Actuaries Summit
Think Differently

21-23 May 2017
Grand Hyatt Melbourne

Actuaries Institute

Thinking about life insurance through a genetic lens

Prepared by Dr Damjan Vukcevic & Jessica Chen

Presented to the Actuaries Institute
Actuaries Summit
21 – 23 May 2017
Melbourne

Institute of Actuaries of Australia
ABN 69 000 423 656
Level 2, 50 Carrington Street, Sydney NSW Australia 2000
t +61 (0) 2 9239 6100 **f** +61 (0) 2 9239 6170
e actuaries@actuaries.asn.au **w** www.actuaries.asn.au

Thinking about life insurance through a genetic lens

Damjan Vukcevic & Jessica Chen

Abstract

Genetic research is booming. The last decade has seen the discovery of thousands of genetic variants that confer greater risk for various common diseases, such as cancer and heart disease. While this has been a breakthrough for medical research, what implications does it pose for life insurance? The ability for individuals to learn more about their long-term health based on a genetic test raises the prospect of greater anti-selection risks.

To help life insurers navigate this new landscape, here we summarise the latest in genetics research as it relates to predicting individual disease risk and analyse the potential impact of genetic testing on the industry. We show that genetic risk prediction is already powerful, but will only become a potential material threat if, or when, such testing becomes more widespread. Finally, we explore some ideas on how we can think differently about evolving the practice of life insurance to cope with greater personal knowledge.

Keywords: genetics, polygenic risk score, life insurance, trauma insurance, anti-selection

Acknowledgements

We wish to thank Martin Mulcare for his time and insights in peer reviewing our paper.

Disclosure

Dr Damjan Vukcevic is employed by The University of Melbourne and the Murdoch Children's Research Institute. Jessica Chen is employed by BT Financial Services. The opinions outlined in this paper are their own and do not necessarily represent the views of their employers.

Table of contents

1. Introduction

The Economist asks, "How has DNA shaped the human race?" (Palmer and Rutherford, 2016)

We ask, "How will DNA shape life insurance?"

Modern-day genetic research has uncovered thousands of genetic variants that are associated with greater risk of many common human diseases, such as cancer, heart disease, Alzheimer's and diabetes. These variants are generally common throughout the population, with all individuals carrying at least some of the 'high risk' genetic variants. This is distinctly different from the previous generation of genetic discoveries, which focused on rare variants and rare diseases, and which form the basis for genetic tests that are currently used in clinical practice.

These advances in genetic knowledge, together with a reduction in the cost of genetic testing, raise the potential for individuals to gain better insights into their current and future health. For life insurance companies, such advances might be expected to increase anti-selection risks and may even challenge the way in which insurance risk is considered and managed.

In this paper, our aim is to provide:

1. A summary of the **latest in genetics research**, focusing on how it can be used to **predict disease risk for individuals** (Section 2). We shine a spotlight in particular on the key diseases that lead to the majority of life insurance claims (Section 4).

2. An overview of the **current state** of the **Australian life insurance industry**, with particular regard to disclosure requirements on genetic information when applying for insurance (Section 3).

3. An analysis of the **'threat' of genetic testing** to the life insurance industry, based on the latest genetic knowledge (Section 4).

4. Some ideas on how insurers can **think differently** to respond to the challenge of customers with **greater knowledge of their personal health risks** (Section 5).

2. Genetics & disease risk prediction

In this section we summarise the progress of genetic research as it relates to the prediction of disease risk for individuals. We begin with a brief introduction to the main concepts in genetics that are necessary to understand this paper, followed by a longer overview of risk prediction using genetic variants. Finally, we discuss some current trends in research and where it might be headed in the future. In the Appendix we provide a more detailed introduction to genetics, as well as a historical perspective on the study of genetic risk factors for diseases.

2.1 Genetics

Genetics is the study of genetic variation and how it influences biology, including physical traits and susceptibility to diseases. Of particular interest is DNA, which are molecules that record and carry genetic information and are present in nearly all of the cells in our body. These form the 'blueprint' for our bodies, encoding

instructions—known as *genes*—for how and when to make certain proteins, and thus are a primary determinant of our physical traits.

Genetic information is inherited by children from their parents through the process of sexual reproduction. In this way, children end up with similar traits to their parents. Skin and eye colour are two readily apparent examples, but other such traits include height, body shape and propensities for various diseases such as cancer and diabetes.

Other factors also play important roles in our development, such as nutrition, social support and education. These are collectively referred to as *environmental* factors. Colloquially, we often speak about 'nature vs nurture', contrasting the genetic factors, which we are born with and cannot change, with environmental factors, which we have an opportunity to shape irrespective of our genes. The extent to which one or the other is more influential for determining a given trait varies across traits. The balance between them can be quantified in various ways; for example, the *heritability* is the proportion of the variance in the trait that is attributable to genetic factors, and the *familial relative risk* is the increase/decrease in risk of a disease for an individual who has an affected family member (typically only looking as far as parents and siblings) compared to the average risk in the population.

Genetic research has made rapid progress over the last few decades. For example, we now know the exact genes that cause a large number of rare but highly heritable diseases. These have led to the development of genetic tests that allow individuals to determine if they carry the disease-predisposing variants to help them make decisions about their future health and/or their children and potential children. Such tests are now routinely available and life insurers already incorporate them as part of underwriting their policies.

More recently, technological improvements have allowed us to measure DNA ever more cheaply and have enabled very large genetic studies—known as *genome-wide association studies* (GWAS)—of numerous diseases, leading to discoveries of thousands of variants associated with disease risk. Unlike in earlier studies, these variants and the diseases they relate to are both common in the population. This has raised the prospect that, via a genetic test, we could predict the predisposition to a whole range of diseases, even from birth. Such a development would have important consequences for how we shape medical care, public health policy, and also life insurance.

While such radical changes are yet to occur, we have already seen the formation of companies such as 23andMe (https://www.23andme.com/) that offer world-wide direct-to-customer genetic testing services, with disease risk prediction included. Some health agencies around the world are also investigating the potential utility of population genetic screening along similar lines. It is early days, with many ethical and practical concerns yet to be resolved before anything like this can become routine. Meanwhile, genetic research has continued apace and shows no sign of slowing down. It is indeed a realistic prospect that, in the longer term, the efforts of health agencies and genetic researchers will coalesce and give rise to genetic testing services that are ubiquitous and informative of future health.

2.2 Risk prediction based on genetics

We now turn to the topic of using genetic information from individuals to predict their risk of various diseases. In particular, we mean using actual measurements of the individuals' DNA, rather than simply using their known family histories or the results of genetic tests from relatives. The findings from genetic studies of diseases can be used to create predictive models for this purpose.

To what extent such models are useful in practice will vary substantially from one disease to another. For monogenic diseases (those controlled by a single gene) that have been thoroughly characterised, such a model will be definitive: it will tell you exactly whether or not you carry the risk variant(s) and whether you will develop the disease. Indeed, such tests are readily available as part of standard clinical care and are routinely used to confirm medical diagnoses and for screening suspected carriers.

For complex disorders (those controlled by many genes), despite the large number of recent discoveries, risk prediction is generally much less conclusive (with some notable exceptions). It is also still an active area of research. The predictions for a given individual can change substantially over the space of a few years as our scientific knowledge progresses, both in terms of our knowledge of the genetic factors at play and the development of more effective predictive modelling approaches.

2.2.1 Why good risk prediction is not an immediate consequence of successful association studies

The success of GWAS has been their ability to find replicable associations between disease traits and specific genetic variants. This is useful for understanding the biological mechanisms of the disease. However, it will only be useful for predicting disease risk if those particular variants lead to big changes in risk. In other words, it requires that the *magnitude* of the associations be large, not just the *strength of evidence* for them (usually measured by their *statistical significance*). It is the difference between showing that smoking leads to increased risk of lung cancer, and actually giving a prediction for a given smoker or non-smoker.

By and large, variants identified by GWAS confer only very modest changes in risk (e.g. relative risks of less than 1.1), and are thus not very predictive on their own. (The studies can nevertheless identify them and amass strong evidence by using very large sample sizes.) Also, unlike for monogenic disorders, these variants will typically not be the actual causal variants but would instead be correlated with them. This reduces the predictive power, as compared to what would be possible if we knew the causal variants.

This highlights the differing aims of association studies and risk prediction:

- Association studies focus on a **specific trait**, such as a disease or a disorder. The aim is to further the **scientific knowledge** about that trait by detecting genetic variants associated with it, usually according to a high standard of evidence. This is to ensure that the finding is likely to be replicated and we have high degree of certainty in the association. Establishing causality is desirable, although not often done as part of the initial study.

- Risk prediction focuses on **individuals**. The aim is to maximise **predictive accuracy** when using genetic variables to infer the occurrence of a trait, such as a disease or disorder, for an individual. Which genetic variants are used does not matter, as long as the predictions are shown to be robust, reliable and useful for making decisions.

More generally, there's a difference between being able to show that a particular trait or disease is heritable (i.e. has a genetic component) and being able to use genetic measurements to actually usefully predict the trait for a specific individual. The latter is much more challenging because it requires detailed knowledge about the genetic variants involved, and that these are strongly predictive on their own. Estimates of heritability allows us to understand the limit of genetic risk prediction, and give us a gauge on how much there is left to discover.

2.2.2 Approaches to genetic risk prediction for complex traits

The earliest, and still commonly considered, approaches to risk prediction involve creating regression models using the genetic variants from published GWAS. Typically, these will be the *single nucleotide polymorphisms* (SNPs) that are declared to be 'genome-wide significant' by the study; this is a strict measure of statistical significance (most often defined as having a p-value less than 5×10^{-8}) that is tailored to this type of study and has been widely adopted in the field.

