UNIVERSITY OF
CAMBRIDGE

Polygenic scores for cancer

PHG
FOUNDATION

making science
work for health

Authors

Dr Chantal Babb de Villiers, Dr Sowmiya Moorthie,
Dr Mark Kroese, Dr Laura Blackburn

Key points

◆ Cancer is a highly variable and complex disease, and each cancer has a variety of factors – e.g. genetics, environment – that contribute to its development. The genetic component of different cancers, and their subtypes, varies.

◆ Between a third and a half of cancer cases could be prevented if current knowledge about risk factors was translated into effective public health actions.

◆ Comprehensive risk prediction models (RPMs) that bring together information across a diverse range of factors can inform prevention strategies. The extent to which such models are used in clinical practice varies.

◆ Polygenic scores are considered a measure of genetic contribution to the risk of developing cancer. They are now being considered as a factor for risk prediction which could be used independently or as part of RPMs for cancer.

◆ Polygenic scores can improve risk prediction in some cancers and some clinical contexts, but the magnitude of the improvement varies between cancers.

◆ Any use of polygenic scores in cancer management will be specific to the cancer and the clinical context. This is not a one size fits all solution.

◆ Wider implementation of risk prediction using polygenic scores requires sufficient understanding of how they will affect clinical care, as well as wider infrastructure considerations for delivery.

◆ Continued effort is needed to gather the appropriate evidence for evaluation and demonstration of utility that would support implementation efforts.

◆ Premature implementation of polygenic scores in cancer risk estimation approaches could undermine these efforts, and risk loss of confidence in this potentially valuable area of population health improvement.

Contents

Risk factors

Risk factors

Cancer is one of the leading causes of death worldwide, accounting for nearly one in six deaths, and the number of new cases is rising. There were over 18 million new cases and almost 10 million deaths from cancer in 2020, with over 28 million cases expected by 2040 due to aging and growing populations[1]. It is therefore a global health priority, especially in terms of improved prevention of the disease.

Cancer is a complex and heterogeneous disease that can occur at any age. It arises from a series of genetic changes that disrupt normal controls over how cells in the body grow and function. The chance of these genetic changes occurring (and hence of disease developing) are affected by many different influences. These include:

◆ **Genetic factors.** Some genetic changes that contribute to cancer risk can be inherited and are often associated with a family history for cancer. Familial or hereditary cancer syndromes occur when a variant that confers a high-risk of cancer is shared between family members. Some cancers have well established genetic links that can explain a portion of the cancers that occur, for example pathogenic variants in *BRCA1* or *BRCA2* for breast and ovarian cancers. The level of increased risk associated with individual genetic variants varies widely. Lower (though potentially still significant) genetic risks of cancer are harder to identify, however knowledge of increased genetic risk of cancer can help inform cancer management.

◆ **Environmental factors.** Many environmental or lifestyle factors can affect cancer risk, most notably diet, alcohol consumption, obesity, exercise, sun exposure, and tobacco use. Some cancers are more strongly associated with specific environmental factors, such as skin cancer and sun exposure. The potential to change or modify environmental factors to decrease risk of cancer is an important element of prevention strategies.

◆ **Infectious agents.** Certain infections can also be a risk factor for cancer development, for example human papilloma virus (HPV) infection and cervical cancer, or hepatitis B and C virus infections and liver cancer. In low and middle-income countries, infectious agents contribute to the development of a relatively high proportion of cancer cases, partly due to the higher burden of infectious diseases in these countries.

Prevention and screening

Prevention and screening

Given the complexity and the highly heterogenous nature of cancer a wide variety of approaches are necessary for its prevention, early detection and management. These initiatives can be cancer specific or relevant across many cancers and focus on primary prevention or secondary prevention, which includes screening.

Primary prevention

Primary prevention seeks to reduce the incidence of a disease within the population by preventing it from ever arising. It typically relies on universal public health interventions targeted at modifiable environmental risk factors.

Primary cancer prevention typically includes measures to promote healthier lifestyles and reduce consumption of unhealthy food, tobacco or alcohol. Protective approaches such as vaccination programmes can reduce the risks conferred by infectious agents. Between a third and a half of cancer cases could be prevented if current knowledge about risk factors - smoking, alcohol use, balanced diet etc - was translated into effective public health actions.

For high-risk sub-groups, specialised interventions may also be feasible, such as surgical removal of the breasts and ovaries for women at very high-risk of breast and ovarian cancer.

However, for some cancers, such as, brain tumours, there are currently no known prevention approaches.

Between a third and a half of cancer cases could be prevented if current knowledge about risk factors was translated into effective public health actions

Secondary prevention

Secondary prevention seeks to identify early stage disease, before the onset of clinical signs and symptoms. Interventions are aimed at slowing or preventing disease progression. In general, if identified early, cancer is more likely to be treatable resulting in a greater probability of survival and less morbidity with improved quality of life.

◆ Early diagnosis focuses on detecting symptomatic patients as early as possible

◆ Screening consists of testing healthy individuals to identify those with cancer before any symptoms appear

Testing for early markers of disease is used to facilitate early diagnosis and treatment. For example, the UK bowel cancer screening programme offers faecal testing for this purpose to men and women aged 60 to 74; those flagged as higher risk will be offered colonoscopy to detect early signs of colorectal cancer.

Health systems focus prevention efforts on population screening to support early detection since many treatments are more successful on early stage cancers.

Cancer biology, disease progression, prognosis and outcome are highly variable across different cancers. For uncommon and difficult to diagnose cancers, such as brain cancer or pancreatic cancer, a combination of their rarity and late appearance of symptoms typically results in low clinical suspicion of the disease, and therefore later diagnosis and poorer outcomes. Other cancers, such as non-melanoma skin cancers, are easier to detect early and treat.

