



Predicting the likelihood of common diseases using polygenic risk scores and the possibility of adverse selection

The latest advances in genomic science are permitting more and more accurate predictions of disease risk

Developments in genetics research over the past decade have enabled an extraordinary expansion in the understanding of the human genome and the genetic causes of common diseases. Early disease detection and disease prevention are critical for extending human longevity, so it follows that continued advances in predicting the likelihood of disease and the advent of personalised treatments – based on genetics – will ultimately lead to improvements in mortality and healthy longevity.

On the other hand, life insurance availability depends on an insurer's capacity to accurately evaluate mortality risk and charge an appropriate premium. The growth in direct-to-consumer (DTC) genetic testing and obtainability of genetic risk information allows consumers to access information about their disease risk that, if non-disclosed or disallowed, might start to signal an imbalance in the availability of critical risk information between life insurers and consumers.

WHAT ARE POLYGENIC RISK SCORES?

Until quite recently, our genetic understanding of the causes of different diseases centred on single, high-penetrance genetic mutations, which cause Mendelian or monogenic disease. But most common diseases have a polygenic architecture, meaning that many genetic variants in combination influence disease risk. Furthermore, major diseases such as cancer, coronary artery disease and diabetes are multifactorial with both genetic and environmental factors affecting the likelihood of suffering from the disease in the future.

MANY GENETIC VARIANTS IN COMBINATION INFLUENCE DISEASE RISK

Genome-wide association studies (GWASs) have been a hugely important tool for scientists to study differences in the DNA sequences of individuals with a disease compared to individuals without the disease. Since 2005, GWASs have been used to identify the most common genetic variants that impact an individual's genetic predisposition to a given disease. These variants are known as single nucleotide polymorphisms (SNPs) – places in the genome at which some individuals in the population have a different DNA nucleotide than the one normally found in that position.

Genetic risk scores, known as polygenic risk scores (PRSs), aggregate risk information from the most important SNPs into a single score that describes an individual's genetic predisposition to a given disease. PRSs may include tens, hundreds, thousands or even millions of SNPs. The PRS simply adds up all the disease-specific SNPs, weighted by their effect size, to produce an overall risk score of developing the given disease. PRSs are often expressed as a percentile: in individuals with a PRS close to the 50th percentile, the genetic risk will be similar to the population's risk, but a person with a PRS in the 91st to 100th percentile would have a much higher than average genetic risk.

THE UK BIOBANK AND POLYGENIC RISK SCORES

PRSs are already proving a powerful means to identify individuals at increased risk of developing certain morbidities, but just how



informative are PRSs for predicting diseases? Researchers are now using large-scale cohort studies like the UK Biobank to explore genetic data alongside traditional health data in order to determine the value of each type of information in predicting diseases.

The UK Biobank is a study of 500,000 participants across the United Kingdom, which was established to identify the determinants of common life-threatening and disabling conditions. The study captured clinical, genetic, environmental, sociodemographic and biomarker information on all participants, including more than 20 million genetic variants.

Researchers from RGA, in collaboration with a team of academics from King's College London, have been studying UK Biobank genetic, environmental and clinical data to determine the predictive value of genetic data alongside other risk information. This research collaboration focuses on assessing the utility of PRSs to predict incidence of diseases after controlling for typical risk factors used in underwriting, such as BMI, blood pressure, smoking and family history.

In the RGA-King's College London research, the UK Biobank cohort was divided into two subgroups based on participants' medical histories. Risk prediction models were then built to study coronary artery disease and breast cancer outcomes in each subpopulation, adjusting for relevant biometric, lifestyle and socioeconomic factors. The results demonstrated an approximately twofold increased risk for those in the top 5% of genetic risk compared to those in the middle (40th to 60th percentiles) for both diseases – which was true for both subgroups. These results establish the importance of PRSs in risk differentiation, alongside, and largely independent of, traditional underwriting risk factors. The study has now been published in the *Annals of Actuarial Science*.¹

REGULATIONS AROUND THE WORLD CONTINUE TO LIMIT THE USE OF GENETIC DATA IN LIFE

POLYGENIC RISK SCORES HAVE GREAT POTENTIAL, BUT GENES ARE NOT FATE

The application of PRSs in clinical medicine holds fantastic potential. Perhaps the most imminent application will be to identify people with higher cancer risk, which could lead to personalised screening programs. A study by Mavadatt and colleagues² developed a PRS for breast cancer risk and showed that women with a PRS in the top 20% had a 17.2% lifetime incidence risk of developing the disease

compared to a 5.3% risk for women with a PRS in the bottom 20%. Earlier or more frequent breast cancer screening might therefore be recommended to those women with higher-than-average genetic risk.

Genes, however, are not fate. A study by Khera and colleagues³ demonstrated that even if a person is shown to have a high PRS for adverse cardiovascular outcomes, the overall risk of disease can be greatly mitigated by adhering to a healthy lifestyle. Identifying those at high genetic risk might prove helpful for targeted public health campaigns and lifestyle interventions.

INSURANCE IMPLICATIONS AND CONCLUSIONS

The use of genetic information in insurance will continue to be a highly sensitive subject for life and protection insurers. Regulations around the world continue to limit the use of genetic data in life, critical illness and other products, so PRSs and other genetic-based risk tools should be considered an emerging risk for our industry. In particular, the rapidly reducing costs of DNA sequencing technology coupled with the growth in direct-to-consumer genetic testing poses immediate concerns around information asymmetry (since it is highly unlikely that PRSs or other genetic risk scores could be used in underwriting) and increased anti-selective behaviours.

At present, anti-selection due to knowledge of genetic risk information is believed to be relatively low, but growing access to this information, through PRSs and other advances, might soon affect insurance-purchasing decisions. The key concern for insurers, therefore, is the adverse financial impact this might have. Yet unquestionably, this is also a time for optimism and opportunity: genetics research continues to grow our understanding of human health, and genomic medicine developments will soon follow, almost certainly bringing improvements in mortality and longevity. ■

Dr. R. Russell is Lead Health Data Scientist in the Global Data & Analytics department within RGA. Richard is a highly experienced Data Scientist, providing specialist resource to enhance RGA's mortality and morbidity modelling capability.



1 – <https://www.cambridge.org/core/journals/annals-of-actuarial-science/article/multifactorial-disorders-and-polygenic-risk-scores-predicting-common-diseases-and-the-possibility-of-adverse-selection-in-life-and-protection-insurance/CDBF32A426F41C65E35DDAC542FED30F>

2 – <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4754625/>

3 – <https://www.nejm.org/doi/full/10.1056/nejmoa1605086>