and understand their perspectives. Whilst not a comprehensive assessment of views, we endeavoured to interview experts with contrasting views on the utility and use of polygenic scores to ensure a clearer understanding of the current debate.

2   Healthcare tests

Application of knowledge from polygenic scores within healthcare necessitates individual level testing. The way in which tests are defined has an impact on their perceived value. In this chapter, we cover what is meant by a test, the need for their regulation and evaluation, and how polygenic scores may fit within these perceptions.

Key points

- Healthcare tests take a variety of forms and vary in their complexity, purpose, providers and users

- A statement of intended purpose has been identified by many as a key parameter of the test development process and is important in assessing their value

- Lack of clarity on the characteristics and purpose of polygenic score-based tests can be a significant limitation to determining the level of evidence required prior to their implementation

2.1   Test categories

Tests in healthcare take a wide variety of forms, and generally involve collecting information to inform a particular question. This can include collection of demographic characteristics, symptoms and signs, physical examination and/or undertaking laboratory or imaging based investigations to gain insights on particular biomarkers. Therefore, testing can be conducted through a variety of mechanisms ranging from questionnaires to laboratory analysis.

Tests are conducted at different time-points and can inform healthcare decision making in a variety of ways. In clinical settings they contribute to informing assessment regarding the probability of a particular diagnosis (the diagnostic process), enabling the clinician to rule in or rule out specific hypothesis about possible diagnoses. This process will involve all available information relevant to a case including clinical features and other test results. Table 1 is an illustration of the commonly stated purpose for different types of tests and their potential utility. The majority of tests are imperfect and do not provide a definitive or conclusive answer. Thus context of use is important in deciding on whether to utilise a particular test and in interpretation of its results.

Healthcare associated tests are obtained and administered from a range of providers within the UK. Though the vast proportion of molecular tests tend to be conducted in laboratories and are provided via the health system, commercial entities or laboratories also provide services directly to individuals. Individuals can also directly access tests that do not require specialised laboratories, such as simple point-of-care tests (e.g. pregnancy tests, HIV antibody testing, glucose monitors) via commercial providers. In general, tests are conducted to obtain information to aid clinical and personal decision making, and can be obtained from a variety of sources.

Table 1: Test categories and their potential utility. Adapted from The Essentials of Diagnostics series: Molecular Diagnostics, AdvaMedDx and DxInsights, 2013[1]

Broad test category	Potential uses	What they do	Potential utility
Screening	Disease risk prediction	Evaluate likelihood of developing a particular condition	Could lead to lifestyle changes or treatment to minimise risk
	Early detection	Identify disease at an early stage	Reduce impact of disease or prevent if amenable to a treatment
Informing diagnosis	Confirmatory diagnosis	Confirm or rule out specific diagnoses	Determine next steps in care
	Staging and prognosis	Determine severity of condition or predicted outcome	Determine treatment decisions
Management	Therapy selection	Predict effectiveness or potential side effects of treatments	Avoid unnecessary treatment
	Monitoring/ treatment assessment	Assess ongoing safety and effectiveness of treatments	Enables timely intervention to adjust or change treatment when necessary

2.2   The test development process

Increased scientific knowledge and the development of a wider array of technological platforms are contributing to the development of a wide variety of tests. The term "diagnostics" is sometimes used as a broad term to refer to this field and to differentiate it from therapeutics development. Just like the drug or therapeutic development process, test development can be a lengthy process[2]. It requires basic research for the development of technology platforms (e.g. sequencing, mass spectrometry, imaging technology) as well as biomarker discovery (e.g. genes, proteins), followed by translation of those biomarkers and discovery technologies into a tool or product. This additional development is needed to ensure that the biomarker and technology platforms meet user needs.

A statement of intended purpose has been identified by many as a key parameter of the test development process. A determination of whether a test has value is linked to the intended objectives of the test and for informing the degree of validation and regulatory oversight that is required[3-4]. The rise of single platforms that can potentially be used for multiple purposes is leading to a greater need to clarify the nature of specific tests and their intended purpose.

2.3   Implementation of tests within health services

Unlike therapy development, test development has no clearly established pathways and processes for integration into health systems. This is in part due to the variety and complexity of tests, their users and providers, but also because development of tests can be in different sectors such as academia, commercial companies and health system laboratories. Nevertheless, the introduction of

novel diagnostics into healthcare pathways, similar to uptake of therapeutics, requires consideration of usefulness, benefits and risks. These are often assessed through evaluation processes and via ensuring compliance with regulatory frameworks. The degree to which these processes are utilised and applied can vary, and are linked to the characteristics of the test, its intended purpose and user group.

The general objective of regulators is to ensure that tests are reliable, replicable and that they are safe. Depending on the test, this is achieved through various degrees of oversight by a range of regulators during the lifecycle of a test (i.e. pre-implementation to post-market surveillance). Chapter 5 provides a fuller discussion of regulation.

Linked to regulation is the process of test evaluation, a technical exercise that enables assessment of the evidence base. This assessment can inform both regulators, as well as other healthcare decision makers (e.g. commissioners of care) of the evidence base with respect to specific tests.

2.4   Distinguishing assays and tests

As described above, ambiguity around the intended purpose may arise when single platforms or technologies can be used for multiple purposes. This is because they may form the basis of a variety of tests that can be used in different healthcare scenarios. The use of different technologies to provide the same applications and in certain instances the conflation of the technology with the application can also create issues for determining intended purpose and therefore the implications for evaluation and regulation[5]. The proposal to make a distinction between the assay and the test was put forward to enable greater clarity and more informative evaluation of genetic tests[6-7], this conceptual framework can also be applied to other molecular tests.

Technically the method used to analyse a substance in a sample is considered the assay. The test is described as the use of that assay:

1. Within a specific disease context
2. In a particular population
3. For a particular purpose

Evaluating an assay can be restricted to validating a particular methodology, e.g. measuring a biomarker. This may demonstrate that a particular method used to analyse a substance is reliable and robust within a range of conditions. These parameters are often described in terms of scientific validity. Evaluating a test is a more complex undertaking that encompasses wider considerations related to the use of the test and its value in addition to its technical performance. Describing the nature of a test as above adds different use case scenarios to assays which can alter their intended purpose (Figure 1). This conceptual framework enables differentiation of these varying purposes, thereby enabling clearer assessment of the implications of an assay versus the test for healthcare pathways.

Figure 1: One assay, multiple tests

This diagram illustrates how a single technology platform – Next Generation Sequencing (NGS), can give rise to multiple assays. Each assay can be used to inform different questions, giving rise to different tests, each with differing purposes.

Technology	Assay	Test
	Germline WGS	Variants: Wide variety Disease: Developmental delay Population: Paediatric Purpose: Diagnosis
NGS	Cancer (somatic) WGS	Variants: Somatic, compared to germline WGS Disease: Cancer Population: Cancer patients Purpose: Prognosis
	Panel sequencing	Variants: CTFR genes Disease: Cystic fibrosis Population: Newborns Purpose: Confirmatory diagnosis

2.5 Polygenic score models, tools and tests

Similar to molecular assays, risk models take a wide variety of forms and can be used to represent a clinical situation. These models provide a set of rules and data that can be applied to individual level data to predict different outcomes. This has led some to refer to them as algorithms or predictive algorithms. To enable ease of use, these models may be transformed into tools. Tools can take a variety of forms, and are the mechanism through which risk models are utilised. They can be considered as mechanisms through which end-users can collate individual level information, apply a predictive model and obtain a particular output. Examples include the QRISK risk calculator for cardiovascular disease which enables application of the QRisk model[8] and the CanRisk Tool for breast cancer which enables use of the BOADICEA model[9]. Tools may take a variety of forms depending on user needs and can potentially be used in different testing scenarios. Therefore, the use of a particular model and its related tool for a particular disease, in a particular population, for a particular purpose can be considered the test[5].

Whilst guidelines exist that provide researchers with criteria on model parameters that need evaluating and assist decision makers on selecting particular models, their integration into care pathways is often ad hoc. Factors contributing to this are the interchangeable use of the terms model and tool, and a lack of clarity in describing their context of use (i.e. defining the test). As stated above, context of use is particularly important in determining intended purpose, which is key for evaluation and regulation. Describing the elements of risk prediction models and tools in terms of assays and tests helps inform evaluation and regulation. Below we describe how these different elements could potentially be described when considering polygenic score analysis.

Polygenic scores can be used as a biomarker to provide an estimate of the genetic liability (risk) for a particular phenotype. They are usually calculated as a weighted sum of a number of risk alleles carried by an individual. Different statistical methods can be used in the selection and weighting of variants to be included in calculating a polygenic score, giving rise to different predictive models. We refer to these as polygenic score models and many have been developed for calculation of polygenic scores for different diseases[10-11].

These models are applied to individual level genotype data to obtain a polygenic score for an individual. In practice, this requires standardised processes for obtaining genotype data followed by application of the model for calculation of the score to that data. Knowledge of the population distribution of this biomarker and its association with particular traits enables interpretation of this information. Thus polygenic score models can be considered as assays that can form the basis of different tests.

Existing risk prediction models that combine information on a variety of risk factors can be adapted to incorporate a polygenic score, or new models can be developed that calculate risk based on a polygenic score and other variables[12-15]. Separate tests may need to be conducted to obtain information on these variables, such as a cholesterol test, a blood pressure test, a family history assessment etc. These models provide a combined risk estimate, thus we refer to them as combined risk models.

- ■ Risk model: Mathematical representation of a clinical situation able to predict different outcomes. Can also be considered as a predictive algorithm. Risk models can be simple or complex

- ■ Tools: Models may be further transformed into tools. Tools can take a variety of forms and are the mechanism through which risk models are utilised to enable individual level testing and risk scoring. Because they have the potential to be used in different contexts, models and tools could be considered as the assay for purposes of evaluation

- ■ Test: The use of a particular model or tool in a specific population for a specific purpose can be described as the test

2.6   Issues in evaluation of models, tools and tests

Similar to other biomarker tests, predictive tests based on polygenic score analysis can be used for a variety of purposes and in different contexts [16-18]. As described above obtaining and utilising a polygenic score requires a number of steps. This includes steps to obtain genotype data and the use of a polygenic model to obtain a score. This score may be further integrated as part of a combined risk model. This means that depending on how polygenic scores are intended to be used, these steps could be viewed as discrete tests or part of a single test system. Further clarity with respect to these elements and in defining the test and its intended purpose is needed for enabling clinical evaluation and determining the appropriate level of regulatory oversight that is needed.

The development of a polygenic score-based test will necessitate development and evaluation of polygenic score models. These models can be used by themselves, incorporated within existing risk models that have been translated into tools, or used in informing the development of novel risk models and tools. Frameworks exist for the evaluation of risk prediction models and cover similar parameters to those applied to diagnostic tests[19]. These same principles can be applied to evaluating polygenic score models, by themselves or as part of wider risk models.

However, model evaluation is only part of the process of developing an implementation ready polygenic score-based test. As described above, this may also require the development of a tool to enable use of risk models. Risk tools take a variety of forms, and combined risk tools usually consist of web-based applications or digital interfaces that enable end users to input their data and receive an output. These may require assessment of alternate factors to model evaluation, for example, the assessment of elements of usability and ensuring risk information is conveyed in an appropriate manner. The extent to which such tools need to be developed may be influenced by requirements of the test and its user.

2.7   Summary

Healthcare tests are conducted for a variety of purposes, can be obtained from a variety of sources and inform both clinical and personal decision making. A determination of whether a test has value is linked to the intended objectives of the test. There can be a lack of clarity as to the intended objectives when a testing technology can be applied in several settings and have different objectives in these settings. Therefore, in determining the value of a test it is important to have clarity about the setting in which it is to be used, including the disease context, population and purpose.