The simplest approach is to take all such SNPs, along with their published estimates of effect on disease risk, and combine them together in a single regression equation. Going one step further would be to re-estimate the risk coefficients by fitting a joint model with all SNPs together. The latter requires having access to the study data, whereas the former can be done by anyone simply with access to the published literature. Both approaches assume that each SNP acts independently to modify disease risk, which turns out to be an adequate assumption in most cases.

More recently, it has been shown that using greater numbers of SNPs, including those that are not highlighted as having the strongest evidence of association, can lead to greater predictive accuracy (Goldstein et al. 2015). The rationale is that many such SNPs will capture true genetic effects (the 'signal') that are simply too weak to reach genome-wide significance. They will be mixed in amongst SNPs that simply appear to have a strong effect by chance (the 'noise'). If the amount of signal is greater than the noise, then there will be benefit in their inclusion.

The techniques that use this idea typically employ some type of regularised regression, such as lasso or ridge regression (Spiliopoulou et al. 2015). Such techniques allow the use of a large number of predictor variables, while constraining them to prevent *overfitting* (the tendency of some models to fit the data too closely and not thus not generalising well to new data).

Each of these models can be used to produce a *polygenic risk score* (PRS), sometimes referred to as a *genomic risk score* (GRS). This is a single number that summarises the contribution of all genetic variants to the trait of interest, according to the model. For example, for a disease (a binary trait), the higher the score the greater the risk of the disease. The advantage of this formulation, rather than a direct mapping to risk probabilities, is that it allows further modelling and analyses. For example, it could be combined together with environmental and lifestyle variables to create a more powerful predictive model.

As with any exercise in predictive modelling, the models need to be calibrated and externally validated before attempting to use them in practice (Abraham and Inouye 2015; Chatterjee, Shi, and García-Closas 2016).

2.2.3 When are genetic tests useful for risk prediction?

There are various reasons to order a genetic test. For example, it might be to: confirm a medical diagnosis based on observed symptoms; determine whether or not you carry the genetic variant(s) that cause a genetic disorder that runs in your family; help decide on which treatment option will be most effective for you for a given disease; screen for likely diseases early in life to treat or prevent them before they develop or become too severe.

To make a test useful for these purposes requires: (i) that the test has enough predictive power to lead to conclusive outcomes or clear decisions; and (ii) that these outcomes are desirable to know or that useful decision options actually exist.

Tests for monogenic diseases will generally be useful in this regard: the outcome will clearly indicate the presence or absence of the causal risk variants, and thus also where an individual will develop the disease. This can then guide options for medical treatment and other personal decisions such as whether to have children.

The usefulness of tests for complex diseases is not so clear cut. The tests will vary both in how predictive they are and whether or not a prediction would actually make a difference to any medical or lifestyle decisions.

We first examine the issue of predictive power. The majority of discovered genetic variants for complex diseases have very small effect. That means that they will not be useful on their own and would need to be combined together into a polygenic risk score (see above) before they have a chance of usefully discriminating between individuals based on risk. How much discrimination is possible varies across diseases. For example, in Section 4 we survey a few important diseases; for coronary artery disease, a genetic test is as informative as knowing a few of the important lifestyle variables, whereas for depression the predictive power is more modest.

A related issue is how much an individual already knows about their risk. For some diseases with a substantial genetic component, if you know the person's family history of that disease then the additional benefit of measuring the same person's genetic variants might be quite minor. This depends on how discriminative the risk prediction models are. For example, for a monogenic disease (where we have essentially perfect discrimination) such testing can distinguish carriers from non-carriers. For complex diseases where we still only know a small number of genetic variants at play, the test may just tell the subject they have higher than average risk, which they already knew; or it may not even do that, if the risk variants at play in the family are not ones that are currently known.

Similarly, for many diseases, the total effect of known genetic variants on disease risk will be smaller than known environmental or modifiable lifestyle risk factors, such as diet, exercise, smoking and alcohol intake. Thus, simply knowing the latter is enough to give a good picture of disease risk that will likely not change much given any genetic information. Moreover, the non-genetic variables might be more relevant with regard to health interventions, further marginalising the need for a genetic test.

For these reasons, risk prediction for complex diseases based on genetic tests is generally not yet a useful tool. Nevertheless, there are at least two situations in which they can be strongly predictive:

1. When there are a large number of genetic variants of small effect, often they occur mostly independently, thus creating a wide distribution of underlying genetic disease risk. Most individuals will be in the middle of such a distribution, but there will be a few unlucky individuals in the upper tail of the distribution: those who happen to inherit the high-risk versions for a large number of those variants. They can end up having a disease risk that is much higher than the general population (e.g. 20% higher).

2. For some diseases, particular genetic variants may in fact have a very strong effect and thus be highly predictive on their own. Two prominent examples are breast cancer and Alzheimer's disease, certain variants of the *BRCA1* and *BRCA2* genes confer very high risk for the former and variants of the *APOE* gene do the same for the latter, although for both diseases there are also many other, independent causal factors (many of which are still unknown).

If we discover large numbers of rare variants with substantial effect, as we hope to do in future studies (and especially if we manage to uncover the actual causal variants), this will improve the usefulness of risk prediction because we will be able to give much stronger indications of risk for certain individuals (similar to situation 1, above). For example, contrast the following two hypothetical scenarios for the known relationship between genetics and the prevalence of a certain disease.

In the first, we have a single common genetic variant associated with disease risk. Half the population carry the low-risk version and half carry the high-risk version. The prevalence of the disease is 1% in the low-risk half and 2% in the high-risk half, giving 1.5% overall.

In the second scenario, imagine we know 20 high-risk rare variants associated with disease risk. Each of these variants occurs independently in 0.1% of individuals. If you have at least one high-risk variant, your chance of getting the disease is 76%, but otherwise is zero. The overall prevalence is 1.5%.

In the first scenario, a genetic test is useless because it will just tell you whether or not your risk is higher or lower by 1 percentage point. In the second, a genetic test would tell you whether you are in the disease-free majority or the minority with very high risk.

One issue worth bearing in mind is that many of the current research findings are based on studies of Europeans. Therefore, their results may not necessarily transfer well to other populations; a fact that has recently been demonstrated empirically (Martin et al. 2017). This issue will only become more pronounced when it comes to studying rare variants, which are more likely to be population-specific.

2.2.4 Direct-to-customer genetic testing

Since 2007, a number of companies have offered genetic testing services directly to customers, rather than via medical professionals or genetic counsellors. Their products were enabled by the development of cost-effective SNP genotyping arrays and the success of GWAS at discovering genetic variants associated with disease risk.

The most well-known such company is 23andMe, which continues to operate to this day. Other companies that were once prominent, but have since been acquired and have discontinued their service, include deCODEme and Navigenics.

These services enable individuals to access information about their disease risks based on their genomes, according to the latest genetic research. In fact, a notable feature of the service is that the risk reports provided to customers can be updated over time as new knowledge comes to light, including adding risk predictions for diseases and conditions that were previously not reported on. However, this also makes these risk reports unstable, which arguably undermines their validity (Krier et al. 2016). Furthermore, the fact that different companies report quite different risks for the same individual (due to the use of different sets of genetic variants, and differing assumptions in the risk models) raises similar questions about validity (Kalf et al. 2013).

The provision of this information directly to individuals, without medical oversight, has been controversial. Medical professionals will generally discourage the use of these services, with the American College of Medicine Genetics issuing a public statement that such 'home kits' should not be used, and much longer statement from the American Society of Human Genetics cautions that such tests need to be carefully regulated. In 2013, the U.S. Food and Drug Administration (FDA) famously ordered 23andMe to stop providing health-related reports to customers until they get FDA approval. As of April 2017, the company finally obtained approval to provide reports of genetic risk for 10 specific diseases (including Parkinson's disease, late-onset Alzheimer's disease and celiac disease). Moreover, the FDA announced provisions for much easier approvals for such tests in the future, meaning that we are likely to see many more of these become available over time.

2.3 Future developments

Genetic research is progressing rapidly, driven by past successes, technological improvements, and strong institutional support. The fact that GWAS have finally made inroads into the problem of complex disease genetics has spurred many lines of 'follow-on' and 'copy-cat' research. Genetic technology companies are innovating at a pace that is faster than the famed Moore's Law that drove the computer revolution (Schaller 1997), allowing larger and ever more extensive study of genetic variants. Collaboration between researchers around the world, particularly the formation of large consortia and 'biobanks' (very large, curated datasets), has led to studies with the power to find the many faint genetic signals that contribute to disease risk.

All of these trends are likely to continue for the foreseeable future. An important factor is the large amount of public and private financial investment in this endeavour. Genetics is seen as the 'next big thing' in medicine, enabling more targeted and timely treatment and prevention due to greater knowledge of individuals' circumstances, a vision known as *personalised medicine* (see the Appendix for more information).

With regard to scientific discovery, the future looks rosy. We are even able to scientifically model the progress of discovery, based on known aspects of genetics and disease processes, to make concrete predictions for planned and future studies (Chatterjee et al. 2013). For the immediate future, simply amassing very large datasets is enough to make progress, and this is happening in many ways.

For risk prediction, we can temper our expectations somewhat because of known limitations: most diseases are due to both genetic and environmental factors, and the predictive power of purely genetic tests is limited by the size of the genetic component. This gives a guide as to the potential or likely accuracy attainable based on progress using our existing study designs. To go beyond this, the next revolution will be driven by more extensive study of the interaction between the two types of factors. Some of this might happen through the latest biobanking projects and some may require future developments in data collection.

There is no doubt that the science of genetics will continue to strike gold. The extent to which these advances will make an impact on society, and the speed with which this will happen, might be more dependent on how fast we can change our institutions and norms to adapt to them than the speed of scientific progress.

With regard to genetic risk prediction, we already know that for certain common diseases this can be helpfully used to improve the targeting of screening and prevention strategies (see Section 4 for examples). Whether widespread genetic screening becomes the reality now depends much more on economic and logistical considerations, as well as ethical and privacy considerations. We will likely work through these given enough time.

