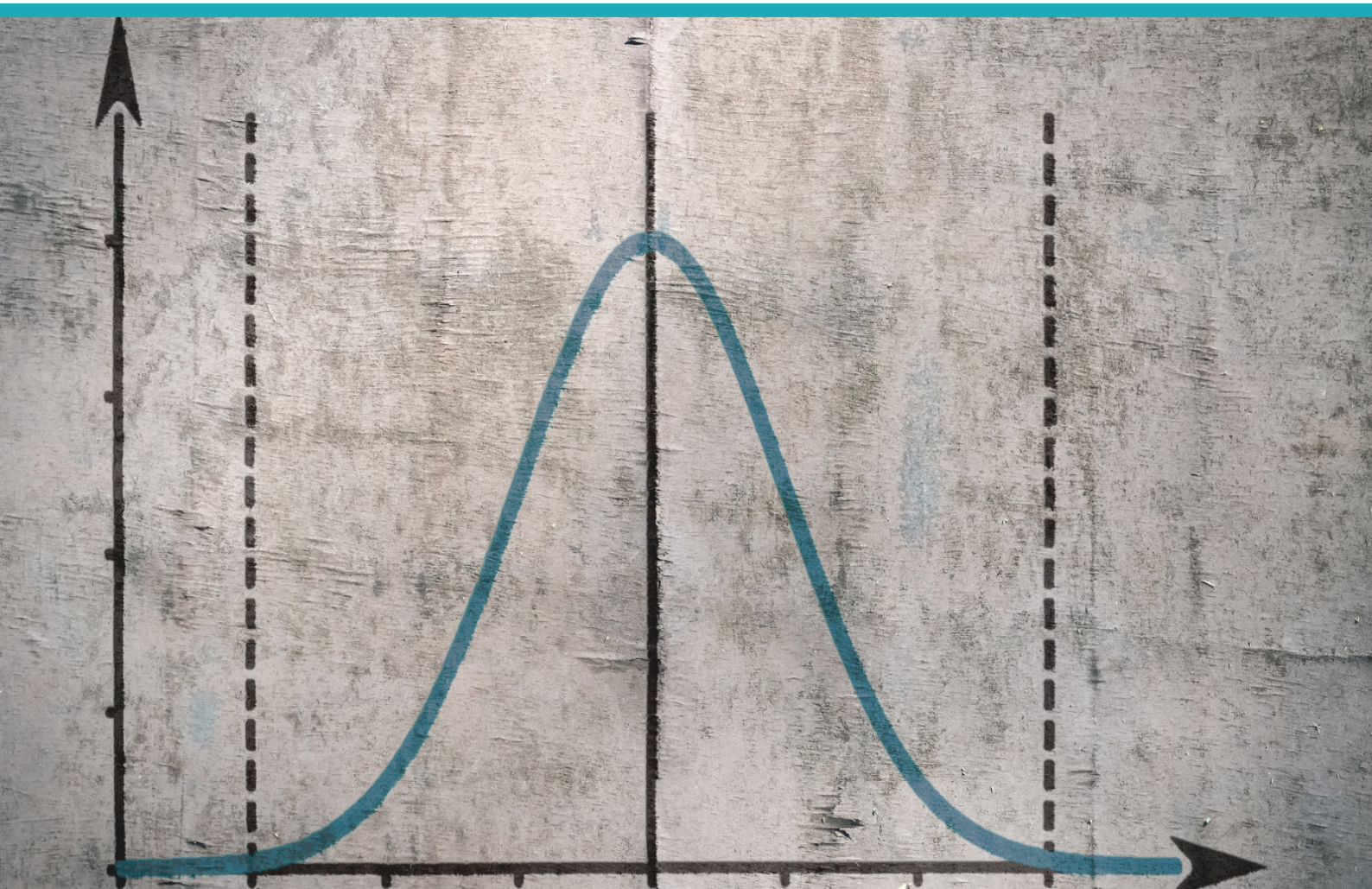
making science work for health

PHG FOUNDATION

# Polygenic scores and public health services

## Moving the debate forward



## Authors

Chantal Babb de Villiers, Pete Mills and Hayley Wilson

## Correspondence to

PHG Foundation can be contacted at intelligence@phgfoundation.org

## Acknowledgments

The authors would like to thank the participants who attended a roundtable meeting on 5 February 2026 for their contributions to the discussion. This report has benefitted from their insights but should not be taken to represent the views of any participant in that event.