These challenges and the imperative to improve cancer prevention and outcomes, mean new tools to support risk prediction, prevention, early detection, screening, diagnosis, and management of cancer are being explored.

Risk prediction

Screening and prevention often uses risk prediction to determine those participants on which to focus efforts. Understanding the factors associated with underlying disease development can provide information to enable risk assessment, which can provide early warning for people at increased risk of developing that disease.

For cancer, risk predictions based on age and sex are useful, as both are strongly associated with risk. However, as knowledge into the causes of disease has improved, additional environmental and biological factors have been identified as associated with disease.

In many common cancers, comprehensive risk prediction models (RPMs) that combine a set of factors, can provide a risk estimate. The contribution of each factor to the risk prediction varies between cancers and the clinical purpose of the risk prediction, affecting their suitability for inclusion in a RPM. The effort and cost necessary to collect the information will also affect decisions on inclusion in risk prediction for use in clinical care.

While pathogenic variants such as those in *BRCA1* and *BRCA2* have been used for some time to support cancer risk prediction in those with cancer, carrying the variant and their families, other genetic approaches are being explored. These include polygenic scores, which are now being considered as a measure of genetic contribution to the risk of developing cancer and that could be used independently or as part of RPMs.
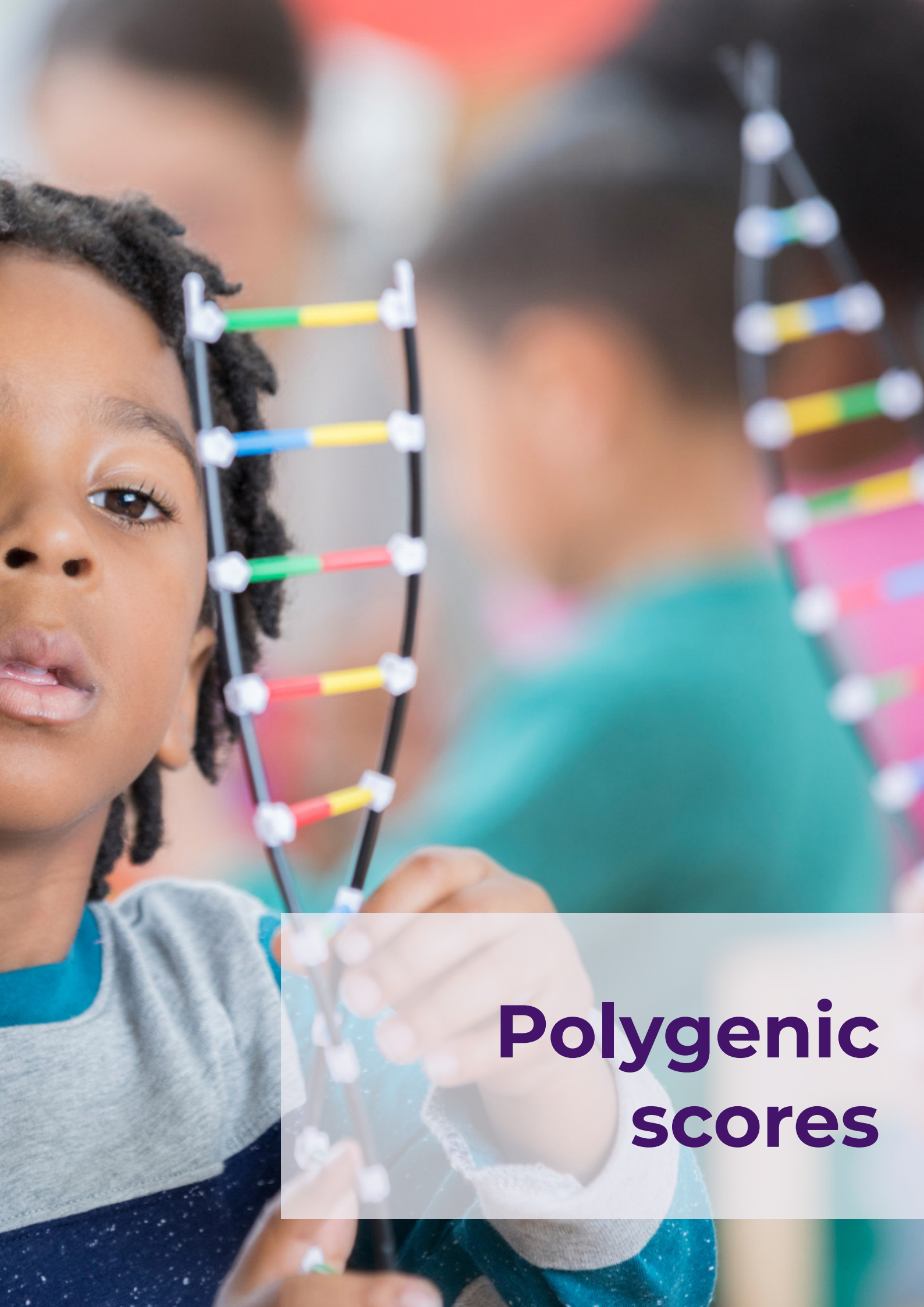
Polygenic scores

Polygenic scores

Common genetic variants contribute to risk of disease. As each variant has a low impact on disease risk, it is useful to look at them in combination.

