Journal of Health Economics 81 (2022) 102556

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/jhealeco

# Genetic risk scores in life insurance underwriting

Richard Karlsson Linnér[a,b,*], Philipp D. Koellinger[a,c]

[a] School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, Amsterdam 1081HV, the Netherlands
[b] Department of Economics, Leiden University, Steenschuur 25, Leiden 2531ES, the Netherlands
[c] La Follette School of Public Affairs, University of Wisconsin-Madison, 1225 Observatory Dr., Madison, WI 53706, USA.

A B S T R A C T

Genetic tests that predict the lifetime risk of common medical conditions are fast becoming more accurate and affordable. The life insurance industry is interested in using predictive genetic tests in the underwriting process, but more research is needed to establish whether this nascent form of genetic testing can refine the process over conventional underwriting factors. Here, we perform Cox regression of survival on a battery of genetic risk scores for common medical conditions and mortality risks in the Health and Retirement Study, without returning results to participants. Adjusted for covariates in a relevant insurance scenario, the scores could improve mortality risk classification by identifying 2.6 years shorter median lifespan in the highest decile of total genetic liability. We conclude that existing genetic risk scores can already improve life insurance underwriting, which stresses the urgency of policymakers to balance competing interests between stakeholders as this technology develops.

## 1. Introduction

Predictive genetic testing for common medical conditions, such as cardiovascular disease, is an emerging technology (Torkamani et al., 2018). The first clinical trials are underway (Fiorentino, 2021), and people can since recently purchase prototype versions directly from consumer genetics services. This development is both a blessing and a curse for the life insurance industry (Rechfeld et al., 2019). On the upside, insurers could potentially use this new source of information to refine the underwriting process. Eventually, precision medicine can improve population health (Zeggini et al., 2019), which could reduce the number of death claims (Lefebre et al., 2019). However, existing bans on insurers' use of genetic information (Bélisle-Pipon et al., 2019), which are meant to prevent genetic discrimination, can result in regulatory adverse selection (Polborn et al., 2006). Due to the limited evidence on whether predictive genetic tests are actuarially justified, an expert group of academics and insurance stakeholders has called for more research (Joly et al., 2014). This study investigates empirically whether polygenic scores, a predictive test based on genetic risk scoring, can improve mortality risk classification in a relevant life insurance scenario.

Polygenic scores are particularly promising for predicting the lifetime risk of developing common medical conditions (Zeggini et al., 2019). Together, this disease category accounts for most of the mortality burden in industrialized countries (Jan et al., 2018), thereby being of great relevance to the life insurance industry. Polygenic scores capture a person's genetic liability towards a specific trait or condition by summarizing the weight of a large number of variants throughout the genome. This approach is motivated by the recent understanding that the substantial heritability of common medical conditions is spread across thousands, or perhaps millions, of variants with individually small effects (Visscher et al., 2021). The trait-specific weights required as inputs are estimated in genome-wide association studies (GWAS), many of which now include genetic data from more than a million people (Tam et al., 2019). Existing polygenic scores based on large-sample GWAS can usually explain up to a few percent of the variation ($R^2$) in the relevant trait, and heritability estimates ($h^2$) are indicative of the asymptotic upper bound. With a single cost-effective genetic assay, additional polygenic scores can be generated at low marginal cost. The life insurance industry monitors this development closely and are interested in using polygenic scores in the underwriting process (Widmer, 2019).

* Corresponding author at: Department of Economics, Leiden University, Steenschuur 25, Leiden 2531ES, the Netherlands.
*E-mail addresses:* r.karlsson.linner@law.leidenuniv.nl (R. Karlsson Linnér), koellinger@wisc.edu (P.D. Koellinger).

https://doi.org/10.1016/j.jhealeco.2021.102556
Received 23 March 2020; Received in revised form 3 November 2021; Accepted 12 November 2021
Available online 15 November 2021

The literature on clinical stratification with polygenic scores is extensive and growing (for a review, see, e.g., Lambert et al., 2019). A handful of studies have shown that polygenic scores not only predict disease incidence, but also mortality rates (Ganna et al., 2013; Marioni et al., 2016; Pilling et al., 2017; McDaid et al., 2017). In particular, Timmers et al. (2019) developed an enhanced polygenic score for parental lifespan (a common proxy in genetic studies on longevity) that could distinguish a 3.5-year difference in median lifespan between the extreme deciles of the score distribution. This finding suggests that polygenic scores may already be actuarially justified, but existing studies have only adjusted estimates for the underwriting factors age and sex. Conceptually, polygenic scores capture a variety of direct and indirect effects, some of which act through other observables that are already available to insurers. Maxwell et al. (2020) found that polygenic scores for breast cancer and cardiovascular disease could improve the prediction of disease incidence above a set of underwriting factors, but to our knowledge, this has not been shown for mortality risk in the context of life insurance. Our main contribution is to test whether a battery of polygenic scores can improve the classification of mortality risk after adjustment for a comprehensive set of underwriting factors.