**June 2026**



DOI: https://doi.org/10.61599/sibl1

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.

# Contents

# Summary

The 10 Year Health Plan for England contains a commitment to establish a new Genomics Population Health Service by 2035, promising to introduce polygenic scores as a component technology. But the use of polygenic scores in healthcare is fiercely contested. Against this background, PHG convened an interdisciplinary roundtable to explore contexts in which a constructive and inclusive discussion about polygenic scores and their role in healthcare might be possible.

We found that the disagreement was not principally about the insufficiency of available data but about the different kinds of benefits and harms, clinical consequences, alternatives, social and ethical implications, evidence, and interpretive frameworks that are appropriate for the different use contexts.

This report examines interconnected aspects of the debate, including: competing approaches to prevention that underlie many of the most entrenched disagreements; the relevance of different evidence and what appropriate evaluation of polygenic scores requires in different contexts; ethical and social questions, particularly around equity and the communication of risk; and governance and regulatory infrastructure that may be needed to support responsible implementation.

From these discussions, three priority questions emerged:

- The first is to determine which bodies should hold responsibility for the evaluation and deployment of polygenic scores in each specific context of use, and to establish coordination between them.
- The second concerns fundamental principles of ethics, equity and justice that should be settled before implementation can proceed.
- The third concerns contextually appropriate but consistent standards for the evaluation of polygenic scores that reflect the tests' implications for both individual and population health.

Addressing these questions may not be sufficient to overcome the controversies surrounding polygenic scores, but doing so would better define a territory on which differing views can engage constructively.

These questions extend beyond what further research alone can resolve. They are irreducibly questions of policy and ethics as well as questions of science and medicine, and they must be addressed discursively and inclusively. This report contains proposals for what needs to happen in each area and who should take it forward.

# Introduction

## Polygenic scores

Polygenic scores provide a weighted aggregate of a large number of genetic variants associated with small individual effects to assign a single estimate of disease risk. The associations are identified through research – genome wide association studies – that associate genetic variants with observed health effects in large populations. A set of identified variants are selected, each is weighted and the result aggregated by risk tools to produce a single score. This can be reported on its own or combined with other risk factors to give an integrated risk score.

The risk tool that produces the score requires a multistage process that involves the acquisition of a biological sample, the extraction and analysis of DNA in the laboratory to produce genomic data, the retrieval of relevant variants in the dataset, the weighting of these using a risk algorithm, and the computation of the final score.

Polygenic scores can be incorporated into different kinds of tests. The significance of the score will depend on the kind of test and a number of important contextual factors. These include the purpose of testing, the population in which the test is applied, the implications of testing for further action (e.g. in a care pathway), who holds responsibility for further action, and the likelihood of the further action resulting in a health benefit or harm.

The evidence base supporting the use of polygenic scores in healthcare is highly variable across applications. As discussions at our roundtable meeting made apparent, the use of polygenic scores in healthcare is a polarising subject.

The areas of contention span a broad range of issues including analytical validity, clinical utility, evidentiary standards, economic justification, service integration, ethical acceptability and social implications. This makes it extremely difficult to have a meaningful discussion about the value of polygenic scores in general.

Even in deliberative settings, the layering of contested assumptions across these dimensions means that without clarity about context and responsibility, meaningful engagement is difficult to achieve. We return to this in drawing out the implications of our discussion at the end of this report.

## The policy impetus

The UK Government's 10 Year Health Plan for England names population health genomics as a priority for achieving the 'Sickness to Prevention' shift. Risk stratification using polygenic scores is one component contributing to the plan's ambition.

- The new ten-year contract for the Genomic Medicine Service (GMS) from April 2026 explicitly specifies population health functions to meet the goals of the 10 Year Health Plan.
- The National Cancer Plan makes commitments to personalised risk in which genomic stratification is expected to play a part.

- Our Future Health (OFH) have partnered with the NHS to deliver the UK's largest research programme, with a target of five million participants, of which 2.5 million are already recruited. They will pilot integrated risk scores that will include a polygenic score component, providing an opportunity to explore at scale whether and, if so, how such scores might ultimately be implemented in healthcare.

The scale of existing investment in research infrastructure, contractual agreements, and public engagement, signals a level of institutional commitment to the prospective use of polygenic scores.