Each polygenic score is a single measurement that combines the effect of a large number of individually low-impact genetic changes. Together, these small changes in risk may pose a greater, potentially significant risk. Polygenic scores are also known as polygenic risk scores (PGS/PRS) or genetic risk scores. They are considered helpful in predicting the chance of disease occurring based on multiple genetic changes.

The information used to develop polygenic scores mainly comes from genome wide association studies (GWAS), which analyse large numbers of common genetic variants and their association with disease.

Since 2005, GWAS have successfully uncovered many common genetic variants associated with a plethora of human characteristics and disorders, including cancer. The majority of these variants are single letter changes to the genetic code known as single nucleotide polymorphisms (SNPs). As these studies grow, the more power they have to reliably identify SNPs associated with a disease.

Individual SNPs associated with disease have a small effect, meaning that they tend not to be useful on their own to predict the risk of developing a particular disease. Polygenic scores sum the effects of multiple SNPs into a single number to estimate the genetic predisposition for a trait, such as cancer.

The selection of which SNPs to include is based on the mathematical methods used to develop the polygenic score model, the parameters and thresholds set for the model, and the strength of the association of a SNP to the disease, which is typically done by weighting each SNP and combining them with others.

There are many ways to calculate a polygenic score... and no single model development method has been found to be the most predictive across all diseases

In general, polygenic score models include only SNPs highly associated with the condition of interest. Alternatively, some models are 'genome-wide' and will include all the SNPs associated with a disease.

There are many ways to calculate a polygenic score and when models are being developed, different methods will be compared to determine which is most predictive for that disease.

No single model development method has been found to be the most predictive across all diseases, and a variety of methods are being used. It is highly probable that many more polygenic score models will be developed and that they will be updated as more data and information become available.

A further complexity when developing cancer polygenic scores arises from cancer heritability, a measure of the genetic influence on cancer risk. Studies have reported contradictory results for the heritability of cancers. This could explain why there are a variety of polygenic score models for the same disease.

However, it is not clear whether these contradictory results arise from the genetics of the cancers, or factors associated with the databases or technologies used in developing polygenic score models[2, 3]. Greater clarity is expected as genetic research studies move to more detailed methods of analysis such as whole genome sequencing.

Between the variety of models and the variability of research findings, selecting the 'best' polygenic score model for a particular disease is not straightforward. Models have been developed through diverse approaches, in separate contexts, for varied reasons and using different populations. There is not a 'one size fits all' solution in terms of developing a cancer polygenic score model.

This is a rapidly growing field in which a lot of research activity is taking place and resources are being developed to monitor and capture some of this activity. One such resource is the PGS Catalog[4], which makes it easier for users to review and compare available polygenic score models, allowing those that are interested to select a model that best suits a particular setting and population.

The PGS Catalog

The PGS Catalog is an open database of published polygenic scores developed for a range of diseases and characteristics (traits). It has been developed by a consortium of researchers from Health Data Research UK, the University of Cambridge, the Baker Institute and the European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI).

A polygenic score must meet several eligibility requirements before it can be listed in the PGS Catalog – these include information from the peer review publication, details of the trait and of the polygenic score development, descriptions of the sample, polygenic score performance metrics, ancestry and dataset descriptions.

Importantly, the PGS Catalog collects information on polygenic scores that have been developed in research. It does not provide evaluations or assessments on the quality of a polygenic score's utility in a clinical setting. This would need to be done by the users of a polygenic score to support clinical care.

Polygenic scores for cancer

The PGS Catalog is currently the foremost resource on polygenic score models. Although not comprehensive it provides a useful indicator of the cancers for which scores have been developed.

Six common cancer groupings in the PGS Catalog
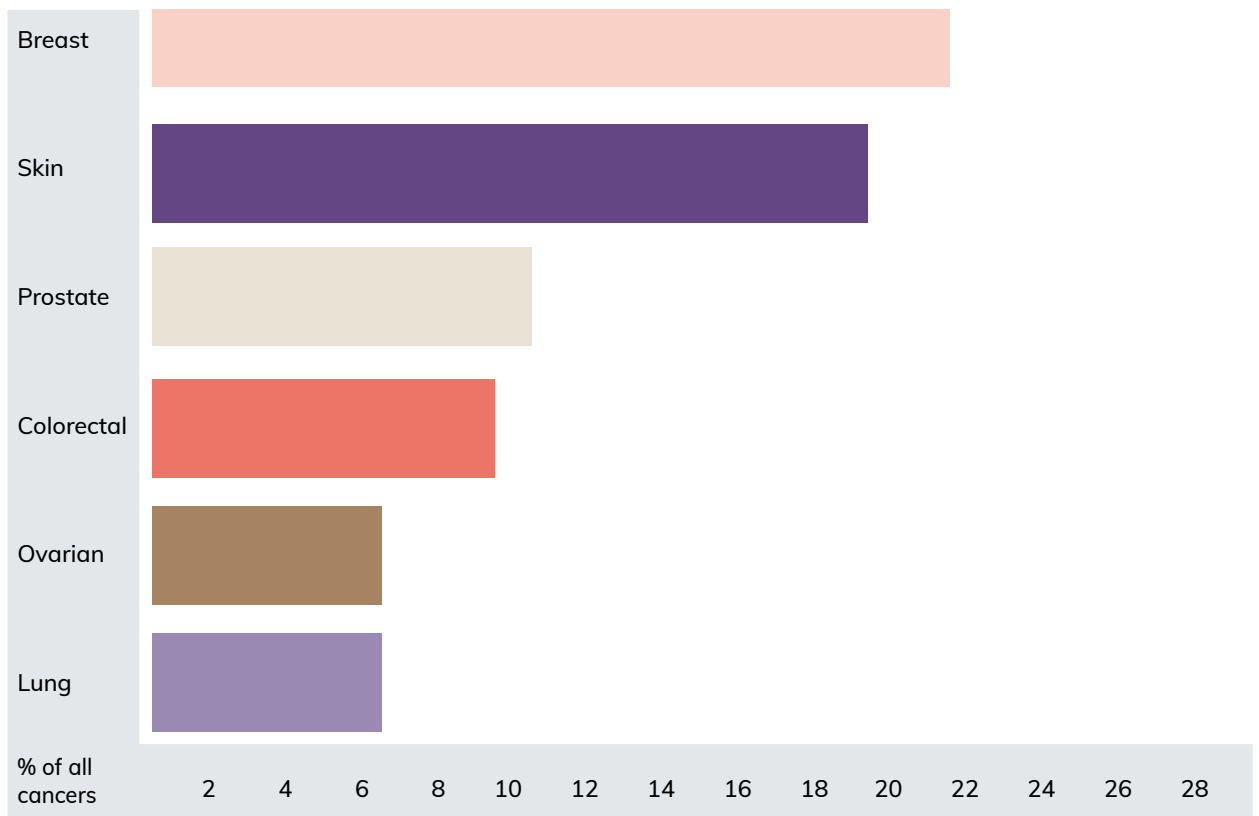
As of May 2022, the 16 most common cancer groupings in the PGS Catalog are:

◆ Breast (~21% of the cancer polygenic score models listed). Includes: breast carcinoma (almost 85% of those listed), triple-negative breast cancer, oestrogen-receptor (ER) negative breast cancer, ER positive breast cancer, HER2 positive breast carcinoma, HER2-receptor negative breast cancer, luminal A breast carcinoma and luminal B breast carcinoma

◆ Skin (19% of the cancer polygenic score models listed). Includes: basal cell carcinoma (~28% of the non-melanoma skin cancers), cutaneous melanoma and melanoma (together 40% of all the skin cancer), skin cancer, skin carcinoma, skin carcinoma *in situ*, and squamous cell carcinoma

◆ Prostate (~10%)

◆ Colorectal (~9%)

◆ Ovarian (6%)

◆ Lung (~6%)

◆ Thyroid, testicular, urinary bladder, brain, leukaemia, lymphoma, tracheal, pancreatic and uterine cancers (each 1–4%)

Most polygenic score development activity is for the commonly diagnosed cancers – breast, prostate, colorectal, lung and skin cancers. This is to be expected, as these common cancers will have bigger patient groups, making it possible to do larger GWAS to detect SNPs associated with disease, which can then be used to develop polygenic score models.

The larger sample sizes available for breast cancer research, and the many genetic factors influencing the disease, mean it is one of the more advanced areas of cancer polygenic score research, with several polygenic score models available for breast cancer sub-types.

Pan-cancer research

Pan-cancer studies look at genetic variants across a range of cancers and provide insights on their genetic profile and how the polygenic score models compare to each other [5]. They also provide valuable insight into which cancers are more promising candidates for polygenic score risk prediction. The studies show that the predictive ability of a polygenic score is not reliant on the number of associated SNPs but on the combined predictive ability of each SNP used to calculate the polygenic score. For example:

◆ Some GWAS have identified a large number of SNPs associated with a cancer, but when used to calculate a polygenic score, it weakly predicts cancer risk. This phenomenon is seen for some cancers that have known strong environmental risk factors and distinct subtypes, such as lung, oropharynx, and oesophageal cancers[6].

◆ Other GWAS with a relatively limited number of samples have nevertheless identified a large number of SNPs associated with a cancer. Due to the strength of the association, the SNPs could still be used in a polygenic score, for example in testicular cancer and chronic lymphocytic leukaemia[6].

◆ Another group is GWAS that had a relatively large number of samples, but only detected a small number of SNPs associated with the cancer. Despite the small number of SNPs, they still provide some risk prediction when used to calculate a polygenic score, for example in thyroid cancer[7].

Pan-cancer studies have demonstrated that the predictive ability of polygenic scores will vary by the cancer type[6]. This is another indication that polygenic scores may be more useful for some cancers and not others.

Limitations

All risk factors, whether a biomarker or environmental, have limitations when used in clinical care. Two important ones are generalisability and gene-environment interactions.

Generalisability

One critical limitation for the translation of polygenic scores for wider application in cancer prediction and prevention is the generalisability of scores across all populations.

There is known to be a bias in genetic research databases towards individuals with European ancestry, with the majority of GWAS undertaken using data from these population groups. Unsurprisingly, there are indications that the predictive performance of polygenic scores decreases when used in populations of non-European ancestries. This is a major issue for applicability of polygenic scores to different population groups: using scores with these could increase health disparities, as they will be less effective in individuals with different ancestries[8].

Statistical steps are being explored that adjust for the population in the polygenic score calculation, but this is not a viable solution for the long term. The use of more representitive genetic databases for the development and testing of polygenic score models in different population groups must be a priority if polygenic scores are to be accurate for all populations of the world.

Gene-environment

Environmental factors also have an impact on polygenic score development and use. Relatively little is known about the complex interactions between environmental and genetic factors in relation to cancer and the impact these may have on risk prediction of cancer.

Gene-environment: weight

Excess weight is linked to several cancers. Weight is considered an environmental risk factor, but there is a genetic component. Whether polygenic scores are associated with the risk factor (e.g. weight), rather than the cancer itself, may need further investigation.

Limited understanding of how genetic and environmental factors interact and contribute to disease risk may result in over or underestimation of risk from polygenic scores.