Evaluation and regulation of a polygenic score-based test to determine its value is not a straightforward process and requires clarity with respect to the test. The fact that different elements can contribute towards this test (genotyping, polygenic score model) and that it can form part of a "multi-test" system in the case of a combined risk model can be a source of confusion. Furthermore, each of these elements maybe further packaged in the form of a tool, which has other additional features. Better definition of the test is needed as well as a clearer understanding of which elements contributing to the test have been evaluated and in what form.

Clarity with respect to this will also contribute to demonstrating transparency, which will be vital in developing sufficient trust and confidence in health professionals and lay users that such tests work as intended, with the appropriate degree of specificity and sensitivity and, more broadly, that the algorithms and devices support decisions that are evidence based and unbiased. Under the new incoming EU regulatory framework, it will be necessary to demonstrate this, and also to capture adverse events, across the lifecycle of the tests from inception, in development, and through post-market surveillance (Chapter 5).

3 Concepts of utility in relation to healthcare tests

Clinical and personal utility are concepts that have been widely discussed in relation to polygenic scores. In this chapter we provide an overview of how these concepts are applied in relation to tests.

Key points

- Although discussion around clinical and personal utility discerns a difference between the two concepts, they are interlinked and provide differing perspectives

- The emphasis placed on particular outcomes may differ depending on whether it is a health system or individual perspective that is taken

- Disagreements and challenges arise due to differences in expectations about the purpose of testing, the nature of evidence and the thresholds required to establish that a test fulfils a specific purpose and differing interpretations of clinical utility

3.1 Clinical utility

Any new change of healthcare practice, including the introduction of tests often requires demonstration that they have value. Clinical utility has no singular or agreed definition, and is a broad term that is used to denote usefulness or value. In the context of healthcare associated tests, two main considerations that influence decisions about usefulness are value judgements, firstly on the information obtained from a test and secondly on the value placed on the practical aspects of obtaining this information. This can be viewed from numerous perspectives: public health, clinical, personal or social and is intricately linked with the purpose and context of testing.

From a public health or clinical perspective, the intended purpose of a test can be to address an unmet health need, or offer an advantage over an existing test (e.g. improved diagnostic yield, more accessible, faster and/or cheaper). The benefits of a test may be realised through its influence on one or more of the following parameters; patient outcomes, clinical decision making or workflow.

Value judgements of the information obtained from tests are influenced by consideration of whether the test fulfils its intended purpose in a specific clinical or public health context, and the implications this will have on care pathways and clinical practice. The purpose of tests intended to confirm or rule out the presence of a particular disease is clear, and the direct benefit of this information on patient outcomes is more straightforward to elucidate. However, the vast majority of tests fall out of this scope and often have indirect benefits on patient outcomes, these benefits may be realised more through impact on subsequent decision making than direct patient outcomes. For example, a test which is used to assess the likelihood of disease development in asymptomatic individuals may impact on the preventive options that are considered (Table 1). Therefore within the sphere of health systems, outcomes or definitions within the spectrum of clinical utility may vary from those limited to clinical endpoints such as mortality and morbidity to more encompassing concepts of net benefit that

include emotional and psychological effects[20–21]. The latter are frequently attributed to the concept of personal utility, which is explored further in section 3.2.

In addition to consideration of the value of information from tests, health systems also need to consider mechanisms of obtaining this information. Depending on the assay and the nature of information it provides, the process may be simple, or more complicated. This is due to the impact these parameters have on wider structural requirements for use of the test and interpreting the information provided by it, which may require different levels of specific expertise. Therefore, different factors may come into the assessment of whether a particular test is considered fit for purpose. These can be divided broadly into evaluation of the test's potential to achieve its purpose and the feasibility of test delivery[20].

This has led some to extend the concept of clinical utility to include organisational aspects, cost-effectiveness analysis and the consideration of ethical, legal and social issues. For instance: test results have the potential to raise concerns about stigma, discrimination and privacy. The potential for these harms to arise can influence individual decisions on whether and when to undergo testing. The social outcomes of testing could affect families, communities and society, and need to be considered in order to evaluate the worth of the test more comprehensively.

Health systems generally aim to provide tests that have value in improving health outcomes on a population basis, but the decision to undergo testing is made at an individual level. Individual decisions to undergo testing may be influenced by personal outcomes that manifest outside of a strictly clinical decision making context. Therefore, whilst individuals and the health system might have a shared understanding of the intended purpose of a test, the value they place on the decision making process and how the results might be used, may differ.

3.2   Personal utility

'Personal utility' is a term that has emerged recently and is being used more widely with respect to genetic tests. As with clinical utility, there is no unified definition, but iterations often include what are considered 'personal' rationales and effects of testing. In other words, personal utility concerns the value of the information to the person being tested. Some consider it a separate concept to clinical utility, whilst others consider it as part of the clinical utility spectrum[22-23].

As concepts of health and wellbeing have shifted to become more holistic, and the associated benefits and harms of testing have broadened, personal utility has become the term used to capture value beyond direct, measurable clinical outcomes. It is often seen as complementary to, or encompassed by clinical utility (depending on how broad a definition of clinical utility is adopted), and to cover 'those benefits or harms that are manifested primarily outside of medical contexts'[24].

The term personal utility is regularly applied to outcomes that may have value to the individual but which are perceived as not having strict clinical relevance and therefore not always incorporated into assessment of clinical utility. This includes both health related outcomes such as increased awareness of health risks, or informing reproductive choices, but also extends further to include outcomes

considered to be unrelated to health such as paternity testing, and testing for certain phenotypic traits (e.g. muscle function).

In certain circumstances there is no meaningful distinction between personal and clinical utility, as 'personal' outcomes can have a clinical impact. For example, although there are no therapeutic options for Huntington's disease, receiving a positive or negative test result may still have psychological, social and practical benefits, and still have clinically relevant utility for individuals. However, in other circumstances a distinction is more feasible. For example, with the rise of consumer genomics, personal utility has been used to encompass 'information for information's sake'. This dimension is often used by direct-to-consumer companies which seek to satisfy curiosity or provide entertainment.

Although discussion around clinical and personal utility discerns a difference between the two concepts it is arguable that the distinction is subjective and heavily dependent on context. Therefore, these are interlinked concepts, where the emphasis placed on particular outcomes may differ depending on whether it is a health system or individual perspective that is taken.

3.3   Decision makers interested in clinical and personal utility

A wide range of groups have an interest in clinical and personal utility. These include the individuals undergoing a test, healthcare professionals, healthcare laboratories, advisors to healthcare systems (for example organisations that set out guidelines or assessment units that evaluate the evidence), payers (e.g. individuals, laboratories, hospitals, departments of health) and regulatory bodies involved in approving healthcare products for use either by consumers or health systems.

Each of these groups may place greater emphasis on different elements of clinical utility. Healthcare professionals may assess utility to determine if the test will be of benefit in helping them make a decision about a patient in their care. Healthcare payers may place greater emphasis on determining value based on improvement in the current process, reduction in costs, and if it will answer a question at a cost they are willing and able to pay. Differing decisions on utility can be reached on the same test depending on who is making the value judgements. Such different groups may also adopt different mechanisms to assess the evidence with respect to utility depending on their perspective and needs. In this document we use the term clinical utility to encompass both these concepts.

3.4   Summary

Clinical and personal utility are broad terms used to make an assessment as to the value of tests. The fact that they have no clear singular definition and can be considered in a narrow or broad sense, means that they are often contentious and context dependent concepts. Decision-makers may place differing emphasis on particular aspects or facets of these concepts, depending on the context of use of a test. In this document we use the term clinical utility to encompass both these concepts.

The fact that decisions about clinical utility are tied to value judgements and priorities means that it is an inherently subjective concept. The evaluation processes used in decision making can also impact considerations of clinical utility (Chapter 4). Finally, even when tests are considered to have clinical utility, this may not lead to a decision to implement within a healthcare setting. Decisions with respect to clinical implementation are likely to be influenced by a broader set of factors including other competing priorities, level of impact, infrastructure requirements, opportunity costs and availability of resources.

4 Evaluating clinical utility

In this chapter, we describe the frameworks for test evaluation that may be used by healthcare decision makers and how clinical utility is approached within them. As the aim of this report is to examine the concept of clinical utility with respect to polygenic scores, we have focussed on genetic test evaluation frameworks.

Key points

- Evaluation frameworks are an important component of evidence based decision making and provide a mechanism for systematic evidence appraisal

- The spectrum of clinical utility considered under different frameworks varies and few readily define the evidence required for clinical utility to be satisfied

- On the basis of the test evaluation, different decisions may be reached with respect to clinical utility and/or implementation

---

4.1 What are test evaluation frameworks?

Test evaluation frameworks are a set of procedures or questions that can act as a guide, informing evidence requirements for clinical tests and their implementation. Stakeholders such as test developers or healthcare decision makers can utilise evaluation frameworks to assist in evidence review, clarifying the scope of assessment and the requirements for evidence, to better understand different aspects of test performance[25]. Although evaluation can take place outside of any specified evaluation framework, systematic application of this process enables more objective appraisal of the evidence-base.

Test evaluation can be applied at various stages of the test development process, and can also help to identify evidence gaps and enable key research questions or next steps to be determined. Frameworks can provide a format for evaluators to clarify, organise and critically appraise the existing evidence base relating to a test and may be used to plan evidence generation when gaps are present.

In healthcare settings, evaluation frameworks are an important component of health technology assessment (HTA) and evidence based decision making. The complexity of testing technologies, together with considerations around opportunity cost and financial pressures on healthcare funders, means that mechanisms are needed for healthcare decision makers to assess the evidence base of particular tests. Such systematic evidence appraisal can then inform decision-making with regards to test implementation.

4.2 Diversity in genetic test evaluation frameworks

Evaluation frameworks exist for many types of tests with a large number developed specifically for molecular and genetic tests. This has been driven by technology development and the need to better

understand the implications of novel diagnostics for healthcare. Evaluation frameworks applied to genetic tests do not differ significantly from those applied to other biomarkers, and aim to assess:

- Whether the test can accurately and reliably measure whether a variant is present
- If it accurately measures or predicts the presence, absence or risk of the clinical disorder
- The positive and negative impacts of carrying out the test
- The cost of testing
- The usefulness of the information retrieved from the test

Like other biomarker tests, genetic testing can be carried out for different purposes including for confirmatory diagnosis, prognostication or susceptibility testing. In 2018, Pitini *et al*.[26] published a systematic review of evaluation frameworks developed between 2000 and 2017 for the assessment of genetic tests. This review identified 29 different frameworks originating from several different countries. Many of these are based on the ACCE (Analytic validity, Clinical validity, Clinical utility, Ethical, legal and social implications) framework or the Health Technology Assessment (HTA) process.

The ACCE and HTA describe two differing, but not mutually exclusive, approaches to the evaluation of tests. Many currently available genetic test evaluation frameworks draw or build upon the concepts and processes outlined in these approaches to create frameworks that are adapted for specific circumstances.

Frameworks for evaluation of genetic tests differ in the "type" of genetic tests they can evaluate. This diversity reflects the different considerations required in testing for different types of genetic variation, clinical situations, and the levels of evidence that may be obtainable. There is also variation in the format of frameworks. For example, several frameworks have been developed as a list of questions which are posed to the potential provider of the test and evaluated by an independent board e.g. ACCE framework; others are conducted in a more flexible manner, where evaluation questions are tailored to the tests after an initial review of evidence availability e.g. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative.

In addition, the intended audience can also differ, for example, evaluations may be performed to:

- Guide policy decisions e.g. cancer screening framework
- Provide guidance for patients (amongst others impacted by the test)
- Guide clinician decisions on test usage e.g. EGAPP
- Provide evidence summaries for general use e.g. companion test assessment tool (CAT)

4.3   Overview of the process for demonstrating clinical utility

As described in the previous chapter, the concept of clinical utility is tied to demonstrating that a test meets user needs and its intended purpose. Figure 2 provides an overview of the processes that can ultimately lead to demonstration of clinical utility. Whilst demonstration of clinical utility is often considered the endpoint, the preceding steps do contribute to an overall assessment of usefulness.