Policy in this area is also under pressure from a variety of contingent factors. Commercial stakeholders have a clear interest in adoption by the NHS and are active advocates for implementation. The reputation and funding of some academic research groups is also staked in the adoption of these tools into clinical practice. Charities and patient organisations push for faster development and earlier adoption.

Despite these pressures, combined with the political momentum around precision medicine and genomic innovation, commissioners and policymakers must nonetheless exercise caution before making potentially disruptive changes to their services.

## About this report

This report is intended for those to whom it will fall to determine whether, when and how polygenic scores could be implemented in diverse contexts within publicly funded initiatives and health systems, specifically the English NHS.

This report draws on the discussions of a roundtable meeting held under the Chatham House Rule in February 2026. The meeting brought together researchers, clinicians, commissioners and ethicists. Our aim was to understand better why it has proved so difficult to have a constructive and inclusive discussion about polygenic scores, and to explore contexts in which such a discussion might be possible.

The views summarised here do not represent the full range of perspectives within the field, nor should they be taken as the position of any individual, institution, or organisation involved in the roundtable discussion. No commercial, industry representatives or patient and public perspectives were present on this occasion.

# Prevention in the polygenic score debate

A recurring tension between different approaches to improving the health of populations underlies many of the disagreements about polygenic scores. Understanding the implications of these differences is a first step towards bringing clarity to some of the most entrenched arguments in the debate.

## Competing frameworks for prevention

Two unresolved questions underlie disagreement about polygenic scores, and they are frequently conflated.

The first is whether the priority for prevention should be to shift the general level of risk across the whole population or, instead, to identify and target those individuals at highest risk. The two approaches are not inherently incompatible but come into conflict as a result of the need to decide how to use limited resources and the existence of opposing views about ways of trading off harms and benefits among individuals and groups.

The second is whether polygenic scores can estimate individual risk reliably enough to justify targeted intervention in a specific clinical context, and whether doing so adds sufficient value over existing approaches to warrant the effort and expense.

While, at first blush, the first question may look like a normative question of policy and the second a question that can be answered by empirical research, when we examine them more closely we see that each is a complex question of both fact and value to which there is no single or simple way of responding. Some of these entanglements appear in a clearer light when we pay closer attention to the perspective from which the questions are addressed.

Taking a population health perspective foregrounds the fact that a prevention strategy focused on identifying and treating only the highest-risk individuals will always miss the majority of future cases. Most cases of disease arise in the large number of people at moderate risk rather than the small number at very high risk, the so-called ‘prevention paradox’. A more accurate risk tool does not resolve this as it sharpens identification of the high-risk group while leaving the moderate-risk majority, where the greatest disease burden sits, largely unaddressed.

Regardless of the accuracy of polygenic scores, investment in genetic tools to stratify risk will, it is argued, distract from, delay or displace population-wide approaches with a more established evidence base.

For example, providing statins and low-dose blood-pressure medication to all eligible adults over 50 could, assuming full adherence, prevent 60-70% of heart attacks and strokes without risk stratification or genetic testing, an outcome that targeted approaches based on current polygenic score performance could not match. In this context, it is appropriate to ask whether adding a polygenic score to an existing healthcare programme would prevent more heart attacks and strokes than simply extending statins and blood pressure treatment to everyone above a given age threshold.

A different question comes to the fore when the perspective is that of clinical risk management: whether polygenic scores can extend and improve clinical risk assessment in contexts where a clinical pathway is already in place.

Genetic information and calculated risks already shape referral thresholds and prophylactic decisions in several clinical settings, and even modest improvements in prediction can have value if they direct an earlier or more intensive intervention to specific patient groups. In the same disease area, the question becomes whether a polygenic score changes what a clinician should do for the specific patient sitting in front of them, perhaps someone with a borderline risk estimate or a strong family history. Here, a more informed picture of risk might shift the clinical decision.

The public health and clinical management perspectives each speak to different kinds of evaluation. This contrast suggests that greater precision about intended application and the evaluation criteria appropriate to each are required to avoid the risk of advocates and sceptics 'talking past' each other.

The question of whether risk stratification is the right instrument for prevention at all has inescapably ethical dimensions to which we will return below.

It is worth noting that public health and clinical management do not adequately describe the range of different possible use cases for polygenic scores: primary care operates across both; opportunistic case-finding represents another mode of use, and the boundaries between all of these are not always clear in practice.