The problem with genetic research

Genetic research, on which polygenic score models are based, is currently biased towards individuals with European ancestry. There are indications that the predictive performance of these models decreases when used in individuals of non-European ancestries.

This is a major issue: using polygenic scores in different population groups could increase health disparities as they will be less effective for populations underrepresented in genetic research.

One example is body weight, where excess weight is correlated with a variety of cancers[10]. Although weight can be considered an environmental risk factor, there is a genetic component. Whether polygenic scores are associated with the risk factor (e.g. weight), rather than the cancer itself, may need further investigation. Limited understanding of how genetic and environmental factors interact and contribute to disease risk may result in over or underestimation of risk from polygenic scores.

Translation into clinical use

Determining how to deploy any test in a clinical setting is a complex undertaking that must account for disease dynamics in the target population as well as the availability of interventions. Disease features vary between cancer types, and implementation strategies to balance the potential benefits and harms will differ between the cancers. The location and natural history of a cancer type (how the disease arises and how quickly it progresses), along with the types and efficacy of treatments available affect how useful a predictive test for cancer might be.

Questions to ask in determining suitability for clinical use include [9]**:**

◆ Does the cancer have a potential for risk stratification either through predisposition genes, polygenic score and/or non-genetic risk predictors?

◆ What is the public health burden of the cancer, including its incidence and mortality?

◆ What is already known about a cancer's natural history?

◆ What biomarkers exist for the cancer (to enable detection and characterisation; can include DNA and other biological molecules)?

◆ What are the existing opportunities for targeted interventions to prevent the cancer in high-risk individuals, such as enhanced screening, chemoprevention or surgical prophylaxis?

Appreciation of the complexities of translating new tests to predict or diagnose cancer into improved public and individual health outcomes is essential. Risk identification is not useful in isolation; it must be combined with evidence-based interventions to reduce risk that are effective, safe and acceptable. If people are identified as being at significantly increased risk of cancer, they should have access to clinical care to mitigate this risk of the cancer occurring, or to enable earlier detection and clinical management.

Clinical implementation of any RPMs will benefit from addressing issues such as communication of risk, and the positive or negative impact polygenic score information may have on behaviour of healthcare professionals and their patients[11].

Applications
in cancer

Applications in cancer

The prospect of using polygenic score information in clinical practice to reduce cancer morbidity and mortality is the focus of ongoing research and development. Various approaches and uses are anticipated and are being explored. These range from refining risk prediction, informing screening, facilitating diagnosis, predicting prognosis and guiding therapeutic interventions.

Risk prediction

The best risk prediction model (RPM) for cancer is one that combines the most suitable risk factors to achieve the most accurate prediction[12]. This could include polygenic scores. The number of risk factors incorporated into a RPM will vary depending on the cancer type. Age and sex are strongly associated with cancer risk and are predictive of general cancer risk, so are suitable to be considered for inclusion.

Some cancers have additional well-established, strongly associated risk factors, such as smoking for lung cancer, or family history of disease for breast or colorectal cancer. Other cancers, such as prostate cancer, have a limited number of established risk factors.

Risk prediction models need associated tools - for example a simple user interface - that allow users to collect and enter the information required for the model to calculate the risk prediction.

A small number of risk prediction tools for cancer exist, including QCancer (for the prediction of up to 11 cancers and an overall risk prediction for cancer); CanRisk and IBIS for breast cancer. Although in clinical settings these tools do not currently include polygenic scores, this is being explored in research.

If a polygenic score demonstrates suitable performance as a predictor of risk for a cancer, it could be integrated into RPMs, along with other risk predictors. Including genetic risk factors in comprehensive RPMs has improved risk prediction in several cancers. However, the improvement varies between cancer types, and the magnitude of improvements in predictive ability is contested[13].

The utility of incorporating polygenic scores in RPMs is also context dependent; in some situations, even a slight improvement in risk prediction performance is considered clinically valuable, whereas in others, a moderate improvement may not be of clinical significance[13].

Examples of polygenic scores for cancer include:

Prediction of tumour subtypes or disease progression

Some cancers have several tumour subtypes, some of which are more aggressive than others. A test that could distinguish an aggressive tumour from those likely to be slow growing and causing limited harm, would be of clinical value. It would enable treatments to be targeted to patients most likely to benefit and possibly reduce overtreating patients and and other associated harms. Polygenic scores to determine risk of certain cancer subtypes are being investigated and some have been developed, such as a polygenic score for triple negative breast cancer, an aggressive subtype of breast cancer.

Prostate cancer is another example, where the tumour can be slow growing and the patient needs minimal treatment. However, for men with a more aggressive, potentially lethal, prostate cancer, developing polygenic scores with the specific goal of identifying more aggressive tumours could improve decisions on treatment selection. It could also reduce the risk of men with indolent cancers being overdiagnosed and overtreated.

Cancer risk prediction in organ transplant recipients

Polygenic scores are being investigated as a tool to improve the accuracy of cancer risk prediction in organ transplant recipients. Compared to the general population, this group is at substantially increased risk of developing cancer post-transplant, particularly non-melanoma skin cancer[15]. Non-melanoma skin cancers are treatable if they are diagnosed in good time, and the clinical management of organ transplant recipients who do develop non-melanoma skin cancers may be altered through switching or reducing the doses of anti-rejection drugs. Being able to accurately identify those at highest risk of skin cancer can provide intervention methods such as regular skin checks and potentially altering immunosuppression treatment regimens to reduce the risk of post-transplant malignancy.