We see two main motivations for studying genetic testing and insurance: (1) the risk of adverse selection as more people acquire private genetic information, and (2) the possibilities and consequences of using genetic information to determine premiums or rejections. The analyses we report are mostly motivated by (2), and we do not provide estimates of adverse selection. We reviewed the economic literature on genetic testing and insurance. In an early study, Tabarrok (1994) introduced the interesting idea of making a genetic insurance mandatory prior to genetic testing, which would mitigate any increased premiums resulting from a burdening result. Theoretical models developed by Doherty and Thistle (1996), Hoy and Polborn (2000), Dionne and Rothschild (2014), Peter et al. (2017), and Bardey et al. (2019) generally show that people are more incentivized to undergo testing under genetic information bans and that such bans should, in theory, intensify adverse selection. However, studies by Polborn et al. (2006), Hoel et al. (2006), and Hoy and Witt (2007) disagree on whether information bans increase consumer welfare. Hoy and Ruse (2005) review moral arguments in support of regulation, but they conclude that existing bans must be reviewed periodically in light of future accumulation of private genetic information.

Further, Macdonald and Tapadar (2010) remark that the economic literature has almost exclusively modelled diagnostic tests for rare genetic disorders (e.g., Huntington's disease) or mutations (e.g., *BRCA* variants with drastic effects on breast cancer). They show that predictive genetic tests for common medical conditions are unlikely to lead to as strong adverse selection. Similarly, Hoy and Lambert (2000) were early in recognizing that genetically multifactorial conditions and the lower accuracy of predictive tests are complicating factors for theoretical models that predict strong adverse selection or insurance market failure. They show that genetic information can improve discrimination between true risk types, but also that even very accurate tests can lead to unfair discrimination of some applicants due to misclassification. Overall, the economic literature emphasizes the risk of adverse selection under genetic information bans, but disagrees on whether predictive genetic tests constitute a major problem and whether there is actually a positive net effect of bans on consumer welfare.

In this preregistered study[a], we generated polygenic scores for 27 common medical conditions and mortality risks among 9,272 genotyped respondents in the Health and Retirement Study. Our main analysis is a series of nested Cox proportional-hazards regressions of survival on the polygenic scores together with covariates for conventional underwriting factors (Cox, 1972). To model a relevant insurance scenario, we deliberately selected covariates to reflect the type of information that is routinely collected in a personal health declaration, which is the main source of information for issuing standard amounts of life insurance to healthy adult applicants (see Section 2. Life insurance and the underwriting process). A retrospective cross-section was inferred, assuming that respondents were observed at age 35 (the modal age of applicants). The coefficients estimated in the Cox regressions were used to combine the polygenic scores into a polygenic hazard index, which could detect a 2.6-year shorter median lifespan in its top decile, independent of the covariates. We generated empirical 10-year cumulative mortality rates and benchmarked the polygenic hazard index against each of some conventional underwriting factors. Our findings show that existing polygenic scores, when combined, can already classify mortality risk similar to some conventional factors, such as sex and former smoking.

Our study also contributes to the economic literature on the determinants of mortality risk, see e.g., Gerdtham and Johannesson (2004), Balia and Jones (2008), or Palme and Sandgren (2008); as well as the actuarial literature on the association between observable characteristics and insurance applications, claims, or pricing, like those by Sijbrands et al. (2009), McCrea and Farrell (2018), Sari et al. (2019), or Maxwell et al. (2020). Most of these studies employed a Cox regression framework similar to ours. Finally, the study is related to a recently expanding literature that applies polygenic scores to investigate inequalities in health or socioeconomic outcomes, many of which analyzed HRS data. For example, , Barth et al., 2020) found that polygenic scores are associated with wealth and financial decision making. Typically, the latter literature concentrates on a single polygenic score for a variable of interest (often educational attainment), while the present study differs by generating polygenic scores for a range of medical conditions and mortality risks to assess their joint capacity to classify mortality risk.

The study is accompanied by **Supplementary Information** that provides a brief introduction to molecular genetics research and reports some materials omitted for brevity. The remainder of the article is organized as follows. Section 2 introduces the reader to modern practices in life insurance underwriting and the personal health declaration that we intend to model in our insurance scenario. Section 3 introduces the data, sample restrictions, and the polygenic scores. Section 4 reports an exploratory univariate survival analysis. In Section 5, we detail our main analysis, which is a series of nested Cox regressions of survival. In Section 6, we utilize the coefficients estimated in the Cox regressions to combine the polygenic scores into a polygenic hazard index. We then

[a] The preregistered study protocol is available at: https://osf.io/c7uem/

benchmark the polygenic hazard index in its capacity to classify mortality risk, generate empirical 10-year cumulative mortality rates, and perform a cross-validation exercise. Section 7 reports discussion, conclusion, and a policy recommendation.