In the future, we can therefore expect that individuals will have readily available access to their personal genetic risk predictions, either through a direct-to-customer service or through standard medical care. They will therefore have the potential to be much better informed about their disease risks, based both on genetics and also lifestyle factors.

3. Life insurance & the use of genetic information

3.1 Life insurance products & eligibility

There are three main distribution channels for obtaining life insurance:

1. Through superannuation plans, referred to as *group insurance*

   Group insurance is sold through superannuation and usually offered by an employer or a large-scale entity, such as an industry fund. There is usually a basic level of cover provided to members of that superannuation scheme before any underwriting is required, referred to as the 'automatic acceptance limit' (AAL). Additional cover on top of the AAL is available, but would require some form of underwriting and/or medical disclosure.

2. Through financial advisers, referred to as *retail insurance*

   Retail insurance is typically sold through financial advisers, who can be either independent or aligned to the company of the insurance product. There is usually some level of underwriting and/or medical disclosure required before a policy is accepted. The level of underwriting would depend on the sum insured.

3. Through direct-to-customer avenues, referred to as *direct insurance*

   Direct insurance is typically marketed and sold directly to customers, for example via the internet, television, telephone and mail. The underwriting

requirements for this channel are low. Correspondingly, the benefits provided are usually also fairly limited. In addition, the size of this channel is small in comparison to group and retail. Therefore we have not discussed it further in this paper.

The types of cover offered under the group and retail distribution channels are below:

Type of Cover	Group	Retail
Death	Y	Y
Total & permanent disability (TPD)	Y	Y
Trauma (also known as 'critical illness')	N	Y
Income protection (short term)*	Y	Y
Income protection (long term)**	N (in most cases)	Y

* Income protection (short term) refers to benefit periods 2 or 5 years
** Income protection (long term) refers to benefit periods to age 55, 65 or longer

Examples of automatic acceptance levels are below:

	AAL Examples	
	Australian Super	SunSuper
Death & TPD	$1.5m (cover above $0.6m is capped at $1.5m or 10 × salary, whichever is lower)	Up to $1.0m, depending on level of salary
Income protection (short term)	Up to $20,000 per month or 85% of your salary, whichever is lower (75% paid to member, 10% paid to superannuation account)	Up to $12,000 per month, depending on level of salary

This shows that Australians can generally obtain some amount of death, TPD, and short term income protection cover within working age, irrespective of current health conditions, through group insurance. The amount of basic cover usually depends on recent salary. However, to obtain cover above the group AAL levels, long term income protection, or trauma cover, customers would require some level of underwriting and disclosure of current health state. The requirements and issues in relation to genetic disclosures and its impact to both the individual and the insurer are discussed in the next section.

It should be noted that for the purposes of this paper, life insurance products offered to people post-retirement are not discussed. While the impacts of genetics research will have implications for this segment of the population, the need for life insurance is much more pronounced for people of working age, particularly while raising a family and/or repaying large debts such as mortgages.

3.2 Legislation & guidance on genetic disclosures

In Australia, the Financial Services Council, an industry body representing the vast majority of life insurance companies in Australia, has an industry standard (Standard No. 11) that applies to genetic testing, disability and trauma insurance. The standard requires that members of the FSC commit to the following provisions for their applicants:

- Will not require you to undergo a genetic test when you apply for insurance.

- Will require that you make available the results of any previously undertaken genetic tests upon request.
- Will not use your genetic test information to assess another family member's risk, for example genetic test information obtained from a parent will not be used to assess an insurance application made by the son or daughter.
- Will take account of the benefits of special medical monitoring, early medical treatment, compliance with treatment and the likelihood of successful medical treatment when assessing overall risk.
- Will ensure that genetic test results are only made available confidentially to the insurer's underwriters and reinsurance companies.
- Will provide, to you or your medical practitioner, reasons for any adjustment to premiums or policy conditions after assessing your application.

Australian life insurance products are guaranteed renewable, which means that once a policy is accepted, further changes in health are no longer required to be provided to the insurer. That said, any changes in cover and/or changes in insurer may require further disclosure. Due to the nature of guaranteed renewable products and the industry standard that requires the results of a genetic test to be made available when requested, it is typical for medical practitioners, such as doctors or genetic counsellors to advise individual to assess and apply for life insurance before obtaining any genetic tests.

Given the disclosure requirement for obtaining life insurance, would it deter people from obtaining a genetic test? Given that genetic testing for common disorders has only recently become readily available, there is limited data which can be used to definitively answer this question. However, the views of genetic counsellors currently are that this would be a small deterrent in stopping people from obtaining genetic tests. A larger consideration is whether an individual is mentally prepared and willing to know their genetic predisposition, particularly for disease with no family history, and what to do with that information once known.

For those who are prepared to understand one's health risks due to predictive genetic testing, a consideration they may have before undertaking a genetic test is how 'persistent' these results will be. In other words, once a person's genome is measured, would further advances in genetic research mean you have pre-emptively consented to all future tests? Under the extreme scenario where the entire genome is sequenced and provided to an insurer, in principle they could use it to conduct their own analyses of disease risks and use this to assess further applications for insurance. Currently, life insurers have not invested in the capability to retrospectively analyse genomic data that was previous provided. In the long term, investment in such capability may depend on the perceived threat of genetics and how the industry wishes to deal with it. The current views on the threat of genetics to life insurance are discussed in the next section.

3.3 Current views on the 'threat' of genetics

What are the current views, within the life insurance industry, of the threat of genetic testing to life insurance business models? We interviewed a number of industry practitioners, including underwriters and Chief Medical Officers, to assess this. While views differed from insurer to insurer, or person to person, the overarching view was

that currently genetics is not seen as a large threat (although it was deemed a potentially emerging risk). The basis for this view was:

1. Medically, there are a limited number of life-debilitating diseases which will definitely occur but that *only* manifest in later life. There are a small number of exceptions, such as Huntington's disease and Parkinson's disease, but the incidence of such disorders is rare enough that they do not materially impact the costs of claims.

2. For predictive genetic tests, rather than deterministic tests, there is a limited ability to use this information to reject applications or vary premium rates, as it may be legally challenged in court. This is because lifestyle factors are often a greater contributor to health outcomes and thus can counter-balance the predicted genetic risk.

3. Based on the current genetic advances, the predictability of a disease may not be substantially different to currently used methods of assessment, such as family history and blood tests.

4. There are very few cases of declined claims due to non-disclosure of genetic test results at the time of application.

Therefore, while the results of any genetic test are required to be disclosed if requested, life insurers currently do not regularly make genetic disclosure requests. In addition, they are rarely used to assess the outcome or change a person's premium. This also seems to be largely the case in the USA as well (Green et al. 2015).

4. How might genetics impact life insurance?

Given the current view that genetic advances are not seen as a major threat to the life insurance industry, this section tests this hypothesis analytically.

Based on life insurance claim experience, for trauma and income protection (IP) products, the top causes of claims, excluding accidents, are:

Cause of claim	Cover most impacted	Predominate gender
Breast cancer	Trauma	Female
Prostate cancer	Trauma	Male
Heart disease	Trauma & IP	Male
Stroke	Trauma & IP	Male
Depression	IP	Male & female

Studies of genetics are wide-ranging and cover a multitude of diseases. Our analysis focuses on these top diseases as they have the greatest impact on the cost of claims.

4.1 Summary of genetic research on the top diseases

The progress of genetic research, including the current capability of risk prediction, for each disease is described below.

Note that, for the most part, the findings below arise from studies of European individuals and the results may not always generalize to other populations.

4.1.1 Coronary artery disease

Coronary artery disease (CAD), also known as coronary heart disease, is a group of diseases that include myocardial infarction, sudden cardiac death, and a few others. It is one of the most common causes of death (Khera & Kathiresan 2017), leading globally to millions of deaths every year. It affects males more often than females, and becomes progressively more common in older age. Due to its high prevalence, CAD is one of the most common conditions that leads to a life insurance claim, especially for men.

CAD has been well-studied by researchers. We know a number of common risk factors (e.g. smoking, high blood pressure, obesity), useful lifestyle interventions (healthy diet, regular exercise, quitting smoking) and effective medical treatments (e.g. statins, which are drugs that reduce cholesterol). We know that CAD is heritable, an observation that dates back all the way to the 1950s. Recent studies estimate its heritability at about 40-50% (Khera & Kathiresan 2017). This means genetics and lifestyle play a roughly equal role in the incidence of CAD.

Since 2007, ever larger GWAS of CAD have successfully discovered ever more genetic variants that confer an increase in risk. At present, we know about 60 such variants with a high degree of certainty (Khera & Kathiresan 2017).

When using some or many of these variants to construct polygenic risk scores for predicting CAD, substantial differences between individual risks were observed. The individuals in the top 20% ('high risk') based on the score were about twice as likely to develop CAD than the bottom 20% ('low risk') (Khera & Kathiresan 2017; Tada et al. 2015). To put this into context, such differences in risk are on par with the effect of key lifestyle variables such as smoking (Khera et al. 2016). Moreover, in all of these studies, the effect of the genetic and lifestyle variables seemed to be largely independent of each other, and they also independent of known family history (Tada et al. 2015).

The roughly 60 variants discovered by GWAS are only the ones that we so far have very strong evidence for. Together, they explain only about half of the estimated heritability (Khera & Kathiresan 2017), meaning that there are other genetic variants still to be discovered. Recently, a polygenic risk score was developed that used about 50,000 SNPs, aiming to capture more of the causal genetic factors (Abraham et al. 2016).

These developments have a clear potential for clinical use. Currently, screening for CAD risk happens late in life, because only then does it become possible to start effectively identifying high risk individuals based on clinical variables. With a polygenic score, such screening can happen much earlier. These scores can now stratify individuals to a large enough extent that many have prompted calls that we should now be focusing on how to incorporate them into clinical practice (Assimes & Goldstein 2016; Khera & Kathiresan 2017).