For example, analytical validity (the ability of a test to correctly detect an analyte) demonstrates one aspect of overall test performance and it is unlikely that tests that perform poorly would have utility. In addition, assessment of scientific validity (biomarker-disease association) and clinical validity (test performance in a clinical setting) are important stepping stones towards demonstration of clinical utility. This is especially pertinent within the context of health services, where tests that are taken forward are usually those that have sufficient evidence of analytical and clinical validity. There is generally an absence of a set threshold for test performance parameters, as these are influenced by the context and purpose of the test.

Final decisions with respect to clinical utility are usually based on holistic assessment of test performance characteristics, as well as consideration of a host of practical/pragmatic factors such as the impact and consequences of the test use on care pathways. Broad areas for further consideration therefore include the safety, effectiveness and efficacy of tests.

Figure 2: Overview of the processes that can ultimately lead to demonstration of clinical utility

Test performance characteristics in laboratory setting

Test performance characteristics in clinical setting

Implication of test use for management

Analytic validity

Clinical validity

Clinical outcomes
Personal outcomes
Ethical, legal and social issues
Work flow
Feasibility of test delivery
Cost-effectiveness

Clinical utility

Demonstrates assay performance

Ensuring test is identifying outcome of interest

Effectiveness on clinical outcomes, impact on clinical decision making

4.4   Gathering evidence of clinical utility

Whilst clinical utility is recognised as an important component of evaluation frameworks, few readily define the evidence required for clinical utility to be satisfied. This may in part be due to the fact that clinical utility is a broad concept and a summative assessment of the practical aspects of how the test will work and provide benefit within a defined system, together with a judgement on value of the information from the test. For this reason, evidence of clinical utility can relate to the different

elements, may come in different forms (i.e. quantitative or qualitative) and potentially come from a variety of studies, as opposed to a single study. Sources of evidence that are deemed acceptable may vary and range from diagnostic accuracy studies, randomised controlled trials (RCTs) to modelling studies and observational data, depending on the intended purpose and context of testing.

Randomised controlled trials are often considered the gold standard for demonstrating clinical effectiveness and utility. However, as described above they may not capture all the information necessary to make a judgement on clinical utility. Furthermore, it is widely recognised that these may not be appropriate or feasible in all instances. Indeed, where tests do not create novel clinical paradigms, i.e. there is an accepted role for a test, impact on patient outcomes may not need to be demonstrated. However, evaluation will still be necessary in some form and will take a linked evidence approach and not require actual clinical outcome data for the test.

The scope and type of evidence required may also vary for different tests and clinical settings. RCTs in the context of diagnostic genetic tests for rare disorders can raise immense challenges and in some cases can be unfeasible because of difficulties in recruitment and trial design[27-28]. Therefore, in such circumstances smaller studies including case studies or a narrative on the outcomes of a test provided by experts such as laboratory directors, clinicians and research scientists may contribute to evidence assessment of clinical utility. However, where a test results in a change to the existing paradigm or has significant population impact, a fuller consideration of benefits and harms may be needed. For example, screening tests within the context of new screening programmes often require evidence from RCTs. In such instances, evidence relating to impact on patient outcomes, as well as feasibility and acceptability of the test are required.

Data on test performance characteristics can be relatively straightforward to gather. However, gathering evidence or data with respect to impact on clinical outcomes and clinical decision making and therefore clinical utility may be more arduous, and is often a neglected aspect of test development. The lack of data with respect to clinical utility can act as a barrier to adoption and implementation.

4.5   Summary

Evaluation is a technical exercise that enables assessment of the evidence base with respect to particular tests. This assessment can inform a wide variety of stakeholders involved in test development and implementation including: test developers, regulators, and other healthcare decision makers (e.g. commissioners of care) of the evidence base with respect to specific tests. There is variability in the extent to which these different decision makers engage with this process and the methods they use for test evaluation.

Different decisions may be reached with respect to clinical utility and/or implementation based on this process. This is because of the existence of a variety of frameworks, heterogeneity in how clinical utility is defined within them, and the mechanisms by which supporting evidence is collected and assessed. This is unsurprising given the breadth of views on the concept of clinical

utility. Furthermore, the spectrum of clinical utility considered under different frameworks may be influenced by the purpose of the evaluation.

5 Test regulation and clinical utility

This chapter sets out an overview of the regulatory framework applying to assays and tests, to illustrate how regulatory approval contributes to considerations of clinical utility.

Key points

- In order for assays and tests to be used as intended, and placed on the market, they must be shown to be safe and effective by complying with regulatory frameworks

- Safety and effectiveness are also key elements within the concept of clinical utility

- There is overlap between the evidence generated for regulatory processes and health care evaluation, which can lead to misunderstandings about clinical utility

- For novel assays and tests, such as for polygenic scores, it may be challenging to identify the nature, quality and quantity of evidence required to ensure sufficient safety and effectiveness to place a test on the market

- To demonstrate the clinical utility of a polygenic score based test, it is important to understand how it might be regulated in addition to how it might be evaluated

5.1 The types of regulation that may impact on tests

Many types of regulation impact on the development of assays and tests. These can apply to the test or assay itself, the setting in which the test is developed, or the expertise of the user offering or administering the test. The key regulations applying to the development of assays and tests are EU Directives relating to medical devices. The application of these Directives is not straightforward. They can apply both at the level of the technology platform used to support the administration of an assay (for example a genomic sequencing platform), and/or the specific test to be utilised for a specified purpose within a population (for example, a specific genomic test used for diagnosis of a genetic disease in a given population). Meeting these regulatory criteria can be a complex process, and similar to the evaluation process requires test developers to consider the intended purpose and nature of the test. This can lead to similarities in the evidence required for regulatory purposes and test evaluation.

In addition, tests may be subject to different types of regulation and degrees of scrutiny depending on their nature, how they are provided or conducted, and their intended purpose. Complying with regulatory requirements is a prerequisite for a product to be available on the market and implementation within health services. Therefore, it is an important element of test evaluation and feeds into considerations of clinical utility and implementation from a health service perspective.

The complexities of the regulatory environment with respect to medical devices, its overlaps with the test evaluation processes and the fact that it is mandatory, can lead some to conflate regulatory approval with clinical utility. Conversely others may not consider regulatory approval as an aspect of

clinical utility. Much of this is due to the similarities in evidence that is considered by regulators and healthcare evaluators. Furthermore, as evaluation processes are often not mandatory, evidence with respect to the value of a test may often be generated to fulfil regulatory purposes as opposed to test evaluation. Therefore an understanding of regulatory processes and the evidence required can aid in a fuller assessment of clinical utility.

5.2   What is a medical device?

When a product is placed on the market in the EU it must have CE marking, namely accreditation that it has met appropriate levels of quality assurance, manufacturing practice and is fit for its intended purpose. This applies to products that are intended for medical use, classified as medical devices. The term medical device applies to a range of products from plasters to X-ray machines and includes healthcare tests. Devices that are intended to be used for analysis of biological samples (blood, urine, tissue) are further classified as in vitro diagnostic medical devices.

The regulatory landscape relating to medical devices is in flux, however in this report we use the EU regulatory landscape to illustrate how regulatory approval contributes to considerations of clinical utility. On leaving the EU, Great Britain (England, Scotland and Wales) will retain the existing Directives[29-32] and an incoming set of Regulations will apply to Northern Ireland and to the rest of Europe. The incoming set of Regulations (Regulation 2017/745 on medical devices (MDR)[33] and Regulation 2017/746 on in vitro diagnostic medical devices (IVDR)[34] are already directly applicable on European Member States and will come fully into force in Europe in 2021 and 2022, respectively. Great Britain will continue to rely on the existing Directive, supplemented by a bespoke Conformity Assessed mark system that will replace the CE mark[35]. There will be a grace period until 1 July 2023 for reassessment of devices for the UK market that already have an EU CE mark.

The primary way in which a device might qualify as a medical device under the Directive is if it is intended for medical purposes. In the new EU Regulations, the definitions of a medical or in vitro diagnostic medical device via Article 2(1) MDR or Article 2(2) IVDR (Table 2), have been broadened to specifically include genetic tests. Under both sets of legislation, if the manufacturer does not intend the device to be used for medical purposes, it will not qualify as a medical device and will be subject to less robust oversight. Most tests used within health systems will qualify either as medical devices or in vitro diagnostic devices, depending on the manufacturer's intended purpose. However, these regulations can apply at the level of technology platform or at the level of test. Similar to the evaluation process, how a test or device is defined will have an impact on evidence requirements and considerations of its risks.

5.3   Risk classification and clinical utility

The Directives classify devices according to their risk profiles on a sliding scale. The level of evidence required depends upon risk classification. Under existing legislation (the Medical Device Directive), 80% of devices (including tests) are subject to light-touch regulation, reliant on self-certification by the manufacturer with no external independent scrutiny. Under the new Regulations which will become operational in Northern Ireland and Europe, only 20% of devices will be eligible for self-

certification. The degree of scrutiny applied will depend on the nature of the test; how that test will be used (e.g. screening, diagnosis etc.) and the nature of the user.

Table 2 EU Definitions of medical and in vitro medical devices[36]

Medical Device Directive 93/42 (as amended by Directive 2007/47/EC, Article 2(1)(a)(i)) Article 1(2)(a)	' "medical device" means any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of: ■ diagnosis, prevention, monitoring, treatment or alleviation of disease ■ diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap ■ investigation, replacement or modification of the anatomy or of a physiological process ■ control of conception and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means;'
Medical Device Regulation (EU) 2017/745 Article 2(1)	' "medical device" means any instrument, apparatus, appliance, software, implant, reagent,material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: ■ diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease ■ diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability ■ investigation, replacement or modification of the anatomy or of a physiological or pathological process or state ■ providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations and which does not achieve its principal intended action by harmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means…'
In Vitro Diagnostic Device Regulation (EU) 2017/746 Article 2(2)	' "in vitro diagnostic medical device" means any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, piece of equipment, software or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information on one or more of the following: (a) concerning a physiological or pathological process or state (b) concerning congenital physical or mental impairments (c) concerning the predisposition to a medical condition or a disease (d) to determine the safety and compatibility with potential recipients (e) to predict treatment response or reactions (f) to define or monitoring therapeutic measures…'

5.4 Regulation of algorithms and software

The development of a polygenic score-based test is often associated with development and evaluation of polygenic score models and related tools. Depending on their intended use, these may qualify as medical devices or in vitro medical devices. The use of algorithms and software within those tests, models or tools creates additional challenges and may also require developers to demonstrate that specific standards have been met.

One such challenge is deciding what constitutes 'a device' for the purposes of regulation, since software may be regulated as part of a device or as a separate device in its own right. Indeed, the definition of an in vitro medical diagnostic device in the In Vitro Diagnostic Medical Devices Directive (IVDD) uses the words "alone or in combination" [29] indicating that regulation can apply at the level of the individual components, or the whole with a key consideration being the interoperability of these components.

While the Directive does not define 'software' as such, its Annexes have been amended and contain further provisions clarifying how algorithms and software should be validated in accordance with the state of the art taking into account the principles of the development life cycle, risk management, validation and verification[30].

Another challenge is the potential for artificial intelligence or machine learning. Machine learning algorithms recognise and apply patterns in training data to new datasets to generate novel findings. However, the dynamic, highly adaptive nature of such algorithms, means they are often opaque and increasingly intractable to traditional regulatory approaches. Machine learning algorithms are therefore prompting new regulatory approaches such as regulatory sandboxes and utilising synthetic data.

Where interventions include multiple different elements, commercial rights over each element in the form of intellectual property rights or trade secrets may inhibit transparency. The complexity of these systems, together with the cluster of rights over each element creates additional regulatory challenges and uncertainties.