# Evidence and the evaluation of polygenic scores

A tendency to conflate different types of evidence runs through the field and was apparent during the roundtable discussion. Moving debates about polygenic scores forward requires that appropriate and agreed standards be found to assess evidence suited to each of the relevant areas. However, it also requires that the different forms of evidence be brought together in ways that can be agreed across the disciplinary perspectives relevant to the debate. Seemingly, both of these conditions are yet to be fully achieved.

## What kinds of evidence matter and how do they matter?

Whether a polygenic score calculates reliably (analytical validity), whether it predicts disease accurately (clinical validity), whether it improves patient outcomes in practice (clinical utility), whether its benefit exceeds that of existing alternatives (relative benefit), whether that benefit justifies the cost and complexity of implementation (health economic value), whether it performs equitably (social justice) and whether its use is ethically permissible or required (ethical value) are distinct questions that use evidence and argument in different ways.

### Openness of evidence

For some, the bar has not been met at the most fundamental level of analytical validity. This may reflect two distinct problems: either the evidence does not yet meet the required threshold, or the underlying data are not transparently available for independent scrutiny. These require different responses: one calls for further evidence generation, the other for open data and reproducibility standards. For others, the unmet standard is clinical utility, where evidence that using the tool in a specific clinical context leads to better outcomes than existing alternatives.

### Use context and establishing clinical utility

Polygenic scores are not used in isolation but within systems operating in specific populations and care settings, and the relevant comparison is not simply between using the tool versus not using the tool but between one configuration of care versus another. The evidence requirements for an evaluation of clinical utility will depend on the context in which a polygenic score is being used.

For example, the criteria applicable to a standalone population screening test will be different from those for a tool used to reclassify risk within an existing clinical pathway in which a risk estimate may trigger a referral, a prescription, or a decision about whether or not to intervene. It is therefore important to consider the implications of using the test in the specific context, and how these implications determine acceptable test performance characteristics.

### Improvement of existing clinical tools

Enthusiasm for polygenic scores often fails to take into account the performance of existing clinical tools. For example, established risk algorithms (such as QRisk for cardiovascular disease, and family history-based assessment more broadly) already stratify risk in NHS settings. Whether adding a polygenic score to these existing tools produces a clinically meaningful improvement, rather than one that is statistically detectable but too small to change clinical decisions, is central to the evaluation challenge.

## Examples of polygenic score use cases

### Stratified breast cancer screening

The evidence base for stratified breast cancer screening is more developed than for many other applications, serving as a useful reference point for what rigorous evaluation requires. Integrated risk tools (such as CanRisk) combine polygenic scores with rare variant status, family history, and other factors to produce personalised risk estimates. The rigorous evaluation process includes independent external validation in prospective cohorts, equity analysis across ancestral groups, iterative co-design with clinical services, full medical device regulatory scrutiny, and post-market surveillance planning.

The challenge involved in replicating this developmental process for other applications is worth noting, as it requires substantial investment, and most models in the literature have not been subjected to equivalent scrutiny. This example also illustrates the importance of prior clinical consensus: the case for intervening was already established before the tool was developed, enabling evaluation against a clear and agreed objective.

### Familial hypercholesterolaemia

Familial hypercholesterolaemia (FH) is an inherited disorder of cholesterol metabolism that, if untreated, leads to substantially elevated lifetime risk of premature heart attack and cardiovascular disease. Despite effective treatment with statins, a significant proportion of cases remain undiagnosed. Polygenic scores can identify individuals with an FH-like phenotype who lack a detectable pathogenic monogenic variant, sometimes termed polygenic FH. This application is best characterised as an area of active clinical investigation.

### Diabetes

Type 2 diabetes highlights questions about how results should be acted upon. While scores achieve reasonable discriminatory performance, the primary intervention prompted would be sustained lifestyle change. A large body of evidence demonstrates that risk communication alone does not reliably produce appreciable change. This limits the clinical utility of a risk stratification approach in this context.

### Mental health

The clinical landscape currently has few objective tools for guiding diagnosis or treatment selection. Polygenic scores might eventually support prescribing decisions or identify individuals presenting with early or sub-threshold psychotic symptoms who are at risk of progressing to a schizophrenia-spectrum disorder.