CAD is in somewhat of a unique position in that safe and effective preventions exist that can demonstrably counteract the genetic risk. Moreover, interventions have been shown to have the most benefit for higher risk individuals. This, together with the fact that CAD is such a common disease, makes it an ideal candidate for a population screening strategy based on genetics (Assimes & Goldstein 2016), a fact that suggests that such testing is not too far into the future.

4.1.2 Breast cancer

Cancer that develops in breast tissue is the most common type of cancer in women, accounting for 25% of cases and affecting 12% of women worldwide. (It can also affect men, but is substantially rarer, so the rest of our discussion will primarily focus on breast cancer in women.) Like many diseases, the risk progressively increases with age, with increasing incidence observed even amongst women in their 30s. Correspondingly, breast cancer is the most common condition that leads to a life insurance claim in women.

Breast cancer has been extensively studied and a large number of risk factors have been identified. Beyond sex and age, some other known risk factors include genetics, lack of childbearing, lack of breastfeeding, smoking, diet (e.g. high alcohol intake), exposure to radiation and certain chemicals. A number of medical procedures are used to treat or manage breast cancer, with varying levels of success. Some prevention strategies also exist, such as maintaining a healthy lifestyle (balanced diet, low alcohol, physically active), which would be expected to prevent 20–40% of cases, and pre-emptive surgery (removal of breasts), which is only done for high-risk individuals.

Family history is an important risk factor, due to inherited genetic variation. Women who have a first-degree relative with a history of breast cancer have about two times greater risk of developing the cancer themselves. The actual genetics of breast cancer are somewhat complex, with a mixture of rare and common genetic variants known to contribute to risk (with many more likely to exist but are currently unknown).

Some of the rare variants lead to a particularly large increase in risk, most notably in the *BRCA1* and *BRCA2* genes. Women with these high-risk variants have about five times greater risk of breast cancer (and 10–30 times greater risk of ovarian cancer), with about 5–10% of overall cases attributable to these two genes. We have known about these two genes since 1994 (Miki et al. 1994); genetic tests for these have long been available and are routinely used to identify individuals of high risk that runs in families. The U.S. Supreme Court famously ruled, in 2013, that the patents on these genes were invalid, leading to a proliferation of much cheaper and readily available gene testing services for these and many other genes (Easton et al. 2015). Many of these rare gene variants also lead to high-risk for other cancers.

Despite their strong effect, the known rare genetic variants (such as *BRCA1* and *BRCA2*) account for less than 20% of the increase in risk due to family history, because they are so rare. More recent studies have identified many common variants which are also associated with risk. A polygenic risk score was developed using 77 such variants (Mavaddat et al. 2015), which was shown to be able to stratify individuals by risk, even beyond known family history. For example, for individuals without a family history and in the top 20% based on the score, the lifetime risk of breast cancer was 17%, compared to only 5% for those in the bottom 20%. For individuals *with* a family history, the corresponding risks were 24% and 9%.

The evidence so far indicates that the genetic factors act largely independently of the lifestyle factors (Mavaddat et al. 2015), meaning that the polygenic risk score remains useful even with knowledge about a person's lifestyle.

According to the UK NICE guidelines, enhanced surveillance is recommended for women with lifetime risk of over 17%. The polygenic risk score on its own can identify

a substantial number of these without considering any other risk factors, with about 17% of cases expected amongst this group (Mavaddat et al. 2015). The score can also be used to make mammographic screening programs more effective, by augmenting the simple age-based criterion currently used. These facts support the notion of population genetic screening, although this would first need to be supported by a health-economic evaluation to see whether the benefits it would bring outweigh the cost of such a program.

4.1.3 Prostate cancer

Cancer that develops in the prostate gland is the most common type of cancer in men. It primarily develops in older men (over the age of 50), with the risk increasing with age. About 1 in every 5 men will be diagnosed with it by the age of 85. Correspondingly, prostate cancer is a common condition that leads to a life insurance claim for men.

Prostate cancer has been extensively studied. Like breast cancer, the most important known risk factors are sex and age (in this case, only males are affected), and also genetics (with risk known to vary across different ethnicities due to differences in the frequency of specific genetic variants). Various potential lifestyle factors have been highlighted by some studies, such as being overweight or having an unbalanced diet, with various degrees of evidence. However, for the most part we still lack convincing evidence that identifies particular lifestyle factors as playing a role (Khankari et al. 2016). Therefore, we still do not have any effective prevention strategies.

The earlier that prostate cancer is treated, the better the health outcomes are for the patient. For this reason, much of the effort is devoted to detecting early signs of cancer. Unfortunately, the tests commonly used for this purpose, such as the Prostate Specific Antigen (PSA) blood test, also pick up a large number of unproblematic cancers (ones that are slow-growing and harmless, and thus will not require treatment because they will not ultimately lead to the patient's death), leading to *overdiagnosis*. When this occurs, a positive finding can lead to unnecessary medical procedures and possibly harmful consequences, which is an undesirable disruption to someone who would otherwise lead a healthy life. This fact has made PSA screening controversial (Burton et al. 2013) and there is a lack of evidence to support a net benefit for widespread screening. Therefore, screening is only recommended for people with higher risk.

Family history is a known risk factor, due to inherited genetic variation. Men who have a first-degree relative with a history of prostate cancer appear to have about two times greater risk of developing the cancer themselves. Our knowledge of the underlying genetics is similar to breast cancer (see above). Some of the rare high-risk variants are known, such as in the *BRCA1* and *BRCA2* genes (which also lead to higher risk for breast and ovarian cancer in women), while more recent studies have led to the discovery of many common variants which each confer smaller increases in risk.

A few studies have developed polygenic risk scores for prostate cancer. For example, a score based on 77 variants was published a few years ago (Eeles et al. 2013), and also a more recent one based on 25 variants (Olama et al. 2015). Both showed the ability to identify some individuals of high risk: those who have scores

amongst the top 1% were shown to have more than four times greater than average risk of developing prostate cancer.

Similar to breast cancer, it was found that the disease surveillance procedures could be improved by changing from an age-based criterion for screening to one that stratifies and targets patients using a polygenic risk score. It could reduce by about 19% the number of men who need to be screened, while only missing about 4% of cancer cases. It should also reduce the rate of overdiagnosis, leading to a better benefit-to-harm ratio (Pashayan et al. 2015). Therefore, population genetic screening is likely to benefit the detection and management of prostate cancer but, as for breast cancer, this would first need a health-economic evaluation to consider the wider benefits and costs of such a program.

4.1.4 Stroke

A stroke occurs when there is an interruption to the blood supply to the brain. This can lead to brain injury, disability and death. About 87% of strokes are caused by a blocked artery, for example by a blood clot; this is known as *ischaemic stroke*. The remaining 13% of cases are due to a blood vessel rupturing and bleeding in the brain; this is known as *haemorrhagic stroke*. (Benjamin et al. 2017)

Stroke is a leading cause of death and disability, second only to heart disease. It is also a leading cause of life insurance claims.

Many risk factors for stroke are known. It occurs more often in men, the risk increases with age, and there are a number of well-known and important lifestyle factors, including high blood pressure, high cholesterol, smoking, being overweight, and drinking too much alcohol. A family history of stroke is also a risk factor, although exactly quantifying this is a challenge because different studies have reported different results (Sacco et al. 1997).

Large genetic studies of stroke have largely focused on ischaemic stroke and have found relatively few genetic variants that modulate risk (when compared to many other diseases). For example, Carty et al. (2015) and Pulit et al. (2016) each report only a handful of variants, and the NHGRI GWAS Catalog (MacArthur et al. 2016)—which aims to tracks discoveries across all published genetic association studies—reports only 10 genetic variants associated with ischaemic stroke as of 24 Apr 2017.

Polygenic risk scores developed for stroke (Malik et al. 2014; Tada et al. 2014; Hachiya et al. 2016) have shown a relatively modest ability to discriminate individuals: the increase in risk when comparing individuals in the top 20% with the bottom 20% of each score range from about 1.2- to 2-fold.

Therefore, genetic risk prediction for stroke is still somewhat immature. To date, we are not aware of studies that have evaluated its potentially utility for clinical use or advocated for widespread genetic screening of stroke.

4.1.5 Depression

Major depressive disorder, also known as depression, is a mental disorder defined as a persistent feeling of sadness or a lack of interest in outside stimuli, lasting for at least two weeks. Several forms of depression are known, which vary in their symptoms.

Depression is fairly common, about 15% of individuals are expected to develop it at some point in their life, and it is about twice as common in women as compared to men (Otte et al. 2016). It accounts for a substantial number of income protection insurance claims.

The causes for depression are unknown but several risk factors have been observed. These include both genetic and environmental factors.

Family history can explain a substantial component of the risk: individuals with a first-degree relative with a history of depression have about a three times greater risk for developing it themselves (Otte et al. 2016). Nevertheless, elucidating the genetic variants that play a role in the disease has proven challenging. A study that looked at more than 9,000 patients, which would be considered quite large and would be expected to discover many variants for other diseases, did not manage to confidently implicate even a single variant for depression (Ripke et al. 2012). One potential explanation for this is that patients with different symptoms were given the same diagnosis, thus diluting any genetic signals that confer risk to only certain types of depression (Levinson et al. 2014).

However, progress has been made on two fronts. First, by focusing on a specific cohort of individuals with severe depression that was relatively similar across patients, two risk variants were discovered (Cai et al. 2015). Second, using a much larger dataset of more than 100,000 patients from 23andMe's database to supplement data from existing studies boosted the power sufficiently to discover another 17 risk variants (Hyde et al. 2016).

To date, we are not aware of any studies that have developed and assessed a polygenic risk score for the purpose of predicting the risk of depression for individuals. Where such scores are calculated, they have been used to study the disease further by relating it to other conditions such as schizophrenia (e.g. Whalley et al. 2016).