5.5 Evidence required by regulatory bodies

As described above evidence required as part of the regulatory process is not described in terms of clinical utility but duplicates elements of it. Under the existing Medical Device Directive and In Vitro Medical Devices Directive, it will be necessary to demonstrate that a device works as intended, but also to capture adverse events, across its lifecycle, from inception, in development, and through post-market surveillance. This requires a target patient group and intended purpose to be clearly specified and for scientific validity, and analytical and clinical performance to have been clearly demonstrated for that population. It will also require a balancing exercise to have taken place, which weighs the risks associated with the use of the device with the benefits to the patient.

The Directive requires limited evidence that a device operates safely and effectively. In contrast, the new EU Regulations that will be fully implemented in Europe in May 2021 and May 2022 require two types of evidence that seem close to the concept of clinical utility: these are performance evaluation and clinical performance. Performance evaluation of a device is defined as 'a continuous process by which data are assessed and analysed to demonstrate the scientific validity, analytical performance and clinical performance of that device for its intended purpose as stated by the manufacturer' (e.g. IVDR Annex XIII section 1). The depth and extent are required to be 'proportionate and appropriate to the characteristics of the device including the risks, risk class, performance and its intended purpose'.

Clinical performance studies are a subset of performance evaluation studies designed to establish or confirm aspects of device performance that cannot be determined by analytical studies, literature or previous experience of diagnostic testing (IVDR Annex XIII section 2). Reliance on different regulatory frameworks in Great Britain and in Northern Ireland and the rest of the EU is likely to cause divergence in the evidence standards used for the development of devices and tests.

All these types of evidence required for regulatory assessment may help to illustrate facets of clinical utility described in Chapter 3 and 4, but are not synonymous with it, since demonstrating clinical utility is not part of the regulatory process. This means that care is required when examining evidence generated for regulatory purposes in the context of healthcare evaluation and vice-versa. Whilst regulatory approval is a pre-requisite for healthcare implementation, it may not capture all the necessary evidence required for implementation such as measures of clinical validity, impact on clinical decision making and clinical utility.

5.6   Further regulations impacting on test development and use

Additional regulations apply to the setting in which the test is developed, and to where it is performed. Molecular diagnostic tests may be developed by laboratories in the health sector (both private and publicly funded), the academic sector, and by commercial organisations. In the US and the UK, clinical laboratory services are overseen by regulators to ensure accurate, reliable and reproducible testing in clinical laboratories. Reliable and reproducible testing requires a shared understanding of the nature of the evidence required, and how it can be generated. These standards are maintained through adopting standard operating procedures and ensuring personnel accreditation and assessment. The International Organisation for Standardisation (ISO) 15189:2012 Medical Laboratories set out requirements for quality and competence which are a key measure of quality assessment.

The standards applicable to the design and development of algorithms is more varied than those applicable to the development of molecular diagnostic tests within a laboratory environment. Various international standards may apply to different parts of the software development process. For example BS EN 62304:2006[37] Software in medical devices applies where software is a medical device or embedded or part of a medical device and recommends a framework of life cycle processes for the safe design and maintenance of medical device software. Standard EN 82304-1:2017[38] applies to health software and recommends requirements for the safety and security of health software

products. Medical software typically goes through highly structured release and documentation. The introduction of machine learning into health fits poorly with this, and for this reason, various regulators have sought to introduce principle based voluntary codes of practice and evidence standards which attempt to make the development of machine learning in health more transparent and consistent (discussed in section 5.8).

5.7   Offering tests direct-to-consumer

Where medical devices are used outside of health systems, such as for monitoring of chronic diseases, evidence of effectiveness, replicability and validation remain important. If they are used either as a point of care test or as a 'direct-to-consumer' test, more onerous requirements apply. The regulatory framework take account of the lack of expertise of potential users by imposing requirements that are more rigorous than where tests are used by professionals. These include additional requirements for labelling and for user information to ensure the tests are used safely and effectively.

Other tests may not have an intended medical purpose, but be used to support the healthy lifestyle and wellbeing of the user. Such tests will not qualify as medical devices, and will be subject to less onerous regulation for them to be placed on the market.

Even if tests fail to meet high standards of effectiveness required by the regulations, for some consumers, tests with a lower standard of evidence may still offer utility from a personal perspective. In such cases, transparency about the evidence underpinning device development remains vital to develop sufficient trust and confidence in both the device and the process of regulatory oversight.

5.8   Regulation and polygenic scores

The use of polygenic scores as a biomarker requires development in many different areas depending on their intended use. This includes the creation of a model to calculate a polygenic score, and/or one that enables combined risk estimation, incorporation within an existing risk tool or development of a novel risk tool and subsequent use for a specific purpose and population. From a regulatory perspective, each step could be viewed as generating a discrete device, and, depending on how it is intended to be used, its own market authorisation. Alternatively, these elements could, for regulatory purposes, form parts of a single device. If the former approach is adopted, there would still be a need to demonstrate interoperability between the different elements.

Given this regulatory flexibility, for novel assays and tests, such as those for polygenic score analysis, it may often be challenging to identify the nature, quality and quantity of evidence required to ensure sufficient safety and effectiveness for market authorisation. A continuing lack of clarity over many aspects that enable polygenic score analysis, testing and implementation is a continuing barrier to effective regulation.

Key priorities include:

- A clear distinction between models, tools and tests
- Increasing clarity about the relation between these three elements
- A clear understanding of the purpose and population for use

Evidence will need to be generated throughout the lifecycle of the test from inception, through development and post-market surveillance. This will be needed both to demonstrate performance - that the test works as intended - but also to capture adverse events. The organisations responsible for detailed assessments of polygenic score based tests - notified bodies, will need to develop sufficient expertise and capacity to understand the nature and purpose of what is being proposed. Such clarity will also contribute to demonstrating transparency.

This will be vital in developing sufficient trust and confidence in health professionals and lay users, that such tests work as intended, with the appropriate degree of specificity and sensitivity and, more broadly, that the algorithms and tools support decisions that are evidence based and fair. In the longer term, demonstrating compliance with standards for algorithm development that combine polygenic score use with machine learning or artificial intelligence tools will also require compliance with appropriate codes of practice from organisations such as NHSX[39]. These require the algorithm or tool to conform to various principles including demonstrating that it is fair, transparent and accountable, providing limitations of the data and algorithm deployed and showing how the algorithm performance will be validated and integrated into health and care provision. This includes meeting appropriate evidence standards from NICE[40]. If the use of polygenic score analysis becomes more routine, organisations such as NICE may also have a role in developing disease specific operational guidance.

5.9. Summary

The regulatory landscape is very dynamic, in part due to recent changes but also because of uncertainty due to the UK's exit from EU regulatory frameworks. This will necessitate continued reliance on the Medical Devices Directive and the In Vitro Diagnostic Devices Directive which are likely to become redundant in Europe as the different Regulations become embedded. Confusion over which rules apply will aggravate regulatory uncertainty. Nevertheless, tests that are provided within the sphere of healthcare and health systems must comply with existing regulations. Regulations are important in ensuring safety, effectiveness and efficacy of tests and apply at multiple levels. The evidence that is gathered for regulatory approval may touch upon elements of clinical utility, but is unlikely to be sufficient for decision-makers who are considering healthcare implementation, and who may want to consider a wider array of factors. These factors include the status of algorithms or software within the device, the nature of potential users, and the overarching intended use and purpose. Therefore, since regulatory approval does not include a full assessment of the clinical utility of a product, it is necessary but not sufficient for an assay or test to be implemented into healthcare.

6 Perspectives on clinical utility and polygenic scores

There is ongoing debate over the use of polygenic scores as a biomarker within clinical practice. In particular there is a lack of consensus in opinion as to the added value and therefore clinical utility provided by the use of polygenic scores. This chapter is a summary of our findings on discussions of clinical utility and polygenic scores elicited from reviewing the academic literature, grey literature and interviews with key individuals in the field who have expressed contrasting views on the use and clinical utility of polygenic scores (Appendix 1: List of interviewees).

Key points

- There are a number of variable factors that influence discussions on the clinical utility of polygenic scores

- Debate over utility is mostly based on the findings of academic research studies, rather than through any comprehensive assessment or evaluation of specific applications taking into consideration specific care pathways

- When both the specific context of polygenic score use and the perspectives of those contributing to the debate are taken into account, there is in fact broad agreement over areas of likely clinical utility

6.1 The range of perspectives

Opinions as to the clinical utility of polygenic scores cover a broad spectrum, from those who believe there is sufficient evidence of clinical utility for implementation, others who are more cautious but optimistic over future utility, and those that argue that polygenic scores will never have clinical utility.

Whilst differences in opinions are to be expected, and can be informative in raising areas of concern, they also act as a barrier to informed decision making if the background underpinning a perspective and the nuances of the debate are unclear. Therefore, an understanding of the spectrum of views along with areas of disagreement can be informative in understanding and anticipating implementation issues that may arise.

Our research and analysis indicate that while it appears on the surface that experts are taking opposing stances, there is often broad agreement over the areas of potential clinical use for polygenic score analysis. However there are differing views held as to the readiness and suitability for implementation of such analysis within specific healthcare settings.

Factors that influence and lead to varied perspectives to this are opinions on:

- The ability of genomics to contribute to prediction of complex traits
- The relevance and validity of polygenic score research
- Interpretations of clinical utility in relation to polygenic scores

- Considerations on the effectiveness of proposed intervention
- The evidence base required to demonstrate clinical utility in different contexts

The discussion surrounding each of these factors is described in more detail below, followed by examples of how specific applications of polygenic scores are viewed as a result of these factors. Whilst we have discussed each of these aspects separately, it is important to note that they are often interrelated.

Areas of consensus

- Genomic biomarkers including polygenic scores are not intrinsically different to any other type of biomarker

- It is not possible to make any overarching claims about the clinical utility of polygenic scores

- Within a healthcare setting the use of a polygenic score-based test must make a difference to clinical and/or public health decision making to have utility

- Some form of evidence is needed to demonstrate clinical utility

- RCTs may be the ideal way to generate evidence to support clinical utility, but are unlikely to be feasible in all circumstances

- Polygenic scores are more useful alongside other biomarkers

---

6.2   Ability of genomics to contribute to prediction of complex traits

The use of polygenic scores as a predictive biomarker is being explored for different diseases, including cancer[41-42], psychiatric disorders[43-45], metabolic disorders (diabetes[46], obesity[47]), and coronary artery disease[12]. Research is also being conducted to examine the relationship between polygenic scores and traits that may be considered outside of the scope of clinical medicine, including educational attainment, personality and social deprivation[48-50].

The underlying principle across all these areas is the same; that variability between the genomes of individuals contributes to phenotypic differences, therefore genomic information can contribute to trait prediction. However, the extent to which genomic information is able to contribute to trait prediction will differ across traits. This is especially the case for common complex traits where genomics is one of many drivers in trait manifestation. In general, the use of genomics for prediction outside the scope of clinical medicine is more controversial and raises many more technical as well as ethical challenges.

Hype and genomic exceptionalism

Genomic hype and popular misconceptions that genetic information is 'unique' and 'superior' to other types of biological markers were identified as factors influencing the debate. Our interviewees unanimously agreed that genomic biomarkers including polygenic scores are not intrinsically different to any other type of biomarker. However, they all stated that misunderstandings about information

from polygenic scores and therefore its predictive ability at the individual and population level need to be addressed.

There are concerns that regarding genomic information as superior to non-genetic biomarkers may result in polygenic scores also being viewed as superior, leading to inadequate scrutiny of their predictive ability. This together with the misconception described above can lead to unrealistic expectations as to the usefulness of a polygenic score-based test as a predictive tool[51]. Conversely, there is also concern that polygenic scores are scrutinised more closely in comparison to other biomarkers due to their genomic nature, leading to greater expectations with respect to their predictive capabilities in comparison to other biomarkers, creating barriers to their implementation.