## 2. Life insurance and the underwriting process

Life insurance protects a beneficiary from economic loss upon the death of the insured. An essential component of life insurance underwriting is mortality risk classification, which stratifies applicants into groups of similar mortality risk (Cummins et al. 1983). The practice of using observable characteristics to determine premiums and rejections is considered vital to prevent adverse selection and to guarantee long-term solvency (Brackenridge et al., 2006). Competition is expected to ensure that premiums remain actuarially fair and that not only the lowest risks can purchase coverage, and there are companies that specialize in insuring severely substandard lives. Although actuarial fairness can be questioned on moral grounds, it is generally viewed to be economically infeasible and inequitable not to price discriminate based on observed risk factors (Cummins et al., 1983; Rothstein, 2004).

An applicant's expected relative mortality is derived using a credit and debit system where the impairment-specific risk of different underwriting factors is summed additively into a risk rating (e.g., +100% relative mortality) (Cummins et al., 1983). If the total risk is deemed insurable and a contract is offered, then the actuarially fair premium for a given rating and age is determined using a sex and birth-year specific life table. A substantially increased mortality risk is required for an applicant to be placed in the first substandard class (often +25–75%), and an increased risk of more than +400% is often considered speculative and may be rejected. Some insurers issue discounted premiums to applicants that present an exceptionally low risk, but this practice is less common. The vast majority (>90%) of life insurance contracts get issued at the standard rate, and only about 2–4% of applications get rejected (Brackenridge et al., 2006).

Life insurance can be purchased at virtually any age, but is most often bought by people in their mid-thirties to early forties (Brackenridge et al., 2006). It has been observed that the average life insurance applicant is healthier and longer-lived than the general population. The age of the applicant and the size of the death benefit jointly determine whether information will be collected through personal health declaration or medical examination. The willingness of insurers to pay for medical examination or other costly examination procedures, e.g., laboratory analyses of blood or urine, increases with the applicant's age and the insured amount. Nowadays, it is standard practice for insurers to issue normal levels of coverage (e.g., up to U.S. $100,000) to overall healthy adult applicants (e.g., up to age 50) without a medical exam (Brackenridge et al., 2006). In the insurance scenario we develop below, because of limited data on medical history, we assumed to model a personal health declaration rather than a full medical examination.

Competition forces the insurance industry to develop better underwriting factors, but the industry is also pressured to apply evidence-based underwriting, which in practice limits the universe of potential factors to those that are actuarially justified by the medical literature (Cummins et al., 1983; Nabholz and Somerville, 2011). Diagnosed medical conditions, as well as severe illness or premature death (e.g., before age 65) among close family members, are particularly important, and this category contains an abundance of risk classes for various medical conditions and their subtypes (Brackenridge et al., 2006). It is standard practice in most countries to request information about family members to infer hereditary disorders. Most questions on a personal health declaration concerns the applicant's health and medical history, while demographic underwriting factors include marital status, education, place of residence, body mass index (BMI), hazardous occupations[b] or military service, smoking, heavy drinking or drug use, a few hazardous activities (e.g., some forms of aviation or sky diving), and sometimes also motor vehicle records. Some information, e.g., on income, is mainly collected to determine insurable interest and that the amount is reasonable in relation to financial status, as remarkably large amounts can signal speculation (Brackenridge et al., 2006).

There are many scholars who argue that insurance companies are already entrusted with handling private and sensitive information, such as medical journals and tests, and that there is no reason to expect them to handle genetic information with less prudence (Rothstein, 2004). But other scholars argue that genetic information must be treated differently so as not to disincentivize people from genetic testing in healthcare or medical research, which could be detrimental to population health (Rothstein, 2018; Bélisle-Pipon et al. 2019). The regulatory landscape varies between countries and is still shifting, with some governments imposing complete bans on genetic information in life insurance, while others permit the use of existing test results for certain disorders or large insurance amounts (Prince, 2019). In many countries, insurers are restricted from requesting new genetic testing of applicants (Rechfeld et al., 2019).

## 3. Data

The Health and Retirement Study (HRS) is a longitudinal household survey of elderly Americans, conducted biannually since 1992 (Sonnega et al., 2014). The purpose of the HRS is to facilitate studies on the socioeconomic environment in relation to health and aging, for which study participants have provided broad consent. We analyzed existing data resources provided in the Public HRS Survey Data and the RAND HRS Longitudinal File 2016[c], together with restricted-access genetic data that are available upon request from the National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP). Overall, 13 survey waves were available, covering the years 1992–2018. In this study, respondents were not recontacted and no information was returned to them.