4.2 Illustrative analysis of impact of genetic testing

We now show an analysis of the potential impact of anti-selection to the profitability of trauma insurance if genetic testing for coronary artery disease, breast cancer and prostate cancer were to be widely adopted. We did not include stroke and depression in the analysis because current genetic research has not yet demonstrated sufficient ability to discriminate individuals' risk for these diseases. The analysis is focused on trauma insurance as these three diseases are most relevant for this product.

Note: the analysis is illustrative as it aims to demonstrate the magnitude of potential change to claim and lapse rates, and is not intended to be a rigorous model.

4.2.1 Model assumptions

The financial analysis we show below is deterministic and is focused on estimating the anti-selection impact of claim and lapse rates. The modelling has identified the following key anti-selection behaviours customers may exhibit from greater understanding of their health with genetic testing:

1. **Disease-specific genetic risk factors & proportion of trauma claims.**
The table below shows risk parameters for each disease based on the current genetic research as summarised in Section 4.1. The proportion of current trauma claims due to each disease is also included.

For simplicity, we treat the genetic test results for each disease as dichotomous, returning either a 'high risk' or a 'not high risk' result. For convenience, we label the latter as 'low risk' for the purpose of this analysis. (The summaries of research findings in the previous section only reported a comparison of, for example, the top 20% against the bottom 20%. To convert this to a comparison of the top 20% against the remaining 80%, we assumed an intermediate level of risk for the middle band of individuals.)

As a further simplification, we grouped all individuals into either an overall 'high risk' group, which includes those with a 'high risk' result for **any** of the three diseases we consider, or otherwise into an overall 'low risk' group, which includes only those who obtain a 'low risk' result for **all** of the three diseases. We assumed that the risk of CAD operated independently to that for the cancers, that the split of males and females was 50% each, that breast cancer only occurred for females and prostate cancer only occurred for males. We also assumed that the relative risk implied by a genetic test operates uniformly across a person's lifetime (e.g. it will impact risk during the younger years, when an insurance policy is in force, not just in much later life, after which most policies would typically expire).

Using these parameters and assumptions, we calculated the overall relative disease risk for the high risk group compared to the low risk group. This can also be interpreted as the relative increase in claims or claim costs relating to these three diseases.

Top 3 diseases	Prop. high risk	Increase in risk relative to the 'low risk' group*	Prop. trauma claims due to condition (ages 35 to 65)**
CAD	20%	45%	12%
Breast cancer	20%	71%	12%
Prostate cancer	1%	61%	10%
Total	**28%**	**31%**	**34%**

*For this analysis, 'low risk' means 'not high risk'.
**Based on the survey by FSC & KPMG (2017). However, the survey provided cancer claims as a whole; the further breakdown into breast and prostate cancer provided here was obtained from AIHW (2017).

2. **The proportion of the population that obtain PRS-based genetic tests.**
Under the base case, this assumption is set to be 0.5%. This is likely to be greater than the current state (note: we are referring to the use of predictive genetic tests based on a polygenic risk score or similar, rather than tests that measure single genes as is more routine for monogenic diseases). As of 7 Apr 2017, 23andMe has over 2 million customers (Hyde 2017). These include both customers which test for medical genetics and for genealogy, and the majority of them are likely to be from the USA given the origin of the company. However, given the recent FDA approval, the assumption of 0.5% of the population is a plausible future state.

We expect this proportion to increase as the cost decreases and the predictive power of the genetic tests increase. Therefore, for claim costs we also provide a sensitivity analysis to changes in this parameter.

We have assumed in our model that no exclusions or premium loadings are applied to applicants who have obtained genetic tests, particularly in relation to the three modelled diseases. This is plausible as the current underwriting practices have tended to not use predictive genetic information, particularly for polygenic diseases that are non-deterministic in illness, when assessing applications.

3. **The proportion of in-force policies that lapse if known to have low genetic risk.**
The model assumes that 20% of in-force policies would lapse if the insured individuals obtained a genetic test and their results show that they are not at a high genetic risk.

   Note that, as a reference, we compare the change in lapse rates to an assumed best-estimate current lapse rate of 15%.

4. **The proportion of non-insured / under-insured population who apply for insurance before obtaining genetic tests.**
The model assumes that 100% of insurable population that are non-insured or under-insured will apply for insurance before obtaining genetic tests. Whilst this is an extreme assumption, as it would depend on the age and financial circumstance of the individual, it is expected that the actual proportion would nevertheless be quite high. This is because medical practitioners, such as doctors or genetic counsellors, are trained to advise individuals to consider their desire for life insurance and inform them that genetic test results could impact their ability to obtain such insurance because the results may be requested during the application process.

5. **Of the new applicants that apply for insurance before obtaining genetic test, the proportion that keep their insurance after receiving test results.**
The previous assumption referred to people who would apply for insurance before obtaining genetic test. Only a proportion of them will retain the policy depending on the result of their test. We assume that this will be only those who find out they have high genetic risk, and we will assume that the others (low risk) will let their policy lapse. Note that we assume no financial impact for those that lapse in this manner because they will have held their policy for a very short time, effectively the same as never having applied in the first place.

   An alternative way of presenting assumptions 4 and 5 is that all eligible non-insured high risk people apply for insurance and get accepted on standard rates. This may be plausible as the tests are direct-to-customer, which would make it difficult for insurers to prove non-disclosure, and therefore easy for customers to anti-select against insurers.

6. **The proportion of the insurable population that is currently insured for trauma.**
The model assumes 8% of the population is currently has trauma insurance, based on Rice Warner (2011).

A summary of these assumptions is below:

Assumptions		
For CAD, breast and prostate cancer		
Prop. people with high genetic risk	[a]	28%
Increase in risk (high risk vs low risk)	[b]	31%
Prop. claims (ages 35 to 65)	[c]	34%
% of population who obtain PRS-based genetic tests	[d]	0.5%
% insured lapse if known to have low genetic risk	[e]	20%
% non-insured who obtain insurance before genetic test	[f]	100%
% of insurable population that is insured for trauma	[g]	8%

4.2.2 Methodology & base results

Based on the assumptions above, we show below an estimate of: (i) the increase in claim costs from new business (NB) anti-selection; and (ii) the increase in lapse rates from in-force selection lapsation.

		Total	High risk	Low risk
Total population	h	100.00		
No. insured	i = h × g	8.00	2.27	5.73
No. non-insured	j = h × (1 - g)	92.00	26.13	65.87
No. non-insured that get tested	k = j × d	0.46	0.13	0.33
No. non-insured with high risk result that obtain policy (leading to NB)	l = k × a × f	0.13	0.13	—
No. insured after test	m = i + l	8.13	2.40	5.73
% increase due to NB	n = (m – i) / i	1.6%	5.8%	0.0%
Relative claim cost per person (low risk = 1)	o, where high risk = low risk × (1 + b × c)		1.11	1
Claims cost before anti-selection	p = i × o	8.24	2.51	5.73
Claims cost after anti-selection	q = m × o	8.34	2.66	5.73
% increase in claim cost	(q – p) / p	**1.8%**	**5.8%**	**0.0%**
No. in-force lapsed due to low risk result	r = i × d × e	0.01		0.01
% of in-force lapsed due to low risk result	t = r / i	0.1%	0.0%	0.1%
% increase in lapse rate (c.f. 15%, assumed current lapse rate)	t / 15%	**0.5%**	**0.0%**	**0.7%**

On the base assumptions the impact on claim lapse rates are small, but no insignificant. The reason for the results is predominately due to the assumption that only 0.5% of the population is seeking this information, a very small proportion. Irrespective of other assumptions, such as the predictive power of genetics, if only a small proportion of the population is using the technology, only those that get tested can chose to anti-select against the company.

4.2.3 Sensitivity analyses

Proportion of the population who obtain genetic tests

The most significant factor that would impact the financial results of insurers is the proportion of the population who obtain genetic tests. See below the sensitivity for variations in this assumption:

	Base	Variation		
% of population who obtain PRS-based genetic tests	**0.5%**	**1%**	**2%**	**5%**
Increase in claims from NB anti-selection	**1.8%**	**3.5%**	**7.0%**	**17.5%**
% of in-force lapsed due to low risk result	0.1%	0.2%	0.3%	0.8%
% increase in lapse rate	**0.5%**	**1.0%**	**1.9%**	**4.8%**

This analysis shows that if genetic tests were to become more widespread, the potential impact on claim costs and write-off of acquisition costs due to lapse could be material. The proportion of people who get tested only needs to rise to 2% to have claims costs increase by as much as 7%.

Predictive power of genetic tests

The estimated impact of changes in the predictive power of genetic tests is show below:

	Base	Variation		
Increase in disease risk (high vs low risk)	31%	131%	231%	331%
Increase in claims from NB anti-selection	**1.8%**	**2.1%**	**2.4%**	**2.6%**
% of in-force lapsed due to low risk result	0.1%	0.1%	0.1%	0.1%
% increase in lapse rate	**0.5%**	**0.5%**	**0.5%**	**0.5%**

This analysis demonstrates that even if the predictive power of genetic tests were to significantly increase, the financial impact remains small if the prevalence of genetic test is at 0.5%. As only 0.5% of the population is assumed to obtain genetic tests, only a small number of policyholders could anti-select against the insurers. In practice, if predictive power were to significantly increase, we would also expect the tests to be more widely adopted, and thus the two variables would be correlated to some extent.

Combination proportion of the population who obtain genetic tests and predictive power of genetics

	Base	Variation		
% of population who obtain PRS-based genetic tests	0.5%	1%	2%	5%
Increase in disease risk (high vs low risk)	31%	131%	231%	331%
Increase in claims from NB anti-selection	**1.8%**	**4.2%**	**9.5%**	**26.3%**
% of in-force lapsed due to low risk result	0.1%	0.2%	0.3%	0.8%
% increase in lapse rate	**0.5%**	**1.0%**	**1.9%**	**4.8%**

This analysis shows that combined together, the amount of genetic tests undertaken and the predictive power anti-selection risk would significantly increase anti-selection risk.