Interpretation of information from polygenic scores

A key point raised by proponents about the usefulness of polygenic scores for disease prediction, is the fact that many research studies have shown associations between polygenic scores and disease status. Furthermore, studies have shown that the predictive performance of polygenic scores for certain diseases such as cardiovascular disease or stroke is comparable to other clinical risk factors, such as smoking or cholesterol[12, 52-53]. Those critical of polygenic scores do not tend to disagree about the associations shown, but about the added value this brings to clinical decision making. There are also concerns that the interpretations of these studies, which are largely exploratory in nature, in some cases exaggerate the ability of genomics to contribute to risk prediction of common disease, especially at the individual level[54] and in non-research settings. In addition, there are concerns that misunderstandings in interpretation of information from polygenic scores, especially conflation of this information with that from rare high penetrance variants, is leading to a lack of clarity as to which clinical scenarios they could inform.

In the context of Mendelian diseases, the impact of particular rare high penetrance genetic variants is often deleterious and sufficient by itself for disease manifestation. This means that in this scenario genomic information, by itself, can be highly predictive of disease risk at the individual level. Furthermore, such information is considered to have clinical utility, especially in the context of clinical genomics services where it can have an impact on patient management. The high predictive ability and clinical utility of genomics in this context is also assumed by some, to apply to polygenic scores.

Our examination of the literature and discussions with experts has led us to conclude that the debate on the clinical utility of polygenic scores reflects to some extent the broader debate on the ability of genomic information to usefully contribute to or improve on complex trait prediction, especially where this is already possible on the basis of other variables[55]. Linked to this are differing views as to whether this predictive ability contributes to prediction at the population or individual level.

6.3   Relevance and validity of polygenic score research

As described in previous sections of this report, we consider the process of deriving a score using a model as an assay. The test is the use of that model for a particular disease, in a particular population for a particular purpose. Our broad review of the field identified that research across different disease

areas is at different stages with respect to development of a polygenic score-based test. The focus of current research efforts has been on development of models and examination of their predictive ability for a particular trait. Often these research studies have not been initiated with the specific aim of developing a test to address a clinical requirement, but from the perspective of exploring potential contributions to healthcare. This means that for many proposed polygenic score applications clearly defined tests are yet to be developed and evaluated, and to date there has been very limited implementation of polygenic scores within healthcare systems[18]. To the best of our knowledge, at this point in time there has been no implementation of polygenic score analysis within publicly funded healthcare systems. In the US some private healthcare clinics are planning to offer polygenic score based tests developed by private companies to assess risk of certain disease such as coronary artery disease and type 2 diabetes[56-57]. Such tests are also available directly to consumers, however, there is currently little reliable evidence to support their clinical validity or utility[58].

Model development is a pre-requisite for test development and our analysis indicates that the progress that is being made in this area is often conflated with availability of a validated test. While these are not mutually exclusive processes, the lack of clarity on this issue leads to criticisms of particular modelling approaches being taken as criticism of a test, where in fact it may be more about their suitability for incorporation into a test or as part of a clinical pathway. This can be illustrated by the questions as to how variants are selected for inclusion in models, when the intended use of the model is not clear[59]. For example, variants included in models may account for risk already measured using phenotypic biomarkers, such as sex, BMI, and cholesterol. This could lead to inaccuracies in predictive ability if the score is then used as a component of an existing risk tool that measure these phenotypic biomarkers. Therefore, the predictive modelling approach used, and its validation are important factors that influence views on clinical utility[18,60-61].

The intended clinical purpose of a novel test needs clear articulation to assess its clinical utility. The lack of clinical context in which many polygenic models are developed has raised concerns amongst critics about their relevance for clinical and public health practice. This also raises concerns about their clinical validity. For example, datasets in which polygenic score models are constructed and validated need to be representative of the population in which their use is intended. Differences between the two, as a result of ethnic mix or different health and age profiles can lead to poor performance or inability to generalise the model[62]. In some instances, models have been developed that function best in specific populations, which raises questions as to the ethics of implementing polygenic score models that are not generalisable[63–64].

6.4   Interpretations of clinical utility

Our examination of the literature and discussions with experts revealed differences in interpretation of the concept of clinical utility. Differing emphasis was placed on particular aspects of clinical utility (e.g. practical delivery of a test, clinical outcomes, personal outcomes) and the extent to which these aspects had been considered in formulating their opinions varied.

Notwithstanding the various definitions of clinical utility, there was agreement, that within a healthcare setting the use of a polygenic score-based test must make a difference to clinical and/or

public health decision making to have utility. There was also agreement that this will often require an intervention or alternate care pathways to be available. We identified that where differences in opinion arose, it was about the degree of change in this decision making that needs to occur and therefore the amount of additional benefit provided by information through polygenic score analysis. In addition, there is clearly a tension between what information was considered useful from an individual test user's perspective and that of a healthcare system.

There was also unilateral agreement from those that we interviewed that it is not possible to make any overarching claims about the clinical utility of polygenic scores. It was generally accepted that clinical utility depends on assessing the specific context and pathways of use, which necessarily demands examination of the nature of information provided by individual scores in these contexts. This is true whether the use of these scores was being put forward for personal use via direct to consumer companies or as part of clinical and public health pathways. Furthermore, whilst there are some arguments that information for information's sake is useful, which all polygenic scores could provide, those we interviewed felt that from a health system perspective this alone is insufficient grounds for implementation.

6.5   Effectiveness of proposed intervention

Opinions over the effectiveness of interventions available to individuals also impacted on opinions of the clinical utility of polygenic scores. Interventions for prevention of common disorders can be broadly classified as either medical (e.g. statins or a screening test such as mammography) or focussed on health promotion through addressing modifiable lifestyle factors. The perceived effectiveness of medical interventions tends to be less controversial as they typically undergo strict evaluation, including guidance on the population that would be eligible based on consideration of harms and benefits. Therefore, where polygenic score analysis are developed to inform decision making with respect to use of such therapies they are broadly considered as useful, and the degree to which they inform decision making, for example through changing eligibility for an intervention is the more contentious issue.

The usefulness of using information from polygenic score analysis to encourage lifestyle modification is more controversial. Proponents argue that such analysis is useful in identifying those at higher risk and therefore most likely to benefit from health promotion initiatives. In breast cancer and coronary artery disease it has been shown that those classified as high risk based on polygenic score analysis would benefit the most from reducing their risk by addressing modifiable risk factors[65-66]. There is also the belief by some that knowledge of genetic risk could have utility at an individual level in motivating those at higher risk to adopt a healthy lifestyle to reduce their overall risk. Similarly, awareness of risk could help motivate patients and clinicians to adhere to prescription guidelines, resulting in better outcomes of current treatment regimens[67].

However, knowledge of risk as a mechanism to influence behaviour with respect to common disease prevention has not been convincingly shown to be effective[67-68]. Furthermore, there are inherent difficulties associated with designing and delivering effective health promotion interventions, especially at an individual level[69]. Therefore, in comparison to medical therapies they can be

considered a more complex endeavour, with more varied views as to their effectiveness. This, together with the fact that all individuals benefit from healthy lifestyles, has led to scepticism about the usefulness of polygenic scores with respect to interventions that relate to health promotion.

6.6 Evidence base required to demonstrate utility in the context of prevention

The level of evidence needed to demonstrate utility is a key factor. Those who believe substantial evidence for the use of polygenic scores is required may be more cautious when considering clinical utility, whereas those willing to accept less evidence may be more optimistic. Nevertheless, there is a general acceptance that some form of evidence is needed to demonstrate clinical utility.

Areas where evidence is considered necessary include scientific evidence with respect to the use of polygenic scores in prediction modelling, as discussed previously. In addition, there is broad agreement that following the development of a test, efforts should be made to demonstrate improved efficiency, improved outcomes, and acceptability of interventions by healthcare professionals and individuals. Therefore, context of use also influences the level of evidence required. This is further influenced by considerations of subsequent decision making influenced by inclusion of polygenic score analysis and potential harms and benefits.

Whilst some believe prospective RCTs would be the ideal way to do this, it is agreed that, for applications related to prevention, the size required and the time scales are too long to make this a viable option. Instead the use of smaller trials, pilot studies and modelling outcomes on a larger population have been suggested. Ultimately the importance of the need for evidence in these different areas will depend upon the requirements of the test user or provider.

6.7 The spectrum of potential applications

Views on the ability of polygenic scores to predict different traits, the relevance and quality of available scores, and the way in which clinical utility is defined and evidenced, feed the debate over which potential applications of polygenic score analysis are considered to have the most and least utility. It is important to remember that, to date, a polygenic score-based test has not been implemented in mainstream healthcare in the UK. Therefore, the debate over utility is mostly based on proposals put forward by research studies, rather than through comprehensive assessment or evaluation of a specific application that takes into consideration specific healthcare pathways.

There are many proposed applications of polygenic score analysis that could be considered to be relevant to clinical management, including aiding diagnosis, informing selection of therapeutic interventions, improvement of risk prediction, informing disease screening, and on a personal level, informing life and lifestyle planning. Table 3 illustrates how some of these applications may fit within the broad test categories described earlier in this report (Chapter 2).

Table 3: Spectrum of potential applications of polygenic scores

Broad testing category		Potential use of polygenic scores	Specific example
Screening	Disease risk prediction	Help to predict the likelihood of developing a particular condition	Predicting the likelihood of cardiovascular disease, therefore better targeting the use of preventive measures such as statins and behavioural change[72]
	Early detection	Risk stratify patients to better target and improve the performance and outcomes of subsequent early detection tests	Predicting the age of onset of aggressive prostate cancer to allow age related stratification into potential prostate cancer screening programmes, improving the efficiency and positive predictive value of prostate cancer prediction[41]
Informing diagnosis	Confirmatory diagnosis	To assist in achieving a disease diagnosis	Use in patients with symptoms of diabetes to more accurately predict their subtype and treat accordingly[71]
	Staging and prognosis	Help to predict the severity of a condition or its likely outcome	Use in patients with bipolar disorders to estimate the likelihood that they may develop psychosis[73]
Management	Therapy selection	Help to predict the likelihood of responding to a drug and can inform decisions on treatment options	Use in patients with schizophrenia to predict the most effective treatment option based on polygenic prediction of the underlying molecular pathways[74]

Of all these, the potential use of polygenic score analysis as a tool in supporting diagnosis or the management of patients already diagnosed with a condition appears to be the least controversial. Polygenic scores in this context are typically proposed to help refine a diagnosis or make treatment decisions in symptomatic individuals within a defined clinical pathway, for which the utility has already been demonstrated. For example in the context of diagnosing an individual with suspected diabetes, a polygenic score may be used to help differentiate Type 1 from Type 2 diabetes, as well from forms of monogenic diabetes[70-71]. Therefore, the value of information from polygenic scores in these contexts is in improving the accuracy and/or diagnostic yield, the benefits of which are easier to envisage. Where doubts over utility arise in symptomatic individuals tend to be in relation to whether the performance of a score is adequate. For example in the context of psychiatry the potential utility of using polygenic scores to help aid in diagnoses of complex diseases with diverse symptoms such as schizophrenia is clear, but as yet it is widely acknowledged that the predictive performance of the scores is inadequate for clinical use for a complex disease such as this[18].

The use of polygenic scores in informing disease prevention in asymptomatic individuals is more contentious, largely because as described above, the use of polygenic scores within the context of specific existing or novel prevention pathways has not been fully explored. What discussion there has been focuses on broad application areas such as screening, or broad disease areas such as cardiovascular disease and cancer, rather than examining their value in the form of a test with a clearly articulated intended purpose. This has led to the use of imprecise language and a lack of clarity over exactly what application is being described, which is fuelling confusion and misunderstanding in debates with respect to clinical utility. For example, there has been extensive debate over the value of polygenic scores as a screening tool. Compounding this debate is the fact discussions may relate to the value of screening for a particular disease, whether polygenic scores are considered a useful addition to existing screening programmes or the parameters that a polygenic score-based test must fulfil in a particular screening setting.