### Agreeing what to measure

What counts as clinically meaningful is itself contested. The absence of consensus on appropriate endpoints is a recognised challenge in the evaluation of complex interventions, and polygenic scores are no exception. Risk reclassification, treatment uptake and mortality are all plausible and relevant endpoints but they differ in their implications. The published literature does not frame questions consistently, which makes cumulative evaluation difficult and means that apparently conflicting findings sometimes reflect disagreement about the endpoint rather than genuine disagreement about the tool.

### The need for prospective criteria

The credibility of evaluation depends on its criteria being fixed in advance and independently of its conclusions. Pre-specification of what success is matters because retrospective criteria are vulnerable to being shaped by the findings they are meant to assess.

### The role of prior policy consensus

One factor that helps to explain why the evidence appears stronger in some disease areas than others is the existence or absence of a prior policy consensus about intervention. Where there is such a consensus (for example, an established national policy such as for breast cancer screening), polygenic scores can be evaluated against the same agreed objective. Where such a consensus does not exist, scientific and value questions of how the tool performs cannot be separated from the prior value question of what a good outcome would look like. Both must be addressed at once, which makes evaluation correspondingly harder.

## Contingent factors affecting evaluation

### Publication bias and techno-optimism

Patterns in what the field publishes and amplifies have become part of the evidence problem. The literature has tended to favour studies reporting novel associations, striking performance metrics, and new applications. The evidence base for polygenic scores therefore leans towards what is technologically exciting, rather than what is demonstrated to be ready for clinical implementation. These tend to overshadow null findings, replication failures, or analyses demonstrating limited performance, particularly in populations with diverse or non-European ancestries.

The cautionary evidence exists and has been published, but it does not appear to carry the same weight. The result is that proponents and sceptics of polygenic scores are often drawing on different bodies of evidence rather than disagreeing about the same data. Those cautioning against rapid expansion are frequently applying the same standards that would be expected of any proposed screening or prevention technology whereas those advocating expansion often apply a more particular standard.

### Evidence that challenges policy ambitions

A related factor is how the field responds to evidence that does not align with technological and policy ambitions. For example, while there is much expectation staked on prevention through voluntary changes in lifestyle and behaviour, such as dietary patterns and physical activity, an accumulating body of evidence suggests that providing phenotypic or genetic risk information does not reliably induce such changes. While this evidence is often acknowledged it may then be set to one side while attention continues to be focussed on refining models or calling for more evidence.

# Ethical and social factors

## Considerations relating to justice

### Individual and population benefit

Polygenic score models for the same condition may produce different results. Differences in training data, ancestry composition, or weighting methods means that two models might assign different scores to the same individual (a high risk rating according to one, say, and moderate risk according to the other) while showing broadly equivalent discriminatory performance at a population level, measured by standard statistical metrics.

The choice between models may therefore appear neutral from a population health perspective but carry potentially significant consequences for individuals. And because polygenic scores assign disease risk rather than detect disease directly, it will not always be possible to identify who has been advantaged or disadvantaged by the choice of model.

Models are therefore not interchangeable, yet there is no indisputable basis on which to choose between them. The choice may also have moral and political implications owing to the relative distribution of benefits. On the other hand, choosing neither model would, in principle, entail an overall disadvantage.

Lack of standardisation and inconsistency between models has direct implications for clinical pathways, patient understanding, and social justice. Furthermore, because model performance varies across ancestry groups, the choice of tool could either deepen or reduce existing health inequalities.

### Equity considerations

Inequitable performance across ancestry groups is a recognised challenge for many clinical tools and risk models. Most existing polygenic score models have been developed using data drawn predominantly from European-ancestry populations. Performance in individuals of African, South Asian, East Asian, and other ancestries is generally lower, sometimes substantially so. And these ancestry groups may be at greater risk for the target diseases than the populations in which the tools have been developed and validated.

A tool that stratifies risk for some groups and poorly for others does not merely underperform but risks directing resources and interventions inequitably across the population it is meant to serve.

Whether equity of performance should function as a threshold condition for deployment (meaning a tool that does not meet defined standards across ancestry groups should not be used, regardless of its overall performance) was a question raised at the roundtable. Some argue that no model should be deployed until equitable performance can be demonstrated across diverse populations; others argue that inequitable performance is a limitation to be addressed iteratively rather than a condition that must be met before initial use.