4.3 Results in relation to current views

The current view is that genetics is an emerging risk but is not seen as a large threat at the moment. Our analysis shows that under the base assumptions, particularly with a very small proportion of the population undertaking genetic tests, this view is valid. However, if the use of such tests grows, it would become a threat.

One characteristic that differentiates genetic tests from other screening tests is that the actual measurements won't change over time, although their interpretation might, in light of any new advances in research. For this reason, some testing services provide updated risk reports to customers when they update their predictive models. This means that customers may only need to take one genetic test in their lifetime for their results to remain valid and updated over time. Therefore, while the impact of genetic tests over a single-year period is small, cumulatively over time, the impact would be larger as the results remain persistent.

Currently, since only a very small proportion of the population have obtained genetic tests, the impact of persistent results is negligible. However, if 2% to 5% of the population were to have undertaken genetic tests, this may be the critical point at which companies should re-consider their pricing, product and underwriting practices.

A further risk consideration for insurers is that increase in claim cost would be predominately due to anti-selection of new applicants. This anti-selection risk can be somewhat managed by an insurer through their underwriting process. However, the risk of selective lapsation on the in-force book may be more difficult to manage as there are fewer actions available. In particular, it would be challenging to retain customers if their perceived need for insurance changes if they have greater awareness of potential future health state.

5. Further discussion and evolving life insurance to cope with greater personal knowledge

In this paper, we have focused on summarising the developments in genetic research and its impacts on life insurance. A key feature of this is that individuals are likely to become better informed of their own health prospects. This raises broader questions about the nature and role of life insurance. We touch on some of these here, with the aim of highlighting ideas for further discussion and consideration.

5.1.1 Opt-in versus opt-out group insurance

Under the current superannuation regulation, insurance cover is based on a mandatory opt-in arrangement. An area of recent discussion has been whether this arrangement should continue, or if it should change to a voluntary opt-in model. Industry submissions from insurers to the Joint Parliamentary Committee have been divided.

The current mandatory opt-in model reduces the risk of anti-selection, as policyholder health risks would be a better representation of the general population

health risk. However, if group insurance were to move to a voluntary opt-in model, policyholders are more likely to anti-select. Insurers could respond by reducing AALs to protect against this risk. A consequence of this response for society at large may be reduced eligibility and/or affordability of insurance to high-health-risk lives.

5.1.2 Ethical considerations

Fundamentally, the broad issues facing the life insurance industry are:

1. Meeting the underlying social need for insurance, particularly from a customer perspective

2. Ensuring the sustainability of the industry

As individuals become better informed, for example via genetic tests, over time this would impact the concept of large pooling of risk, which currently underpins the tools used to design and price insurance products. However, life insurance supports the social need for financial security. Therefore, there is a fundamental ethical tension between the desire to be inclusive and not discriminate insurance applicants based on genetic information, particularly when one's genetics are determined at birth, and the desire to protect the integrity of insurance companies' business models in the presence of information asymmetry and potential anti-selection.

5.1.3 Long term state for insurance

Over time, as individual health outcomes become more tailored, for example through better predictive genetic tests or other forms of early disease diagnosis prior to onset of symptoms, the concept of risk pooling, fundamentally used insurance pricing would need to adapt.

In the longer term there may be many possible states to which the Australian life insurance industry may veer towards. Some of these might be:

1. More-tailored premiums, based on smaller pools of individuals, whereby individuals may differentiate themselves by providing updated lifestyle data

2. Much larger pools and restrictions on 'tailiorability', possibly mandated by legislation (Green et al. 2015)

3. Some restrictions on factors that are allowed to be used for pricing premiums and setting exclusions, especially with respect to genetic tests.

As an industry, it is important to consider which long-term state we want to transition to, and how best to assess and create a path to it. Invariably the path would involve consultations with many facets of society such as insurance companies, regulators, medical professions and customers, over a period of time. However, in steering towards a desired long term state, the industry should set a clear set of principles and a framework to meeting these, which underpins the life insurance industry. In starting the discussion for desired future states that copes with greater personal knowledge, a potential overarching principle might go along the lines of: "Accepting that everyone has some level of health risks and ensuring that the Australian population has access to an affordable basic level of cover, should there be an insurance need".

6. Conclusions

Genetic research is advancing rapidly; useful predictive genetic tests are available for a number of diseases.

Genetic variants associated with disease risk for many common diseases are being discovered in unprecedented numbers. These are being used to build predictive models, which can form the basis for genetic tests for individual disease risk. For a number of diseases, including important ones for life insurance such as coronary artery disease, breast and prostate cancer, these have reached the point that they are as useful for prediction as the key known lifestyle factors. Moreover, the information they provide seems to be largely independent of such factors, and also informative beyond any knowledge of family history. As these tests develop further, we expect to see wider adoption, whether by incorporation into standard clinical practice or through interested individuals seeking them out of their own accord.

While genetic test results are required to be provided when requested by an insurer, life insurers currently do not regularly make genetic disclosure requests.

Australians are able to obtain a certain amount of insurance cover without disclosing any medical information, typically through Group insurance. For additional cover where medical disclosure is required, the FSC has prescribed guidelines on disclosure requirements. The guidelines state that genetic test results are required to be provided when required by an insurer. However, life insurers currently do not regularly make genetic disclosure requests.

Genetic risk is not viewed as a current threat to life insurance. However, it may become a threat if genetic testing becomes more prevalent.

Our illustrative analysis confirmed that genetic testing should not currently have a substantial impact on the life insurance industry. This main reason is because a very small minority of the population has so far undertaken genetic tests, and therefore only very few people exist who have the information to anti-select against insurers.

However, a characteristic that differentiates genetic tests from other screening tests is that they only need to be done once, since a persons' genome does not change over time. Therefore, even if only a small proportion of the population get tested each year, the cumulative impact over time might become significant. We suggest that once about 2% to 5% of the population has undertaken genetic tests, we may see a material impact on the life insurance industry. A particular concern is the direct-to-customer nature of many of the new genetic testing companies, which would make it difficult for insurers to prove non-disclosure, and therefore easy for customers to anti-select against insurers.

A further risk consideration for insurers is that increase in claim cost would be predominately due to anti-selection of new applicants. This anti-selection risk can be somewhat managed by an insurer through their underwriting process. However, the risk of selective lapsation on the in-force book may be more difficult to manage as there are less directly management actions available. In particular, it would be challenging to retain customers if their perceived need for insurance changes if they have greater awareness of their potential future health state.

We may need to think differently about how we structure life insurance in the future

If individuals were to become much better informed about their health prospects, for example via genetic tests, our basis for designing life insurance products might need to adapt to suit. In particular, finding an appropriate way to pool risk, that is both socially acceptable and financially sustainable, will be a challenge.

Appendix A. Genetics & epidemiology

A.1 Fundamentals of genetics

A.1.1 DNA

Deoxyribonucleic acid (DNA) is the molecule that carries our genetic information. Each of our cells has a number of such molecules, which are called *chromosomes*. The set of all chromosomes in a living cell is called its *genome*. Each chromosome is a very long molecule, a chain consisting of units called *nucleotides*. There are four different DNA nucleotides, usually referred to by the letters A, C, G, and T. It is the sequence of these nucleotides that forms the genetic information in the genome.

Different parts of the genome have different roles. An important one are the parts called *genes*, whose function is to store the instructions for how to make proteins, which are fundamental building blocks of living organisms. The sequence of nucleotides in a gene gets interpreted in a complex but consistent way to form proteins, somewhat analogous to how the instructions written by a computer programmer get converted into computer software.

Only a minority of the genome is actually comprised of genes. Other parts serve different roles, such as helping to regulate which genes should be turned 'on' or 'off', or how much protein should be made from each one. Much of the genome seems to have no function, or at least no role we have yet to elucidate. This led to the term 'junk DNA' being applied to such regions; this terminology is now discouraged as we slowly learn the function of some of these regions, although it is still popular in colloquial usage.

A.1.2 Inheritance

DNA plays a key role in *heredity*, the passing of genetic information from parents to children. When a human egg is fertilised by a sperm, each contributes half the number of chromosomes that will form the resulting embryo. These two halves are created in a process called *meiosis*, during the formation of the egg and sperm. This involves a number of cell divisions that end up halving the usual number of chromosomes in a cell. *Genetic recombination* also occurs during meiosis, a process of somewhat random exchange of genetic material between chromosomes, resulting in chromosomes that contain portions from both the mother and father of that individual. The net result is that, while each of us inherits half of our genome from each of our mother and father, the contribution from each grandparent is random (but, on average, is one quarter).

A.1.3 Genetic variation

A human genome is about 3 billion nucleotides long (6 billion when counting the contribution from each parent). Any two individuals would have about 99.9% of this

in common. This is unsurprising since it is this commonality that defines us as being human, as compared to the genomes of, for example, a kangaroo or an onion. Where differences exist between individuals, they take a number of forms.

The simplest type of genetic variants, and the most studied, are SNPs. These are places in the genome where one nucleotide differs between two individuals, but the surrounding genetic sequence is identical. For example, I might have a T and you might have a G at the same genetic location. More complex genetic variants include insertions and deletions (sequences of nucleotides that are present in one individual but not another), copy number variants (sequences that occur multiple times, with differing numbers of repeats in different individuals) and inversions (sequences that occur in the reverse order in some individuals).

SNPs are relatively common throughout the human genome. The current best estimate is that there are about 10 million, which is on average 1 in every 300 nucleotides. Note that this refers to all variants that might exist between some pair of individuals. If we actually take any specific pair of individuals and compare their whole genomes, the number of variants between them would be much smaller, closer to 1 in 1,000.