Our analysis indicated that much of the disagreement over the clinical utility of polygenic scores appears to arise without explicitly acknowledging specific use case scenarios. Below we describe broad application areas of polygenic score and some of the debate that has arisen around them. Many of these applications may at first glance appear the same, either in terms of their use for a particular disease or for the same application, such as screening, but can be substantially different when viewed from the contexts of particular use-case scenarios.

Early life screening for risk of common disease

Probably the most hotly debated issue is the proposed use of polygenic score analysis in early life to identify those who are potentially at high risk for common diseases, such as coronary artery disease[12], obesity[75] and type 2 diabetes[76]. The fact that polygenic scores can be calculated from birth are characteristics that set them apart from many other biomarkers, and is one of the aspects that some proponents cite as a key way in which they can contribute to risk prediction. It is believed that this characteristic will enable identification of individuals at high risk, early in life before any symptoms develop. This in turn can enable more effective prevention through early interventions such as increased monitoring, programmes for behaviour change and potentially even early therapeutic interventions. For example, in the context of cardiovascular disease, this might mean targeting those at the very highest risk based on their polygenic score to treat them proactively with statins and supported behaviour change prior to the emergence of other risk factors.

The term 'early life' is used to refer to different scenarios from screening newborns to screening at other time points or life stages through childhood and early adulthood. Each of these different time points give rise to different screening scenarios, which will necessitate different considerations of benefits and harms. For example, concerns over the ethics of newborn screening are likely to influence opinions of using polygenic scores at this life stage. These concerns may not apply to screening at later time points such as early adulthood.

Beliefs about the predictive nature of genomics significantly shape the debate in this area. Many are sceptical of the use of polygenic scores alone to predict risk for common disease with a known large environmental component[46,51,55]. The fact that risk can vary as a result of environmental exposure

over time also needs to be considered. There are also questions as to whether targeting interventions to those at highest risk of a disease on the basis of a polygenic score, would be an effective public health strategy, when the majority of disease risk is due to external factors which could be tackled on a population wide level. Finally, as discussed earlier there is general doubt and uncertainty over whether interventions taken would be appropriate and effective, especially where they relate to lifestyle change.

Related to this is the concept of a "genomics first approach", where non-genomic biomarkers are interpreted in light of genomic information. If population-wide whole genome sequencing is routinely implemented, a polygenic score could act as a biomarker that is incorporated into health records, similar to age, cholesterol etc. This information could then be interpreted alongside the use of other biomarkers as and when needed for different conditions. However, some experts we interviewed indicated that whilst this may be a vision this is currently not a possibility, given our limited understanding of the biological role of variants that are included in polygenic scores in disease causal pathways. In addition to the scientific challenges of this approach and for early life screening in general, there are ethical concerns over the collection, storage, and communication of genomic data to otherwise healthy people.

Polygenic scores as part of risk tools

In the near future, indications are that the most likely use of polygenic scores will be alongside other risk factors to provide a combined risk estimate. Risk prediction tools already exist and are widely used in different disease areas to aid clinical decision making, especially in relation to medical therapies. There are a variety of contexts and disease areas where the incorporation of polygenic scores into existing risk tools has been proposed. Incorporating a polygenic score for coronary artery disease into the existing QRISK tool for cardiovascular disease is one of the most well recognised potential applications. It is widely acknowledged that incorporation of polygenic scores can lead to improvements in existing risk tools [12,67]. The main disagreements in this area occur over the extent to which improvements in performance of the tool are needed to justify incorporation of a polygenic score, and the type and amount of evidence required to evaluate this[77,78].

In addition, as discussed earlier, a potential issue identified is that some variants used in polygenic scores may account for risk already measured using phenotypic biomarkers, such as sex, BMI, and cholesterol, leading to inaccuracies in predictive ability. Therefore, the predictive modelling approach used, and its validation are important factors that influence considerations about the incorporation of polygenic scores into existing risk tools.

Polygenic scores for refining risk prediction in the context of Mendelian diseases

The use of polygenic scores to aid in refining risk predictions associated with Mendelian diseases is an application area broadly thought to have clinical benefit[79,80]. In this scenario polygenic scores could be used within the context of clinical genetics services to help provide a more accurate assessment of disease risk. Studies have shown that polygenic scores can help refine risk estimates for Mendelian forms of breast cancer, cardiovascular disease and colorectal cancer[81–84]. Whilst these individuals are usually classified as high risk, there can be substantial variation in the risk profiles of

people presenting at clinical genetics settings. More accurate and refined risk assessment could result in individuals being placed in different risk categories and therefore offered different prevention options[84].

Part of the reason this use is expected to have utility is that genetic risk prediction and counselling for many Mendelian diseases is already established practice. Therefore, similar to their proposed use as part of existing risk tools, it is believed polygenic scores will have utility if they can improve on this process. As with the arguments made with common complex disorders, it is cautioned that more research may be needed to understand the interplay between risk calculated via polygenic scores and via known pathogenic variants[85]. This could aid in improving the accuracy of scores before they can be used in the clinic.

Polygenic scores as a screening stratification tool

The use of polygenic scores to stratify individuals into different risk categories and offer a screening test on the basis of this risk has been proposed[16,86–87]. Risk stratified screening programmes aim to screen those at higher risk sooner and/or more frequently, whilst those at lower risk would be screened later/less frequently. This risk-based approach to screening has been mostly discussed in the context of cancer, where it is anticipated by some to lead to improved benefits and a more efficient programme. This is through more accurately targeting a screening test to those with an increased probability of developing disease, which can lead to increased identification of cases and reductions in overdiagnosis. In this context, polygenic score analysis does not function as the screening test, but rather a mechanism to identify eligible sub-populations for a particular screening test.

Use of polygenic score analysis for enabling stratified screening is an example of an application for which views of clinical utility vary widely depending on specific contextual factors. These include whether the polygenic score is intended for use alongside an existing screening programme or a new one, how well the score performs for a specific trait, the effectiveness of the subsequent screening strategies, availability of interventions, and the infrastructure and costs associated with a specific programme. Those we interviewed felt that it would be easier to demonstrate the clinical utility of polygenic scores in the context of existing screening programmes, as the utility of a screening programme and test is already established. In such use cases, the incorporation of information from polygenic scores is a modification that could lead to a better balance of benefits and efficiency. In addition, the existence of infrastructure for screening is viewed by some as contributing to the clinical utility as the barriers to implementation are much lower than if an entirely new screening programme had to be created.

Breast and colorectal cancer screening are often cited as examples of established screening programmes, where there is general consensus that a move towards a risk-stratified screening programme is likely to lead to improved benefits and a more efficient programme[60,88]. Nevertheless, as much of the current evidence comes from modelling studies some still consider there to be gaps in evidence with respect to clinical utility[89–91]. There are also doubts as to whether such an approach will be acceptable from an individual and societal perspective and the logistical, economical, and

technical challenges associated with such an approach[60,88,92]. Furthermore, studies conducted in the US and Australia suggest that polygenic score-based stratification is currently unlikely to be cost-effective, leading some to postulate that they lack utility[93]. However, this does not exclude the possibility that use of polygenic scores will have utility in future, if parameters of the score generation or the screening programme are altered to make it more cost effective. For example, the score becoming more predictive, the cost of generating the score reducing, or more people participating in screening[94,95].

In comparison to breast and colorectal cancer screening, views with respect to clinical utility of polygenic scores for prostate cancer screening are more varied. This is due to the different proposed use case scenarios for a prostate cancer polygenic score, which are often discussed interchangeably. There is evidence that polygenic score models available for prostate cancer risk stratification perform well in predicting if individuals are more or less likely to develop disease or aggressive forms of the disease[41,86,96]. It has been postulated that they can therefore be used in a similar way to polygenic score models for breast and colorectal cancer, in population stratification.

Currently in the UK no national screening programme for prostate cancer exists, due to lack of a suitable screening test. While the prostate specific antigen (PSA) test is available, it is not deemed accurate enough to be used in a population screening programme in the UK. It has been suggested that the accuracy of the PSA test could be improved if combined with stratification on the basis of polygenic scores. This would require developing the evidence base in support of a novel screening programme for prostate cancer. The lack of clarity on what would constitute the screening test and subsequent interventions or care pathways that would be available for those classified at higher risk, are factors that lead some to consider polygenic scores not to have utility for this use

An alternate use of polygenic scores is in primary care to aid in interpretation of the PSA test or in targeting those who would benefit from it. The PSA test is more accurate in higher risk individuals[97]. However, in this case the score would be used as a triage tool in clinical decision making in those suspected of having prostate cancer. This "screening" use is often conflated with a population-based screening programme.

Glaucoma is another example where a polygenic score model has been developed that performs well in terms of risk stratification, for early identification of at-risk individuals. It has been suggested that this could be used for screening the population for those at-risk of glaucoma[98]. However, there are doubts over the utility of this use case scenario, again because there is insufficient evidence that this would result in better treatment and outcomes for glaucoma patients[79].

6.8   Summary

Our broad examination of the field of polygenic score research and discussions around clinical utility has identified that the discourse is often in relation to the field as a whole with few instances with respect to specific applications or use case scenarios. Context and application are important in considering the utility of information from polygenic scores, much like for other biomarkers. In a

healthcare scenario, most view the ability of a test to improve clinical decision making and patient outcomes as a key requirement of clinical utility.

However, there is recognition that the parameters to achieve this will vary for each specific context of test use. For example, variations in clinical pathways and types of interventions which a polygenic score-based test could influence will impact on acceptable thresholds for test sensitivity and specificity. This has led to the view that it is impossible to say polygenic scores will have utility without considering the context in which they are used and the nature of information they are providing in these scenarios.

Stratification is a process that can be incorporated into population screening programmes or used as part of clinical practice to inform decision making. Polygenic scores have been shown to be useful in risk stratification, however, our analysis indicates that the ability to accurately stratify is often conflated with utility, with different use cases of polygenic scores often discussed as being equivalent. Furthermore, as polygenic scores are a mechanism of stratification, they are assumed to be a screening test, when this may not be the case. In certain contexts, a polygenic score-based test does not function as the screening test, but rather a mechanism to identify eligible sub-populations for a screening test. Lack of clarity on how such tests are viewed and evaluated in the context of existing or novel population-based screening programmes also leads to non-specific discussions with respect to clinical utility in the context of screening.

Also influencing the debate are the individual perspectives of stakeholders who have different backgrounds and different interests in aspects of polygenic scores. Different perspectives are useful and necessary for a rounded assessment of polygenic scores but have on occasion led to confusion and debate at cross purposes. It appears that when both the specific context of polygenic score use and individual perspectives are taken into account, there is in fact broad agreement over areas of likely clinical utility. It has to be taken into consideration that for this report we were only able to interview a limited number of people, all of whom had backgrounds in academic and/or clinical research and some of whom also had commercial interests. Similarly, when gathering additional information on perspectives the majority of sources used were based on academic literature. Whilst these backgrounds and perspectives may represent the majority of those currently developing polygenic scores and who are debating utility, they do not represent all stakeholders, for example patients, healthcare providers and commissioners. In future when making a rounded assessment of perspectives it will be important to consider the views of all stakeholders.

Our analysis indicates that thus far much of the debate with respect to polygenic scores has not openly acknowledged the presence of these differing perspectives nor the subjective nature of conclusions on clinical utility. Lack of clarity about the intended purpose of polygenic score analysis for specific traits and expected outcomes is also adding to unstructured discussions of clinical utility.

7   Key findings

This report has examined the concept of clinical utility of health care tests from the health system and regulatory perspectives. This was followed by an examination of polygenic scores and their potential clinical utility.