## Implications arising from the nature of the test

### Disaggregation of test elements

In the background of all these discussions is an ever evolving concept of what constitutes a 'test' in modern medicine. Genomic data can now be acquired in multiple contexts, and outside a predefined clinical pathway. The rationale for acquiring a biological sample, for running an assay and for analysing the data generated may be treated, in principle, as independent and separated in time and purpose. For example, a polygenic score for cardiovascular disease may be generated for two individuals, in one case using sequencing data obtained decades earlier and retrieved from that individual's health record, and, in the other case, through a newly ordered test following a medical consultation, sample collection and analysis.

This disaggregation and potential recombining of test elements for different purposes adds a further source of potential inequity (between those whose sequence data are in the record and those whose are not). It also further incentivises data acquisition (and more comprehensive data acquisition, when opportunities arise), raising significant considerations about the security, cost and appropriateness of the retention of personal genomic data. As the example shows, it is not enough to evaluate these merely in relation to distinct applications; to appreciate their full implications they must be considered in relation to alternative imaginaries for the future of the whole health system.

### Communicating probabilistic risk

There is a persistent risk that probabilistic genetic information may be received as more determinative ('genetic determinism') than it actually is: as a diagnosis or a destiny rather than a risk factor. A polygenic score places an individual on a distribution of risk across a population. It is not a diagnosis or an individual prediction. The risk categories applied to that score (e.g. high, intermediate, low) are assigned according to predetermined thresholds that are not fixed but reflect contingent factors including the availability of interventions, resourcing decisions, and clinical conventions.

How this information can be communicated in ways that are meaningful, accurate, and actionable remains an open question. What is actionable depends on the availability of interventions and the clinical context, meaning that a score that cannot currently be acted upon may become clinically relevant as circumstances change, and vice versa. This has practical implications for what results are returned and by whom, in what clinical context, and whether the information serves or burdens the people who receive it.

# Requirements for governance

For the use of polygenic scores to be subjected to proper governance two kinds of procedures need to be in place. The first concerns the standards and processes by which polygenic scores are evaluated before deployment. The second concerns how their use is monitored and regulated once in clinical practice. Both remain underdeveloped.

## The absence of agreed pathways for implementation

### Ambiguity and confusion over governance responsibilities

Currently, no standing body has a clear remit to establish application-specific evidential standards for polygenic scores. The UK National Screening Committee provides a relevant model for applications that fall within the scope of population screening; NICE provides a model for health technology assessment more broadly; genomic tests, including those currently in use through the NHS Genomic Medicine Service, are commissioned through NHS England and the National Genomic Test Directory, which have their own criteria and processes. None of these procedures is straightforwardly adapted to the specific features of polygenic score evaluation, which include a multiplicity of models for the same disease, the rapid evolution of the evidence base, and the multi-disease value that may arise from a single genotyping event.

### Multi-disease testing

Where whole genome sequencing is used to collect genotyping information, a single blood sample can support the use of multiple polygenic score models and the application of models for multiple diseases, whether run at the time of collection or applied later as new models are developed, without the need to return to the individual for another sample.

The repurposing of genomic data is made possible by the distinctive characteristic that, unlike other biomarkers, they are invariant over the lifetime of an individual. This separation of the upfront costs and challenges of data acquisition, and the marginal subsequent costs of data use creates novel challenges for health economic assessment and equity of access to services, as data acquisition may occur through a variety of routes in a way that is non-standard across a population.

Managing the cumulative complexity of what is returned to individuals, and ensuring they understand what each element of that information means, represents a further governance challenge that existing frameworks were not designed to address.

### Standardisation of tools and outputs

A further governance challenge, which has not yet been systematically addressed, is technical standardisation: ensuring that polygenic score tools deployed across different platforms and settings produce consistent, comparable outputs, and that the composition of commercially available tools is transparent and subject to oversight.

## Regulatory barriers to deployment

Polygenic score tools that are embedded in clinical decision-making fall within the scope of medical device regulation and, in the UK, must align with regulations governing medical devices or in vitro diagnostic (IVD) medical devices (in Great Britain these are contained in the UK Medical Devices Regulations 2002).

### Sustainability challenges for academic developers

The regulatory framework was designed around a commercial model. As such, it makes no concessions to enable non-commercial tools to be sustained in clinical use over time. As any update of a tool requires that the full regulatory approvals process is begun again, for academic developers without ongoing commercial revenue, the costs and organisational demands of maintaining regulatory compliance can be prohibitive. The box below outlines two examples of this.