During recombination, portions of DNA from parental chromosomes are incorporated together as large blocks. This means that nearby genetic variants tend to be inherited together, with close relatives sharing large tracts of common sequence (thus, we see *long-range correlation* across their genomes). As we compare more distant relatives, being 'separated' by many more meioses, the sizes of these shared blocks decreases. In the extreme, if we took two random people from the population, we would not see such large blocks. However, we would nevertheless still observe short-range correlation across the genome, the remnant signature of inheritance after a large number of meioses. This correlation is referred to as *linkage disequilibrium* (LD).

The existence of LD and the abundance of SNPs has been exploited to allow study of the genome without having to actually determine whole genetic sequences (see Section A.2.2). By measuring a large number of SNPs across the genome, we get indirect measurements of nearly the whole genome. The reasoning is that, if another genetic variant is a causal factor for disease predisposition, due to local LD it will be correlated with a nearby SNP. As long as we measure some of these local SNPs, we will be able to detect an association between them and the occurrence of the disease. Such SNPs are referred to as *proxy* or *tag* SNPs, and are said to be *tagging* the underlying casual variants.

A.1.4 The relationship between genetics and traits

The genome instructs our bodies how to grow and operate, which means it plays a key role in determining our physical traits. Our surrounding environment also plays a key role, for example through our diet and social interactions. The extent to which one or the other is more important varies by trait. For example, eye colour is completely determined genetically, while the ability to speak Spanish will depend mainly on how much you are exposed to and encouraged to learn it.

Many traits will depend on both genetic and environmental factors. A good example of this is height. We all know that taller parents tend to have taller children, and this is known to have a genetic basis. However, a malnourished child is unlikely

to realise their potential height even if they have a genetic propensity to be tall. The extent to which genetics determine the trait is known as the trait's *heritability*. This is formally defined as the proportion of variance of the trait that is attributable to genetic factors. For example, human height is 60–80% heritable; i.e. mainly genetic, but with a substantial portion depending on the environment.

Traits also vary in how many genes are involved in determining, or influencing, them. For example, eye colour is controlled by only a handful of genes, whereas at least many hundreds of genetic variants influence height (this is based on current discoveries only, there are likely to be yet more still).

Traits controlled by only a single gene are called *monogenic* or *Mendelian*. Most examples of such traits are diseases, for example Huntington's disease and sickle-cell disease; such diseases can be thought of as essentially being caused by a defective gene. Eye colour and hair colour were previously thought to be monogenic, but are now known to be the result of multiple genes.

Traits that are influenced by a large number of genes are called *polygenic* or *complex*. Such traits will generally also depend on environmental factors. Many disease are complex in this regard, with predisposition running in families but also with known lifestyle risk factors. Some examples include cancer, diabetes and heart disease.

A.1.5 Beyond DNA

While the genome encodes the instructions for our bodies, the basic DNA sequence is not the only unit of biological inheritence. Chromosomes can be chemically and physically modified in ways that are stable and which can be passed on to children. Such modifications are called *epigenetics*. Unlike the DNA sequence, epigenetics can vary between cells in the body and change over the lifespan of an individual. They interact with the DNA itself to control, for example, how and when each gene is used to make a protein. Epigenetics are one clear mechanism for how environmental factors can affect and interact with genetic factors in determining a trait. The study of epigenetics, especially their effect on diseases, is a relatively new and still developing area.

A.2 Brief history of genetic epidemiology

Epidemiology is the study of determinants of health and the incidence of diseases, an endeavour that goes back many centuries. *Genetic epidemiology* is a relatively recent branch of this field, focussing on how inherited factors, such as genes, relate to health and disease.

Before the discovery of DNA, researchers studied the occurrence of traits within families. Using mathematical models of genetic inheritance, they were able to estimate the heritability of these traits. As we developed methods to measure DNA, we were able to start studying the genetic variants themselves. The large scale of the genome was an ever present challenge and for a long time hampered efforts to locate the variants relevant for any given trait. Nevertheless, through continued innovation in both technology and study design, we have been able to uncover thousands of genetic variants that either cause or predispose us to various diseases. What follows is a brief historical account of the different phases of discovery.

A.2.1 The early years: family-based studies

Genetic epidemiology was first described in 1954, a time when directly measuring genetic variants was not yet possible (Seyerle and Avery 2013). The state-of-the-art at the time was studying the patterns of inheritance in families to try to determine the extent and nature of their underlying genetic variants, through what were known as *familial aggregation* and *segregation* studies. For example, early analyses indicated that breast and ovarian cancer had a strong genetic component (Go et al. 1983).

The 1980s saw the development of *linkage analysis* (Botstein et al. 1980), which combined the idea of looking at patterns of inheritance in families together with advances in technology that allowed the measurement of genetic variants. This enabled researchers to localise the causal genetic variant(s) to specific, but broad, areas in the genome. Such studies led to successful identification of the genes responsible for many monogenic diseases. Notable examples include Huntington's disease and cystic fibrosis.

A.2.2 Population-based studies and the genome-wide era

Despite their successes, linkage studies were not able to shed much light on the genetic variants that underlie complex diseases. This led researchers to shift to a different approach called an *association study*, which involves directly comparing genetic variants and the outcomes of interest. This approach was shown to be more powerful than linkage analysis for complex traits (Risch & Merikangas 1996). Moreover, it could be done with population-based samples (rather than family-based), which is cheaper and more feasible to do on a large scale.

Initially, studies of this type were based on candidate genes, meaning they only measured genetic variants in genes that were thought to be responsible for the disease of interest. This was the only practical approach at the time. To a large extent, this method was also not particularly successful.

The publication of the Human Genome Project (Lander et al. 2001; Venter et al. 2001) and subsequent projects that catalogued human genetic variation (HapMap 2003), raised the prospect of conducting association studies that used genetic variants across the whole genome. A key enabler of this was the development of *SNP genotyping arrays*, a technology that allowed for cheap and timely measurement of hundreds of thousands of SNPs. By designing these to have SNPs that cover the genome (by exploiting LD to tag other variants, see Section A.1.3), and deploying them across large samples of individuals, we reached the era of *genome-wide association studies* (GWAS).

The early GWAS caused great excitement in the field, finally producing discoveries of specific genetic variants that are associated with complex diseases. This continued apace; over the space of a few years, ever larger GWAS were conducted, leading to many thousands of discovered genetic risk variants for hundreds of different diseases and traits. Many of the discoveries related to biological mechanisms that were unexpected or surprising for each disease, explaining why the candidate gene approach often did not lead to success.

A notable feature of the vast majority of the discovered variants is that their impact on disease risk is, by and large, very small. This meant that they typically had very little predictive power and could explain only a small fraction of the overall

incidence of the disease of interest. In fact, they also only explained a small fraction of the heritability of each disease, as previously estimated from family studies. This prompted the question of where the 'missing heritability' lies (Manolio et al. 2009). Many theories have been put forward and the exploration of this question continues to drive research to this day.

Another feature of the discoveries is that they are mostly genetic variants that are relatively common in the population (e.g. carried by more than 1% of people). This is not surprising because the nature of the GWAS design entails greater statistical power for more common variants. As a result, one theory is that much of the missing heritability lies within rare variants.

The fact that many large-scale GWAS were funded and were very expensive, and that only a small proportion of heritability was explained by their results, led to criticisms that they were a waste of money. A key strand of this criticism relates to the lack of biological insight from many of the discoveries and the inability to immediately translate them to clinical use. In response, it is usually pointed out that these studies are genuine breakthroughs in our knowledge about complex disorders, having finally given us vital pointers into which parts of the (extremely long) genome to focus on in follow-up research. Furthermore, the discoveries so far are just the 'tip of the iceberg', with more expected in future, including further biological insights, leading hopefully to better methods for early diagnosis and treatment.

Time will tell whether this long-term vision is fulfilled. However, some of the consequences are already clear. In particular, as research continues, we are likely to uncover ever more of the 'genetic architecture' of diseases and thus likely also improve our ability to predict disease risk. Moreover, we will also improve our understanding of how genetic and environmental factors interact with each other. For example, genetic variants whose presence leads to an individual being highly responsive, or not responsive, to certain lifestyle interventions or medical treatments (for example, the anti-HIV/AIDS drug abacavir causes a severe skin rash for a minority of patients, but most of them can be identified by a genetic test, so genetic pre-screening is now recommended before prescription of this drug). This is exciting from the point of view of public health. It also has clear implications for life insurance.

A.2.3 The sequencing revolution and hopes for personalised medicine

Over the last decade, genetic sequencing technology has been substantially revolutionised. It is now possible to read the sequence of a complete human genome in the space of a few days for less than $1,000. This stands in stark contrast to the assembly of the first human genome reference sequence, which took 13 years and cost $3 billion (USD).

In contrast to SNP arrays, sequencing allows for the measurement of all genetic variants in an individual, without relying on LD to tag unmeasured variation. This is particularly helpful for studying rare genetic variants, one of the 'gaps' in the current GWAS design.

Sequencing is still too expensive to replace SNP arrays for very large studies, but presumably this will change in the future given the current pace of technological progress. Meanwhile, it is being used in various ways that are currently feasible, such as for developing very detailed catalogues of genetic variation, for studies that only

look at genes (rather than the whole genome, which is much longer and therefore more expensive to sequence), and for smaller studies of rarer diseases.

Much of the current and future research using sequencing will focus on trying to find rarer genetic variants with a greater effect on disease risk. If many of these are discovered, it will go some way to solving the missing heritability problem. It would also increase the predictive power of genetic tests based on discovered genetic variants.

The success of GWAS and the rapid progress of genetic technology have led to the idea that in the future we would routinely measure someone's genome and on that basis prescribe very specific treatments and interventions that would be suitable and effective for that individual. This idea is known as *personalised medicine*.