Contralateral breast cancer

Due to the high incidence of breast cancer and improving survival rates, an increasing number of breast cancer survivors are at risk of developing contralateral breast cancer (i.e. in the other, previously unaffected breast). The risk is particularly high in women with a heritable genetic predisposition for cancer. A few studies have developed polygenic scores for risk prediction of contralateral breast cancer[16].

PROSTATE CANCER

Identifying aggressive cancers

Prostate cancer is one of the most common cancers in men. It can progress slowly, and men can live for decades without symptoms or needing treatment. However, for some it is a serious disease where it progresses quickly and can lead to death. Currently the biomarkers that are available are not very accurate at predicting which men will develop an aggressive prostate cancer.

Development of a polygenic score with the specific goal of predicting who is likely to develop an aggressive, and potentially lethal prostate cancer, is likely to improve treatment decisions and reduce the risk of men with slow growing cancers being overdiagnosed and overtreated.

Risk prediction in familial or hereditary cancers

Within family history or genetics clinics there are often two groups of individuals; those that have a pathogenic variant identified and those that have a strong family history for a cancer but that do not have a pathogenic variant identified. Polygenic scores in these situations can refine and improve the risk prediction processes already undertaken as part of the management of this high-risk group.

◆ **Refining risk prediction in carriers of known pathogenic variants in cancer predisposition genes.** Specific genetic variants conferring a high-risk of disease are associated with hereditary cancers, for example breast and ovarian cancer predisposition genes *BRCA1* and *BRCA2*. It has been shown that the use of polygenic scores can improve the accuracy of the risk estimate for cancer in carriers of these variants for breast cancer and colorectal cancer[17], which will aid in decisions on risk management.

◆ **Refining risk prediction in individuals with a family history of cancer.** This would predominantly be of value to those that have a strong family history but where no high risk predisposing genetic variants were identified. Polygenic scores can have value by improving the risk prediction.

Improvements in the accuracy of such risk prediction would assist clinicians and patients to make potentially significant decisions regarding preventative strategies – for example, deciding between undergoing a mastectomy versus regular mammography to manage breast cancer risk. The PRiMo trial is investigating the efficacy and feasibility of a personalised risk assessment for breast and ovarian cancer, which includes offering a polygenic score to women undergoing genetic testing at family history cancer clinics.

It is unlikely that polygenic scores developed to predict common cancers in the general population will predict hereditary (familial) cancer syndromes. For example, Lynch syndrome, also called hereditary non-polyposis colorectal cancer, is a familial cancer syndrome caused by variants in a small number of related genes. One study found that, whereas polygenic scores are predictive of colorectal cancer in the general population, they do not predict familial genetic predisposition to Lynch syndrome colorectal cancer cases[18].

If polygenic scores are to be used to help predict cancer risk in the general population, strategies to identify people with inherited genetic variants that confer strong predisposition towards cancer will still be necessary.

Screening

The utility of screening is regularly debated and assessed, primarily to balance the benefits of reducing cancer mortality against harms such as overdiagnosis, false positive results and overtreatment[19].

To improve the risk-to-harm ratio, more selective (targeted or stratified) approaches for eligibility for screening could be achieved by improved risk predictions for cancer. Hence, there is a considerable interest in the development of new cost-effective screening programmes that integrate a greater range of risk factors, including indicators of environmental and genetic risk, and improve identification of asymptomatic individuals with a high-risk of developing cancer. Polygenic scores could contribute to these efforts.

There are many points along the prevention and screening pathway where risk prediction using polygenic scores could be considered. This can include determining the frequency of screening, which screening tests to use, or if to continue testing after a positive screening result.

Polygenic scores have been developed in cancers for which there are existing screening programmes, such as breast cancer. It has been suggested that if polygenic scores are shown to improve current risk prediction methods, then they could also contribute to better outcomes from screening programmes.

The use of cancer risk prediction approaches including polygenic scores within population screening programmes has been proposed for risk stratified screening and current screening tests.

Risk stratified screening

Risk stratified screening whereby a personalised risk assessment determines:

◆ eligibility for screening

◆ age at which screening will be offered, based on a risk assessment done earlier than the current screening age

◆ type, amount or level of screening offered based on risk

◆ identification of those at the highest risk (for example, of the top 2%)

There is increasing evidence that risk stratified approaches could result in a lower rate of false-positive results and over-diagnosis of disease, thereby improving efficiency

Globally there are currently no risk stratified screening programmes, but there is increasing evidence that risk stratified approaches could result in a lower rate of false-positive results and over-diagnosis of disease, thereby improving efficiency[19]. Combining a risk stratified screening approach with better risk prediction could improve the overall performance of screening programmes.

The use of comprehensive RPMs that include genetic risk factors, including polygenic scores, demographic and lifestyle risk factors, are the focus of the following prospective evaluations and trials of screening populations:

◆ WISDOM and MyPebs trials for breast cancer

◆ BARCODE pilot trial for prostate cancer

◆ SCRIPT trial for colorectal cancer

Ultimately, the routine adoption of a polygenic score within a risk stratified screening approach will depend on the clinical impact and cost-effectiveness of the relevant screening programme. Initial economic analyses looking at the impact of risk-stratified strategies on screening programmes indicate this approach could be cost-effective, but that this would rely on the use of accurate RPMs[20].

An improved RPM would more accurately identify those suitable for screening compared to existing approaches. It remains to be determined whether the inclusion of polygenic scores does improve the performance of RPMs for use in screening.

Numerous research studies are exploring multiple aspects pertaining to the feasibility and acceptability of personalised risk assessment in the general population, including the Personalised RISk-based MAmma screening study (PRISMA), the Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA), and the Predicting the Risk of Cancer at Screening study (PROCAS).