### Examples of non-commercial tools with high regulatory maintenance costs

**The CanRisk tool for breast cancer risk**

The CanRisk tool for breast cancer risk is an illustrative case. This tool has been through full regulatory scrutiny, holds appropriate device status, appears in guidelines and has post-market surveillance in place. This process was described as costly, complex, and logistically demanding, particularly for tools developed in academic rather than commercial settings. This raised questions about whether the current framework is fit for purpose when applied to publicly funded, non-commercial tools that nonetheless meet clinical needs.

**The familial hypercholesterolaemia (FH) risk tool**

The familial hypercholesterolaemia (FH) risk tool (a similar clinical decision support instrument that incorporates a polygenic score model alongside lipid measurements and family history to identify individuals with an FH-like phenotype) is a further example: it had to be moved from a university setting into a separate entity in order to be deployable in NHS systems, and any subsequent update would require the full regulatory process to begin again. The practical consequence is that improvements to the tool, whether driven by new evidence or broader clinical application, carry a regulatory cost that may simply be unaffordable for the teams responsible for it, meaning that patients and clinicians could be left using an older version of a tool long after better evidence has become available.

These examples point to a structural gap: a regulatory framework that is well defined in terms of requirements but designed around commercial developers who have the infrastructure, revenue, and organisational capacity to meet them on an ongoing basis. For publicly funded academic science, those conditions are rarely in place.

# Conclusion

Polygenic scores sit at the intersection of several well-established interpretive disciplines that may apply different standards, pose different questions, and suggest different conclusions from the same body of evidence. These are complex judgements relying on both facts and values, which the discussion at the roundtable helped to expose.

There is substantial disagreement about how to approach prevention of disease and what standards are appropriate for evaluating tools designed to support it. These are irreducibly questions of policy and ethics as well as questions of science and medicine that must be addressed discursively and inclusively, through deliberation involving relevant perspectives. The fact that they are as yet unresolved is not a reason to halt the scientific work being done. Addressing these questions is a precondition for ensuring that the findings of research can be acted on when they emerge.

It should be noted that some participants in the roundtable meeting took the view that the evidence already available is sufficient to conclude that population-level deployment of polygenic scores is not warranted, and that further evaluation risks diverting resources and attention from underused population-wide prevention approaches with a stronger and more established evidence base. This concern should be taken seriously, particularly where polygenic scores are proposed as a component of population-wide prevention programmes.

At the same time, polygenic scores are beginning to appear in the NHS via research programmes and clinical trials, and there is considerable ambition to explore the potential of the tools in a wider range of contexts. Commissioners, clinicians, and the public are already facing questions about how potential use cases can be prioritised, shaped and evaluated. Importantly, they are having to consider whether and how these tools can practically integrate with, and adapt to, frontline NHS services.

## Areas of emerging agreement

Several points of convergence emerged from the discussion at the roundtable.

One was that polygenic scores cannot usefully be evaluated as a single technology, that the same standards of evidence should apply to them as to any comparable screening or prevention tool, and that the polarisation of the debate has obscured more than it has revealed.

There was broad acceptance that polygenic scores offer a small incremental improvement in risk prediction in some disease contexts, though disagreement remains over what clinical value can be derived from these, as well as over how large any improvement would need to be to justify the cost, complexity, and risk of implementation.

There was substantial agreement that the polarisation of the debate on polygenic scores obscures more than it reveals. Several participants noted that most people in the room were asking that genomic medicine, and specifically polygenic scores, be held to the same standards of evaluation as are applied to any comparable screening or prevention technology in the same context of use. Much of what has appeared as fundamental disagreement about the evidence is better understood as the evaluation criteria being applied to different intended uses.

It was also common ground that polygenic scores cannot usefully be discussed as a single technology: the question of utility is inseparable from the specific disease, clinical pathway, and intervention in question, and a finding in one context does not necessarily imply a similar finding in another.

There was broad recognition that polygenic scores are frequently presented as a package (technology, infrastructure, and future promise bundled together) in ways that create pressure to accept or reject the whole rather than to optimise the combination of component factors for contexts in which they may be capable of delivering value.

Finally, there was agreement about the need for greater consistency, coordination and transparency across the evaluation frameworks that already exist for different purposes, and an entreaty that, where those frameworks reach different conclusions about the same application, the divergence is explicitly justified rather than left unresolved.