The basic notion is actually not revolutionary: we already use personal characteristics such as sex, age, height and weight, and other measurements, such as blood tests, to help recommend the best course of treatment. We are simply adding a new measurement tool, albeit a very powerful one, to our toolbox. The extent to which this would help will vary depending on the medical scenario, but it will no doubt be very useful for many scenarios. It has been noted that there is a tension between trying to tailor information to the characteristics of each patient, but simultaneously combine information together across patients to demonstrate reproducible conclusions (Hunter 2016).

The more radical aspect of personalised medicine is that it will require substantial changes to our medical systems, including investment in new infrastructure, education of the medical workforce, an overhaul of regulations, and consideration of ethical and privacy issues. Progress on these matters will likely be slower than progress in genetic research itself.

References

Abraham, G., Havulinna, A. S., Bhalala, O. G., Byars, S. G., Livera, A. M. D., Yetukuri, L., et al. (2016, Sep). Genomic prediction of coronary heart disease. *European Heart Journal*, 37 (43), 3267–3278.

Abraham, G., & Inouye, M. (2015, Aug). Genomic risk prediction of complex human disease and its clinical application. *Current Opinion in Genetics & Development*, 33, 10–16.

Australian Institute of Health and Welfare (AIHW) (2017, Feb). Australian Cancer Incidence and Mortality (ACIM) books. http://www.aihw.gov.au/acim-books/

Assimes, T. L., & Goldstein, B. A. (2016, Oct). Genetic cardiovascular risk prediction: are we already there? *European Heart Journal*, 37 (43), 3279–3281.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017, Jan). Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*, 135 (10), e146–e603.

Botstein, D., White, R., Skolnick, M., & Davis, R. (1980, May). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics*, 32, 314–331.

Burton, H., Chowdhury, S., Dent, T., Hall, A., Pashayan, N., & Pharoah, P. (2013, Mar). Public health implications from COGS and potential for risk stratification and screening. *Nature Genetics*, 45 (4), 349–351.

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., et al. (2015, Jul). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, 523 (7562), 588–591.

Carty, C. L., Keene, K. L., Cheng, Y.-C., Meschia, J. F., Chen, W.-M., Nalls, M., et al. (2015, Jun). Meta-Analysis of Genome-Wide Association Studies Identifies Genetic Risk Factors for Stroke in African Americans. *Stroke*, 46 (8), 2063–2068.

Chatterjee, N., Shi, J., & García-Closas, M. (2016, May). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, 17 (7), 392–406.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J.-H. (2013, Mar). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, 45 (4), 400–405.

Easton, D. F., Pharoah, P. D., Antoniou, A. C., Tischkowitz, M., Tavtigian, S. V., Nathanson, K. L., et al. (2015, Jun). Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *New England Journal of Medicine*, 372 (23), 2243–2257.

Eeles, R. A., Olama, A. A. A., Benlloch, S., Saunders, E. J., Leongamornlert, D. A., Tymrakiewicz, M., et al. (2013, Mar). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature Genetics*, 45 (4), 385–391.

The Financial Services Council (FSC) & KPMG (2017). Retail Lump Sum Experience Investigation 2010-2012.

Go, R., King, M., Bailey-Wilson, J., Elston, R., & Lynch, H. (1983, Sep). Genetic epidemiology of breast cancer and associated cancers in high-risk families. I. Segregation analysis. *Journal of the National Cancer Institute*, 71, 455–461.

Goldstein, B. A., Yang, L., Salfati, E., & Assimes, T. L. (2015, Jul). Contemporary Considerations for Constructing a Genetic Risk Score: An Empirical Approach. *Genetic Epidemiology*, 39 (6), 439–445.

Green, R., Lautenbach, D., & McGuire, A. (2015, Jan). GINA, genetic discrimination, and genomic medicine. *New England Journal of Medicine*, 372, 397–399.

Hachiya, T., Kamatani, Y., Takahashi, A., Hata, J., Furukawa, R., Shiwa, Y., et al. (2016, Dec). Genetic Predisposition to Ischemic Stroke. *Stroke*, 48 (2), 253–258.

Hunter, D. J. (2016, Aug). Uncertainty in the Era of Precision Medicine. *New England Journal of Medicine*, 375 (8), 711–713.

Hyde, C. (2017, Apr). 23andMe Wins FDA Approval to Sell Genetic Tests. Bloomberg. https://www.bloomberg.com/news/videos/2017-04-07/23andme-wins-fda-approval-to-sell-genetic-tests-video

Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., et al. (2016, Aug). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, 48 (9), 1031–1036.

The International HapMap Project. (2003, Dec). *Nature*, 426, 789–796.

Kalf, R. R., Mihaescu, R., Kundu, S., Knijff, P. de, Green, R. C., & Janssens, A. C. J. (2013, Jun). Variations in predicted risks in personal genome testing for common complex diseases. *Genetics in Medicine*, 16 (1), 85–91.

Khankari, N. K., Murff, H. J., Zeng, C., Wen, W., Eeles, R. A., Easton, D. F., et al. (2016, Aug). Polyunsaturated fatty acids and prostate cancer risk: a Mendelian randomisation analysis from the PRACTICAL consortium. *British Journal of Cancer*, 115 (5), 624–631.

Khera, A. V., Emdin, C. A., Drake, I., Natarajan, P., Bick, A. G., Cook, N. R., et al. (2016, Dec). Genetic Risk Adherence to a Healthy Lifestyle, and Coronary Disease. *New England Journal of Medicine*, 375 (24), 2349–2358.

Khera, A. V., & Kathiresan, S. (2017, Mar). Genetics of coronary artery disease: discovery biology and clinical translation. *Nature Reviews Genetics*, advance online publication.

Krier, J., Barfield, R., Green, R. C., & Kraft, P. (2016, Feb). Reclassification of genetic-based risk predictions as GWAS data accumulate. *Genome Medicine*, 8 (1).

Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., et al. (2001, Feb). Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921.

Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R., et al. (2014, Oct). Genetic Studies of Major Depressive Disorder: Why Are There No Genome-wide Association Study Findings and What Can We Do About It? *Biological Psychiatry*, 76 (7), 510–512.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., et al. (2016, Nov). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45 (D1), D896–D901.

Malik, R., Bevan, S., Nalls, M. A., Holliday, E. G., Devan, W. J., Cheng, Y.-C., et al. (2014, Jan). Multilocus Genetic Risk Score Associates With Ischemic Stroke in Case-Control and Prospective Cohort Studies. *Stroke*, 45 (2), 394–402.

Manolio, T., Collins, F., Cox, N., Goldstein, D., Hindorff, L., Hunter, D., et al. (2009, Oct). Finding the missing heritability of complex diseases. *Nature*, 461, 747–753.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017, Apr). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics*, 100 (4), 635–649.

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., et al. (2015, Apr). Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. *Journal of the National Cancer Institute*, 107 (5), djv036–djv036.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P., Harshman, K., Tavtigian, S., et al. (1994, Oct). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266 (5182), 66–71.

Olama, A. A. A., Benlloch, S., Antoniou, A. C., Giles, G. G., Severi, G., Neal, D. E., et al. (2015, Apr). Risk Analysis of Prostate Cancer in PRACTICAL a Multinational

Consortium Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiology Biomarkers & Prevention*, 24 (7), 1121–1129.

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., et al. (2016, Sep). Major depressive disorder. *Nature Reviews Disease Primers*, 2, 16065.

Palmer, J., & Rutherford, A. (2016, Aug). How has DNA shaped the human race? *The Economist*. http://www.economist.com/news/international/21705401-jason-palmer-editor-espresso-daily-briefing-app-joined-geneticist-and-broadcaster-adam

Pashayan, N., Duffy, S. W., Neal, D. E., Hamdy, F. C., Donovan, J. L., Martin, R. M., et al. (2015, Jan). Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genetics in Medicine*, 17 (10), 789–795.

Pulit, S. L., McArdle, P. F., Wong, Q., Malik, R., Gwinn, K., Achterberg, S., et al. (2016, Feb). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *The Lancet Neurology*, 15 (2), 174–184.

Rice Warner (2011). Underinsurance in Australia Research Report 2011.

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., et al. (2012, Apr). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, 18 (4), 497–511.

Risch, N., & Merikangas, K. (1996, Sep). The future of genetic studies of complex human diseases. *Science*, 273, 1516–1517.

Sacco, R. L., Benjamin, E. J., Broderick, J. P., Dyken, M., Easton, J. D., Feinberg, W. M., et al. (1997, Jul). Risk Factors. *Stroke*, 28 (7), 1507–1517.

Schaller, R. (1997, Jun). Moore's law: past present and future. *IEEE Spectrum*, 34 (6), 52–59.

Seyerle, A., & Avery, C. (2013, Nov). Genetic epidemiology: the potential benefits and challenges of using genetic information to improve human health. *North Carolina Medical Journal*, 74, 505–508.

Spiliopoulou, A., Nagy, R., Bermingham, M. L., Human, J. E., Hayward, C., Vitart, V., et al. (2015, Apr). Genomic prediction of complex human traits: relatedness trait architecture and predictive meta-models. *Human Molecular Genetics*, 24 (14), 4167–4182.

Tada, H., Melander, O., Louie, J. Z., Catanese, J. J., Rowland, C. M., Devlin, J. J., et al. (2015, Sep). Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *European Heart Journal*, 37 (6), 561–567.

Tada, H., Shiffman, D., Smith, J. G., Sj□bgren, M., Lubitz, S. A., Ellinor, P. T., et al. (2014, Aug). Twelve–Single Nucleotide Polymorphism Genetic Risk Score Identifies Individuals at Increased Risk for Future Atrial Fibrillation and Stroke. *Stroke*, 45 (10), 2856–2862.

Venter, J., Adams, M., Myers, E., Li, P., Mural, R., Sutton, G., et al. (2001, Feb). The sequence of the human genome. *Science*, 291, 1304–1351.