Clinical utility and personal utility

Clinical utility has no agreed or singular definition, but is broadly recognised as the value (relevance/ usefulness) of a particular practice in improving health outcomes. In the context of healthcare associated tests, two main considerations that influence decisions about usefulness are value judgements with respect to the information obtained from a test and related to this, the value placed on this information in terms of further action or considerations. Utility is ultimately a subjective and summative assessment influenced by the context of its use and the perspective of the assessor. Judgements of utility can be viewed from different perspectives: public health, clinical, personal or social and is intricately linked with the purpose and context of testing.

Personal utility is a concept that has been put forward by many, especially in the context of genomics. It is often differentiated from clinical utility. Our conclusion is that although discussion around clinical and personal utility discerns a difference between the two notions, these are interlinked concepts, where the emphasis placed on particular outcomes may differ depending on whether it is a health system or individual perspective that is taken. This means that evaluation from each of these perspectives may lead to different conclusions.

Clinical utility in relation to evaluation and regulation of tests

Evidence appraisal is an important component of healthcare decision making and many frameworks exist to assist with systematic appraisal of a healthcare test. Within these frameworks there is heterogeneity in how clinical utility is defined and mechanisms by which supporting evidence is collected and assessed. This is unsurprising given the differing views with respect to the concept of clinical utility. Furthermore, the spectrum of clinical utility considered under different frameworks is influenced by the purpose of the evaluation. This assessment can inform a wide variety of stakeholders involved in test development and implementation including test developers, regulators, as well as other healthcare decision makers (e.g. commissioners of care) of the evidence base with respect to specific tests.

Different decision makers may place differing emphasis on different aspects of clinical utility (e.g. clinical and personal outcomes, cost-effectiveness, feasibility and test delivery) which can lead to contrasting decisions on clinical utility. Furthermore, decisions with respect to clinical implementation are likely to be influenced by a broader set of factors including other competing priorities and availability of resources.

Demonstrating the safety of assays and tests to be placed on the market is a key element of both regulatory and healthcare evaluation frameworks. Whilst the evidence generated for regulatory

processes and healthcare evaluation overlaps, regulatory approval does not include a full assessment of the clinical utility of a product. Thus regulatory approval is necessary but not sufficient for an assay or test to be implemented into healthcare. However, confusion may arise as a result of a lack of clarity as to where there are overlaps or gaps in the evidence generated for these two processes.

Perspectives on clinical utility of polygenic scores

Our broad examination of the field of polygenic score research and discussions around clinical utility has identified that the discourse is often in relation to the field as a whole with few instances examining specific applications or use case scenarios. Context and application are important in considering the utility of information from polygenic scores, much like for other biomarkers.

Also influencing the debate are the individual perspectives of stakeholders who have different backgrounds and different interests in aspects of polygenic scores. Different perspectives are useful and necessary for a rounded assessment of polygenic scores but have on occasion led to confusion and debate at cross purposes. When both the specific context of polygenic score use and individual perspectives are taken into account, there is in fact broad agreement over areas of likely clinical utility. In addition, there is broad agreement that it is impossible to say polygenic scores will have utility without considering the context in which they are used and the nature of information they are providing in these scenarios.

Thus far much of the debate with respect to polygenic scores has not openly acknowledged the presence of differing perspectives nor the subjective nature of interpretations of clinical utility.

Clinical utility and polygenic scores

Polygenic scores are of interest as a biomarker to inform clinical and personal decision making. Methods for obtaining information on this biomarker and its interpretation are rapidly evolving. Whilst there is broad agreement that polygenic scores could have potential in informing decision making (either personal, clinical or public health), their evaluation and evidence appraisal requires greater clarity in a number of areas.

Firstly, a shared understanding of the concept of clinical utility. As clinical utility is a value judgement, differences in opinions are to be expected. Nevertheless, recognition of this fact is important for open and transparent decision making. Decision makers may place differing emphasis on the different elements that contribute to clinical utility and evaluation from each of these perspectives could lead to different conclusions. Linked with this is ensuring that there is a clear understanding of the nature of information from polygenic scores and how they function as a biomarker. This may aid in ensuring that a balanced view is taken when considering their added value in decision making.

Secondly, ensuring the appropriate evaluation and level of regulation of any polygenic score based application will require clarity on the product or test that is being proposed. The development of these applications will necessitate development and evaluation of polygenic score models. These models can be used by themselves, incorporated within existing risk models that have been translated into tools, or used in informing the development of novel risk models and tools.

This means that different elements contribute towards enabling the use of polygenic scores as a biomarker. These include processes for genotyping, calculating the polygenic score (polygenic score models), their integration into combined risk models and risk tools. The way in which these elements come together in the form of a test or test system needs to be more clearly articulated.

Finally, clarity about the intended purpose and use cases of polygenic score analysis are also important going forward to enable more structured discussions of their clinical utility. Similar to many other tests used in healthcare, polygenic score analysis will not provide definitive answers. Nevertheless, as outlined in Table 3, they can inform different healthcare questions. Determining their utility is linked to assessing how they function in these different scenarios. This is because thresholds of acceptability on test performance characteristics will be influenced by the use case[99]. Furthermore, differing sets of benefits and harms may be associated with different use cases. Thus, as well as validating the variants that are incorporated within the score, and in evaluating performance of models, more research is needed as to how such models and any subsequent tools developed to enable use of these models might fit with existing infrastructure and personnel.

For example, if polygenic score information is to be incorporated within the NHS Health Check programme for assessing cardiovascular disease risk, it will be used in adults aged 40 or above. Information from polygenic scores in combination with results on other risk factors could be used to better target interventions (medication or behaviour change) to those at highest risk, who can benefit the most from these interventions. In addition to the performance of the test used in this scenario, consideration will also need to be given to any changes that may be required to the way in which the programme is delivered to enable incorporation of polygenic score data.

There are likely to be different considerations with respect to different disease areas and clinical pathways, that impact on requirements for evidence generation. Furthermore, healthcare implementation will also be influenced by wider factors including competing priorities or opportunities. A clear understanding of intended purpose and use case can also enable consideration of the practical aspects of implementation.

Conclusion

The development of polygenic-score based tests is ongoing, with certain applications holding promise for use within healthcare. Demonstrating clinical utility is considered a key requirement prior to implementation within health services. However, this is not an easy task as the concept of clinical utility encompasses a range of factors and generating evidence in support of all of these may not be feasible or warranted. Its assessment is also perspective dependent and the context of use will influence thresholds that are deemed acceptable with respect to the evidence base. The use of polygenic score analysis as part of care pathways for different diseases needs to be further investigated in order for a judgement to be made on their utility. Research initiatives such as the Accelerated Detection of Disease Research Platform in the UK and other ongoing research initiatives, implementation studies as well as clinical trials are mechanisms through which insights on the utility of such tests in particular pathways will be gained[9,100–102].

Appendix

The following experts were interviewed for this report.

- Prof Peter Donnelly, CEO of Genomics Plc. Prof of Statistical Science at University of Oxford
- Prof Jon Emery, Herman Chair of Primary Care Cancer Research, University of Melbourne
- Prof Cecile Janssens, Prof of Translational Epidemiology, Emory University, Atlanta
- Prof Anneke Lucassen, Prof of Clinical Genetics, University of Southampton, Honorary Consultant in Clinical Genetics, Wessex Clinical Genetics Service,
- Prof Mark McCarthy, Senior Director and Staff Scientist, Human Genetics, Genentech
- Prof Caroline Wright, Prof of Genomic Medicine, University of Exeter

References

1. AdvaMedDx D. Introduction to Molecular Diagnostics: The Essentials of Diagnostics Series. 2013 https://dx.advamed.org/research/resource-center/introduction-molecular-diagnostics-essentials-diagnostics-series-pdf2

2. Walter FM, Thompson MJ, Wellwood I et al. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. BMC Cancer. 2019 Jun 14; 19(1): 586.

3. Horvath AR, Lord SJ, StJohn A et al. From biomarkers to medical tests: the changing landscape of test evaluation. Clin Chim Acta. 2014 Jan 1; 427: 49-57.

4. Cocco P, Ayaz-Shah A, Messenger MP et al. Target Product Profiles for medical tests: a systematic review of current methods. BMC Med. 2020 May 11; 18(1): 119.

5. Burke W, Zimmern RL and Kroese M. Defining purpose: a key step in genetic test evaluation. Genet Med. 2007 Oct; 9(10): 675-81.

6. Zimmern RL and Kroese M. The evaluation of genetic tests. Journal of Public Health. 2007 Sep; 29(3): 246-50.

7. Kroese M, Zimmern RL and Sanderson S. Genetic tests and their evaluation: Can we answer the key questions? Genetics in Medicine. 2004 Nov-Dec; 6(6): 475-80.

8. Hippisley-Cox J, Coupland C and Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017 357: j2099.

9. Archer S, Babb de Villiers C, Scheibl F et al. Evaluating clinician acceptability of the prototype CanRisk tool for predicting risk of breast and ovarian cancer: A multi-methods study. PLoS One. 2020 15(3): e0229999.

10. Kulm S, Mezey J and Elemento O. Benchmarking the Accuracy of Polygenic Risk Scores and their Generative Methods. https://www.medrxiv.org/content/10.1101/2020.04.06.20055574v1

11. Babb de Villiers C, Kroese M and Moorthie S. Understanding polygenic models, their development and the potential application of polygenic scores in healthcare. J Med Genet. 2020 May 6:

12. Inouye M, Abraham G, Nelson CP et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 2018 Oct 16; 72(16): 1883-93.

13. Lee A, Mavaddat N, Wilcox AN et al. BOADICEA: a comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors. Genetics in Medicine. 2019 2019/08/01; 21(8): 1708-18.

14. McGeoch L, Saunders CL, Griffin SJ et al. Risk Prediction Models for Colorectal Cancer Incorporating Common Genetic Variants: A Systematic Review. Cancer Epidemiol Biomarkers Prev. 2019 Oct; 28(10): 1580-93.

15. Riveros-Mckay F, Weale ME, Moore R et al. An integrated polygenic and clinical risk tool enhances coronary artery disease prediction. medRxiv preprint doi: https://www.medrxiv.org/content/10.1101/2020.06.01.20119297v1

16. Torkamani A, Wineinger NE and Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018 Sep; 19(9): 581-90.

17. Lambert SA, Abraham G and Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet. 2019 Nov 21; 28(R2): R133-R42.

18. Lewis CM and Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020 May 18; 12(1): 44.

19. Steyerberg EW, Vickers AJ, Cook NR et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010 Jan; 21(1): 128-38.

20. Burke W and Zimmern R. Moving beyond ACCE: an expanded framework for genetic test evaluation. PHG Foundation. Available from: https://www.phgfoundation.org/report/moving-beyond-acce-an-expanded-framework-for-genetic-test-evaluation

21. Grosse SD and Khoury MJ. What is the clinical utility of genetic testing? Genet Med. 2006 Jul; 8(7): 448-50.

22. Bunnik EM, Janssens AC and Schermer MH. Personal utility in genomic testing: is there such a thing? J Med Ethics. 2015 Apr; 41(4): 322-6.

23. Kohler JN, Turbitt E and Biesecker BB. Personal utility in genomic testing: a systematic literature review. Eur J Hum Genet. 2017 Jun; 25(6): 662-8.

24. Foster MW, Mulvihill JJ and Sharp RR. Evaluating the utility of personal genomic information. Genet Med. 2009 Aug; 11(8): 570-4.

25. National Academies of Sciences E, and Medicine; Health and Medicine Division; Board on Health Care Services; Board on the Health of Select Populations; Committee on the Evidence Base for Genetic Testing. Genetic Test Assessment. In. An Evidence Framework for Genetic Testing. National Academies Press (US); 2017.

26. Pitini E, De Vito C, Marzuillo C et al. How is genetic testing evaluated? A systematic review of the literature. Eur J Hum Genet. 2018 May; 26(5): 605-15.