## Priority questions that should be addressed

Several questions emerged from the roundtable that, if they are not addressed, could lead proponents and critics of polygenic scores to continue disagreeing without really engaging with the other's perspective. The most likely consequences of this would be either an immobilising impasse or the impatient overriding of valid concerns, neither of which is a satisfactory outcome.

### Determining responsibility for contextual deployment

The first challenge is to clearly distinguish the different contexts for potential deployment of polygenic scores. These range from population screening programmes to adventitious case finding, to different points in clinical pathways.

This task falls in parallel to determining which body or agency should have responsibility for consideration of polygenic scores in each context. This is not a simple matter. Different pathways may lead to the same testing context, and the circumstances in which people arrive will vary. Moreover, the elements of a test (acquisition of samples, biochemical assay, etc) may be undertaken across more than one context, with none having oversight of the whole.

Thus, as well as to agree in the first place who is responsible for deciding when a polygenic score might be deployed in different contexts, coordination among relevant bodies and agencies is desirable. This should be considered an ongoing basis to address, and either justify or redress, inconsistencies, gaps and overlaps that might arise with the use of polygenic scores in different contexts.

No standing mechanism currently coordinates the range of bodies that have an interest in polygenic score evaluation. To address this, the National Genomics Board, which provides strategic oversight and supports delivery of the genomic healthcare strategy, could consider the desirability and design of such a mechanism, consistent with the UK Government's commitments to regulatory alignment.

### Addressing ethical challenges

The potential for inconsistencies to arise as a result of both the way in which people access tests and the features of the tests themselves raises ethical and social issues that demand to be addressed prior to any deployment of polygenic scores in the NHS. These include, for example, what standardisation of tools is required so that the selection of one tool in preference to another does not create arbitrary interpersonal injustices.

Other questions relate to equity among groups, such as whether deployment is acceptable in population subgroups for which evidence supports their use, even if doing so risks widening health inequalities. On the other hand, should the need to address the evidence gaps for some groups (or make compensatory provisions) be sufficient grounds to delay deployment where benefits can be expected for other populations?

These questions require careful ethical analysis and societal engagement, demonstrating attention to the concerns and experiences of those whose interests may be affected.

It would be helpful to convene an independent group or, possibly, a working party reporting to a body such as the (currently suspended) NHS Genomics Ethics, Equity and Legal Advisory Group to establish working principles and any general criteria that should be met before any potential deployment of polygenic scores in the NHS, and advise those responsible for each testing context accordingly. This group would need to have the remit, and be given the resources and expertise to engage with communities and groups affected (e.g. communities at risk and charities focussed on particular disease areas).

### Defining standards for service evaluation

For any context within which prior research, and social and ethical analysis, suggests polygenic scores meet the threshold for deployment, the question of context-specific, but generally consistent, standards and procedures for service evaluation arises. This would generally include the selection of relevant comparators against which polygenic scores are to be assessed and alignment with standards of evidence available for them, including agreement about relevant endpoints (e.g. entry into treatment, mortality etc).

While existing evaluation frameworks may be appropriate, clarity is needed about which framework is appropriate for each category of application, and what information is required to support it, though it should be applied in each case with the same rigour that would be expected of any proposed screening or diagnostic technology.

Concrete proposals for service evaluation should be the responsibility of the contextually proper authority in each case, developed through a process that is transparent, consultative, and attentive to the full range of relevant expertise. Once proposed, the standards and procedures can be consulted on and publicised as the policy of the appropriate national body so that test developers and users have a common standard in view.

*

Polygenic scores have long been discussed as a clinical application of genomic knowledge but controversies about whether and how they might be used, and the nature and balance of the potential benefits and harms they entail, continue.

These controversies will have to be confronted if polygenic scores are to play the role that is envisaged for them in the genomic population health service, to be established under the Government's 10 Year Health Plan for England.

By attending more closely to the proposed context of use, the distribution of responsibilities that belong to each context, and the multidimensional forms of the evidence that would help to discharge those responsibilities in each case, it may be hoped that some of these controversies can now begin to be addressed.

**Authors:** Chantal Babb de Villiers, Pete Mills and Hayley Wilson

**Contact:** intelligence@phgfoundation.org

June 2026

DOI: https://doi.org/10.61599/sibl1