Improving performance of current screening tests

It has been proposed that polygenic scores could be combined with an existing population screening test, such as PSA testing (prostate cancer), or the faecal immunochemical test (FIT) (colorectal cancer)[21], to improve screening strategies. This could be by combining a polygenic score with a population screening test, where a positive screening test (for example PSA or FIT) in individuals with a high polygenic score triggers the use of the more expensive and/or invasive confirmatory test, e.g. biopsy or colonoscopy.

Such a strategy would need to be clear about:

◆ which polygenic score test to use

◆ how to use it with the current screening test

◆ its effectiveness within a screening programme

There are also indications that polygenic scores have potential when used in combination with other risk factors and biomarkers. Combining the existing screening test with a polygenic score test could identify individuals who would not be detected by existing screening programmes. For example, one study showed that a high polygenic score was associated with an increased risk of being diagnosed with breast cancer in the period between screening tests (interval breast cancer) [22]. When compared to breast cancer diagnosed during routine screening, interval cancers are more aggressive and associated with poorer prognosis, so identifying these individuals could have significant benefits.

Guiding medical interventions

Informing risk prediction of subtypes and selection of therapeutic interventions

The use of polygenic scores has been suggested as a possible method of identifying individuals who may benefit from specific preventative drug treatments[23].

Risk-reducing medications, such as tamoxifen for ER-positive cancer, are mainly prescribed to women with a breast cancer diagnosis. However, tamoxifen, and similar drugs, could also be used to prevent the disease in asymptomatic women at high-risk. If these women can be identified they could be offered the drug as a preventive therapy. Polygenic scores to identify individuals at risk of ER-positive, HER2-positive or HER2-negative breast cancers are being developed.

Health promotion

There is interest in the use of cancer risk tools to aid individualised health promotion activities by providing personalised information on risk. This would be relevant across the whole prevention pathway.

In practice, however, it is difficult to achieve improvements in individuals' health related behaviours even after receiving personalised risk information; this also applies to other common conditions such as cardiovascular disease. Nevertheless, the availability of risk prediction tools that provide accurate and accessible information on risk can facilitate conversations with individuals who are then motivated to act on this information[24]. One study, using 20 cancer site-specific polygenic scores, demonstrated it was possible to identify individuals at highest overall risk of cancer and to facilitate decision making about lifestyle modifications for personalised prevention[25].

Towards implementation

Towards implementation

Professional bodies do not currently endorse the use of polygenic scores in a healthcare setting. For example, the US National Comprehensive Cancer Network's (NCCN) Guidelines advise against providing polygenic risk scores outside of clinical trials and state that 'there are significant limitations in interpretation of polygenic risk scores'[26]. Wider implementation of polygenic scores will require sufficient understanding of how they will affect clinical care in specific contexts.

Despite the challenges and uncertainties, direct to consumer companies are known to be interested in, or are already using, polygenic scores. These include Myriad (breast cancer), 23&me and Ambry Genetics. However, Ambry Genetics suspended use of polygenic scores in early 2021 until they are suitable for 'pan-ancestry' use and Myriad Genetics is recalibrating their breast cancer polygenic score for all ancestries. Other companies, namely Genomics Plc and Allelica, are providing polygenic score services to analyse existing genetic data, primarily for research.

As costs of genotyping and sequencing decrease, it will probably become more feasible to use polygenic score prediction of disease risk in a clinical setting[27].

There will be different options for test delivery. It is possible that one specific polygenic score test based on predetermined SNPs for a particular cancer is offered. Or when whole genome sequencing is used, a larger number of SNPs will allow for a wider range of polygenic scores to be analysed. This means that an individual could receive a range of cancer risk predictions based on polygenic scores in one report. The implications of using polygenic scores in this way remain to be determined.

It is possible that one specific polygenic score test based on predetermined SNPs for a particular cancer is offered. Alternatively a larger number of SNPs are tested, allowing for analysis of polygenic scores for different diseases. In this instance, an individual could receive a single report identifying polygenic score-based risk predictions for several cancers

To gather further evidence of the potential utility of polygenic scores in healthcare, additional research efforts are underway:

◆ University Hospital Tübingen's Germany project, Ge-Med, will replace diagnostic exome sequencing with whole genome sequencing across a variety of genetic conditions, including familial cancer syndromes and rare genetic disorders and will explore the use of polygenic scores in a routine clinical setting.

◆ Melanoma Genomics' Managing Your Risk Study, a randomised control trial looking at risk of melanoma to evaluate the impact of personal melanoma genomic risk information on sun-related behaviours and psychological health in Australia.

◆ The Our Future Health project in the UK is a large health research programme that will include research into the implementation of polygenic scores for the risk prediction of chronic diseases.

◆ Partnerships with private companies and healthcare organisations. The Institut Curie in France is working with PrediLife, a company developing risk prediction solutions, to provide breast cancer prediction using five risk factors, including polygenic scores.

Others are working to make access to polygenic score methods easier for healthcare professionals. This includes developing tools for educational and clinical purposes as well as enabling interpretation of the scores, such as GenoPred, MyGeneRank and ImputeMe.

The Cancer PRSWeb is an extensive online repository based at the University of Michigan that integrates freely available GWAS summary statistics to develop over 500 polygenic score models for 35 common cancer traits[2]. It aims to facilitate scientific collaboration on polygenic score research, with the ultimate goal to provide absolute risk metrics for each individual in electronic health records, thereby translating GWAS findings to inform patient care. The objective is for the physician to have easy access to this information to inform cancer management decisions for their patients.