27. Augustine EF, Adams HR and Mink JW. Clinical trials in rare disease: challenges and opportunities. Journal of child neurology. 2013 28(9): 1142-50.

28. National Academies of Sciences E, and Medicine; Health and Medicine Division; Board on Health Care Services; Board on the Health of Select Populations; Committee on the Evidence Base for Genetic Testing. An Evidence Framework for Genetic Testing. In. National Academies Press (US); 2017.

29. Council Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices [1998]. OJ L 331/1. ,

30. Council Directive 93/42/EEC of 14 June 1993 concerning medical devices [1993] OJ L 169/1.

31. Council Directive 90/385/EEC of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices. [1990] OJ L 189/17.

32. Council Directive 2000/70/EC of the European Parliament and of the Council of 16 November 2000 amending Council Directive 93/42/EEC as regards medical devices incorporating stable derivates of human blood or human plasma. [2000] OJ L 313/22.

33. Council Regulation 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. OJ L 117/1.

34. Council Regulation 2017/746 of the European Parliament and of the Council on in vitro diagnostic medical devices [2017] OJ L117/176.

35. UK Government. Using the UKCA mark from 1 Jan 2021 Available from: https://www.gov.uk/guidance/using-the-ukca-mark-from-1-january-2021

36. Ordish J, Murfet H, Hall A. Algorithms as medical devices. PHG Foundation. Available from: https://www.phgfoundation.org/report/algorithms-as-medical-devices

37. British Standards Institution (2015) BS EN 62304:2006 + A1:2015 Medical device software – Software life-cycle processes. In: British Standards Institution.

38. British Standards Institution (2017) BS EN 82304-1:2017. Health Software - Part 1: General requirements for product safety. In: British Standards Institution.

39. Department of Health and Social Care. Code of conduct for data-driven health and care technology. 2019 Available from: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

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

41. Seibert TM, Fan CC, Wang Y et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ. 2018 Jan 10; 360: j5757.

42. Fritsche LG, Beesley LJ, VandeHaar P et al. Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan genomics initiative and the UK Biobank with a visual catalog: PRSWeb. PLoS Genet. 2019 Jun; 15(6): e1008202.

43. Vassos E, Di Forti M, Coleman J et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry. 2017 Mar 15; 81(6): 470-7.

44. Chasioti D, Yan J, Nho K et al. Progress in polygenic composite scores in Alzheimer's and other complex diseases. Trends in Genetics. 2019 May; 35(5): 371-82.

45. Mullins N, Bigdeli TB, Børglum AD et al. GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. American journal of psychiatry. 2019: appi.ajp. 2019.18080957.

46. McCarthy MI and Mahajan A. The value of genetic risk scores in precision medicine for diabetes. Expert Review of Precision Medicine and Drug Development. 2018 3(5): 279-81.

47. Belsky DW, Moffitt TE, Sugden K et al. Development and evaluation of a genetic risk score for obesity. Biodemography Soc Biol. 2013 59(1): 85-100.

48. Belsky DW, Domingue BW, Wedow R et al. Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences. 2018 115(31): E7275-E84.

49. Lee JJ, Wedow R, Okbay A et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018 Jul 23; 50(8): 1112-21.

50. Smith-Woolley E, Selzam S and Plomin R. Polygenic Score for Educational Attainment Captures DNA Variants Shared Between Personality Traits and Educational Achievement. Journal of Personality and Social Psychology. 2019 Dec; 117(6): 1145-63.

51. Wald NJ and Old R. The illusion of polygenic disease risk prediction. Genet Med. 2019 Aug; 21(8): 1705-7.

52. Weng LC, Preis SR, Hulme OL et al. Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. Circulation. 2018 Mar 6;137(10):1027-1038.

53. Abraham G, Malik R, Yonova-Doing E et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nat Commun. 2019 Dec 20; 10(1): 5819.

54. Curtis D. Clinical relevance of genome-wide polygenic score may be less than claimed. Annals of Human Genetics. 2019 83(4): 274-7.

55. Figueroa JF, Frakt AB and Jha AK. Addressing Social Determinants of Health: Time for a Polysocial Risk Score. JAMA. 2020 Apr 3:

56. NorthShore integrating genetic tests into primary care, Modern Healthcare; 2020 Available from: https://www.modernhealthcare.com/patient-care/northshore-integrating-genetic-tests-primary-care.

57. Polygenic risk: What's the score? Nature Research; 2020 Available from: https://www.nature.com/articles/d42473-019-00270-w

58. 23andMe Offers New Genetic Report on Type 2 Diabetes. 23andMe; 2019 Available from: https://blog.23andme.com/health-traits/type-2-diabetes/

59. Janssens ACJW. Validity of polygenic risk scores: are we measuring what we think we are? Human Molecular Genetics. 2019 28(R2): R143-R50.

60. Yanes T, Young MA, Meiser B et al. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. Breast Cancer Res. 2020 Feb 17; 22(1): 21.

61. Murthy VL, Xia R, Baldridge AS et al. Polygenic Risk, Fitness, and Obesity in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. JAMA Cardiol. 2020 Jan 8:

62. De La Vega FM and Bustamante CD. Polygenic risk scores: a biased prediction? Genome Med. 2018 Dec 27; 10(1): 100.

63. Duncan L, Shen H, Gelaye B et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nature Communications. 2019 2019/07/25; 10(1): 3328.

64. Martin AR, Kanai M, Kamatani Y et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019 Apr; 51(4): 584-91.

65. Maas P and Overvad K. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States (vol 2, pg 1295, 2016). JAMA Oncology. 2016 Oct 1; 2(10): 1374-.

66. Khera AV, Emdin CA, Drake I et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016 Dec 15; 375(24): 2349-58.

67. Rao AS and Knowles JW. Polygenic risk scores in coronary artery disease. Current Opinion in Cardiology. 2019 34(4): 435-40.

68. French DP, Cameron E, Benton JS et al. Can Communicating Personalised Disease Risk Promote Healthy Behaviour Change? A Systematic Review of Systematic Reviews. Ann Behav Med. 2017 Oct; 51(5): 718-29.

69. Conner M and Norman P. Health behaviour: Current issues and challenges. Psychology & Health. 2017 2017/08/03; 32(8): 895-906.

70. Padilla-Martinez F, Collin F, Kwasniewski M et al. Systematic Review of Polygenic Risk Scores for Type 1 and Type 2 Diabetes. Int J Mol Sci. 2020 Mar 2; 21(5):

71. Oram RA, Patel K, Hill A et al. A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. Diabetes Care 2016 Mar;39(3):337-44.

72. Khera AV, Chaffin M, Aragam KG et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018 Sep; 50(9): 1219-24.

73. Allardyce J, Leonenko G, Hamshere M et al. Association Between Schizophrenia-Related Polygenic Liability and the Occurrence and Level of Mood-Incongruent Psychotic Symptoms in Bipolar Disorder. JAMA Psychiatry. 2018 Jan 1; 75(1): 28-35.

74. Reay WR, Atkins JR, Carr VJ et al. Pharmacological enrichment of polygenic risk for precision medicine in complex disorders. Scientific Reports. 2020 2020/01/21; 10(1): 879.

75. Khera AV, Chaffin M, Wade KH et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019 Apr 18; 177(3): 587-96 e9.

76. Udler MS, McCarthy MI, Florez JC et al. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. Endocrine reviews. 2019 40(6): 1500-20.

77. Elliott J, Bodinier B, Bond TA et al. Predictive Accuracy of a Polygenic Risk Score–Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. JAMA. 2020 323(7): 636-45.

78. Mosley JD, Gupta DK, Tan J et al. Predictive Accuracy of a Polygenic Risk Score Compared With a Clinical Risk Score for Incident Coronary Heart Disease. JAMA. 2020 323(7): 627-35.

79. Oetjens MT, Kelly MA, Sturm AC et al. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. Nature Communications. 2019 2019/10/25; 10(1): 4897.

80. Fahed AC, Wang M, Homburger JR et al. Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. medRxiv. 2019: 19013086.

81. Evans DG, Brentnall A, Byers H et al. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. J Med Genet. 2017 Feb; 54(2): 111-3.

82. Gallagher S, Hughes E, Wagner S et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. JAMA Netw Open. 2020 Jul 1; 3(7): e208501.

83. Lecarpentier J, Silvestri V, Kuchenbaecker KB et al. Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. J Clin Oncol. 2017 Jul 10; 35(20): 2240-50.

84. Fahed AC, Wang M, Homburger JR et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020 Aug 20; 11(1): 3635.

85. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines) Genetic/Familial High-Risk Assessment: Breast, Ovarian and Pancreatic. Version 1.2021. 2020. Available from: https://www.genomeweb.com/sites/default/files/nccn_genetic_cancer_risk_assessment.pdf

86. Shi Z, Yu H, Wu Y et al. Systematic evaluation of cancer-specific genetic risk score for 11 types of cancer in The Cancer Genome Atlas and Electronic Medical Records and Genomics cohorts. Cancer Med. 2019 Jun; 8(6): 3196-205.

87. Jia G, Lu Y, Wen W et al. Evaluating the Utility of Polygenic Risk Scores in Identifying High-Risk Individuals for Eight Common Cancers. JNCI Cancer Spectrum. 2020:

88. Frampton MJE, Law P, Litchfield K et al. Implications of polygenic risk for personalised colorectal cancer screening. Annals of Oncology. 2016 27(3): 429-34.

89. Mavaddat N, Pharoah PD, Michailidou K et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015 May; 107(5):

90. Weigl K, Chang-Claude J, Knebel P et al. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. Clinical epidemiology. 2018 10: 143-52.

91. Mavaddat N, Michailidou K, Dennis J et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. American journal of human genetics. 2019 104(1): 21-34.

92. Henneman L, Timmermans DR, Bouwman CM et al. 'A low risk is still a risk': exploring women's attitudes towards genetic testing for breast cancer susceptibility in order to target disease prevention. Public Health Genomics. 2011 14(4-5): 238-47.

93. Cenin DR, Naber SK, de Weerdt AC et al. Cost-Effectiveness of Personalized Screening for Colorectal Cancer Based on Polygenic Risk and Family History. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2020 29(1): 10-21.

94. Naber SK, Kundu S, Kuntz KM et al. Cost-Effectiveness of Risk-Stratified Colorectal Cancer Screening Based on Polygenic Risk: Current Status and Future Potential. JNCI Cancer Spectrum. 2019 4(1):

95. Callender T, Emberton M, Morris S et al. Polygenic risk-tailored screening for prostate cancer: A benefit-harm and cost-effectiveness modelling study. PLoS Med. 2019 Dec 20;16(12):e1002998

96. Helfand BT, Kearns J, Conran C et al. Clinical validity and utility of genetic risk scores in prostate cancer. Asian journal of andrology. 2016 Jul-Aug; 18(4): 509-14.

97. Callender T, Emberton M, Morris S et al. Polygenic risk-tailored screening for prostate cancer: A benefit–harm and cost-effectiveness modelling study. PLOS Medicine. 2019 16(12): e1002998.

98. Craig JE, Han X, Qassim A et al. Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nat Genet. 2020 Feb; 52(2): 160-6.

99. Sox HC. Assessment of diagnostic technology in health care : rationale, methods, problems, and directions. National Academy Press; 1989.

100. Clinical Implementation of a Polygenic Risk Score (PRS) for Breast Cancer ClinicalTirals.gov; 2020 Available from: https://clinicaltrials.gov/ct2/show/NCT03688204

101. Saya S, Emery JD, Dowty JG et al. The impact of a comprehensive risk prediction model for colorectal cancer on a population screening program. JNCI Cancer Spectrum. 2020:

102. Department of Health and Social Care. GENOME UK The future of healthcare. 2020. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920378/Genome_UK_-_the_future_of_healthcare.pdf