Summary

Summary

Polygenic scores are being widely promoted as having the potential to contribute significantly to the prevention, identification and management of cancer, and there is considerable research and development underway for a variety of cancers. Whilst there are significant issues still to be addressed, such as generalisability of scores, there are indications that polygenic scores can improve risk prediction in some cancers and clinical contexts.

One area of particular interest is commonly diagnosed cancers that have existing screening programmes, primarily to determine if a polygenic score can improve the clinical effectiveness and efficiency of such programmes. How polygenic scores might contribute to a stratified screening strategy, particularly as part of a cancer risk prediction model, is also an important question. Trials are underway to determine the role that stratified screening using comprehensive risk prediction models (that include polygenic scores) could have as part of screening programmes.

The interest of using polygenic scores in screening of common cancers means they could potentially impact a large number of people. However, there are also other rarer cancers for which polygenic scores may also be valuable, especially considering cancers where there are a limited number of known risk factors and biomarkers.

The development, validation and regulation of new tests, as well as infrastructural needs and evidence of clinical utility are still to be fully ascertained before wider implementation becomes an option. Premature implementation of polygenic scores in cancer risk estimation approaches is likely to undermine these efforts, and risk loss of confidence in this potentially valuable area of health development.

This is an exciting area of research that could have potential in specific contexts and for specific purposes but further sustained research and translation efforts are required to adequately assess the potential role of polygenic scores in improved cancer prediction, prevention and management.

Can a polygenic score improve the clinical effectiveness and efficiencies of existing screening programmes for commonly diagnosed cancers?

References

1. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May. 71(3):209-249. doi: 10.3322/caac.21660.

2. Fritsche LG et al. Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. The American Journal of Human Genetics, 2020. 107(5): p. 815-836.

3. Lindstrom S et al. Quantifying the Genetic Correlation between Multiple Cancer Types. Cancer Epidemiol Biomarkers Prev, 2017. 26(9): p. 1427-1435.

4. Lambert SA et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nature Genetics, 2021.

5. Sud A, et al. Will polygenic risk scores for cancer ever be clinically useful? npj Precision Oncology, 2021. 5(1): p. 40.

6. Zhang YD et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. Nat Commun, 2020. 11(1): p. 3353.

7. Liyanarachchi S et al. Assessing thyroid cancer risk using polygenic risk scores. Proc Natl Acad Sci USA, 2020. 117(11): p. 5997-6002.

8. Martin AR et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet, 2019. 51(4): p. 584-591.

9. Turnbull C, et al. Cancer genetics, precision prevention and a call to action. Nat Genet, 2018. 50(9): p. 1212-1218.

10. Graff RE et al. Cross-cancer evaluation of polygenic risk scores for 16 cancer types in two large cohorts. Nat Commun, 2021. 12(1): p. 970.

11. Brigden T et al. Implementing polygenic scores for cardiovascular disease into NHS Health Checks. PHG Foundation report, 2021: p. 57.

12. Kachuri L et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. Nat Commun, 2020. 11(1): p. 6084.

13. Moorthie S et al. Polygenic scores and clinical utility. PHG Foundation report, 2021.

14. Klein RJ et al. Prostate cancer polygenic risk score and prediction of lethal prostate cancer. npj Precision Oncology, 2022. 6(1): p. 25.

15. Stapleton CP et al. Polygenic risk score of non-melanoma skin cancer predicts post-transplant skin cancer across multiple organ types. Clinical Transplantation, 2020. 34(8): p. e13904.

16. Kramer I et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. Am J Hum Genet, 2020. 107(5): p. 837-848.

17. Gallagher S et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. JAMA Network Open, 2020. 3(7): p. e208501-e208501.

18. Jenkins MA et al. Assessment of a Polygenic Risk Score for Colorectal Cancer to Predict Risk of Lynch Syndrome Colorectal Cancer. JNCI Cancer Spectrum, 2021. 5(2).

19. Lebrett MB et al. Targeting lung cancer screening to individuals at greatest risk: the role of genetic factors. J Med Genet, 2021. 58(4): p. 217-226.

20. Pashayan N et al. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncology, 2018. 4(11): p. 1504-1510.

21. Arnau-Collell C. et al. Evaluating the Potential of Polygenic Risk Score to Improve Colorectal Cancer Screening. Cancer Epidemiology, Biomarkers & Prevention, 2022: p. OF1-OF8.

22. Li J et al. Breast cancer genetic risk profile is differentially associated with interval and screen-detected breast cancers. Ann Oncol, 2015. 26(3): p. 517-22.

23. Kim JO et al. Impact of a breast cancer (BC) polygenic risk score (PRS) on the decision to take preventive endocrine therapy (ET): The Genetic Risk Estimate (GENRE) trial. Journal of Clinical Oncology, 2019. 37(15_suppl): p. 1501-1501.

24. Usher-Smith JA et al. Effect of interventions incorporating personalised cancer risk information on intentions and behaviour: a systematic review and meta-analysis of randomised controlled trials. BMJ Open, 2018. 8(1): p. e017717.

25. Zhu M et al. Genetic Risk for Overall Cancer and the Benefit of Adherence to a Healthy Lifestyle. Cancer Res, 2021.

26. NCCN Updates Guidelines Advising Against Using Polygenic Risk Scores Outside of Clinical Trials. Precision Oncology News, 2020.

27. Lello L et al. Genomic Prediction of 16 Complex Disease Risks Including Heart Attack, Diabetes, Breast and Prostate Cancer. Sci Rep, 2019. 9(1): p. 15286.

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

intelligence@phgfoundation.org

PHG
FOUNDATION

making science
work for health

UNIVERSITY OF
CAMBRIDGE