[b] Occupation is becoming a less important underwriting factor and few modern occupations lead to higher premiums because of the steady decline in hazardous occupations in industrialized countries (Brackenridge et al., 2006).

[c] The data are available at https://hrsdata.isr.umich.edu/data-products/public-survey-data

**Table 1**
Sample descriptive statistics.

| | | The genotyped sample of European HRS respondents | | All European HRS respondents | |
|---|---|---|---|---|---|
| Variable | Type | Complete observations | %, Mean (SD), or range | Complete observations | %, Mean (SD), or range |
| Sex (0 = Male; 1 = Female) | Binary | 9,272 | 56.47% | 27,345 | 55.18% |
| Birth year | Categorical | 9,272 | 1920–1955 | 27,345 | 1890–1995 |
| Age at genotyping | Continuous | 9,272 | 68.7 (9.1) | | |
| Deceased by most recent wave (2016–2018) | Binary | 9,272 | 25.15% | 27,345 | 41.32% |
| Deceased before genotype data collection (<2006) | Binary | 9,272 | 0.0% | 27,345 | 21.72% |
| Years of education (17+ coded as 17 years) | Continuous | 9,259 | 13.2 (2.6) | 27,199 | 12.8 (2.9) |
| Ever married | Binary | 9,272 | 97.06% | 26,225 | 95.91% |
| Ever active military service | Binary | 9,266 | 25.40% | 27,333 | 23.56% |
| Heavy drinking (≥5 drinks on a day when drinking) | Binary | 9,272 | 2.21% | 27,345 | 2.70% |
| Former smoker | Categorical | 9,272 | 38.05% | 27,345 | 37.77% |
| Current smoker | Categorical | 9,272 | 18.86% | 27,345 | 21.03% |
| Body Mass Index (BMI) | Continuous | 9,268 | 27.1 (5.1) | 27,284 | 26.9 (5.4) |
| Maternal mortality by age 65 | Binary | 9,272 | 14.51% | 27,345 | 16.44% |
| Paternal mortality by age 65 | Binary | 9,272 | 24.81% | 27,345 | 24.81% |
| Self-rated health growing up: Excellent | Categorical | 9,269 | 53.41% | 24,241 | 51.73% |
| Self-rated health growing up: Very good | Categorical | 9,269 | 27.12% | 24,241 | 27.05% |
| Self-rated health growing up: Good | Categorical | 9,269 | 14.01% | 24,241 | 15.29% |
| Self-rated health growing up: Fair | Categorical | 9,269 | 4.23% | 24,241 | 4.44% |
| Self-rated health growing up: Poor | Categorical | 9,269 | 1.22% | 24,241 | 1.49% |
| Childhood disability | Binary | 8,944 | 4.28% | 16,981 | 4.16% |
| Childhood serious health problem | Binary | 8,952 | 18.64% | 16,986 | 18.57% |
| Diabetes by age 35 | Binary | 9,272 | 0.86% | 27,345 | 0.71% |
| Allergic condition in childhood | Binary | 6,266 | 8.55% | 13,541 | 12.09% |
| Asthma in childhood | Binary | 6,271 | 3.86% | 13,551 | 5.20% |
| Chicken pox in childhood | Binary | 5,989 | 82.23% | 13,005 | 82.08% |
| Depression in childhood | Binary | 6,270 | 2.63% | 13,547 | 3.97% |
| Epilepsy or seizures in childhood | Binary | 6,275 | 0.59% | 13,567 | 0.97% |
| Severe headaches or migraines in childhood | Binary | 6,272 | 5.74% | 13,561 | 6.50% |
| Heart problems in childhood | Binary | 6,271 | 2.18% | 13,559 | 2.30% |
| Measles in childhood | Binary | 6,068 | 88.81% | 13,064 | 76.55% |
| Mumps in childhood | Binary | 6,020 | 67.66% | 12,973 | 60.69% |
| Respiratory disorder in childhood | Binary | 6,268 | 12.09% | 13,543 | 15.01% |
| Stomach problems in childhood | Binary | 6,273 | 3.35% | 13,559 | 4.39% |

*Notes:* **Supplementary Table 1** provides additional details, as well as descriptive statistics for the two variables "Job code for job with longest tenure" and "Census division" (i.e., region of residence).

The HRS has collected longitudinal data from the demographic; family; health and medical; education, occupation, income and wealth; and retirement domains. The quality of the HRS mortality data has been validated by Weir (2016) and judged effectively complete. As described by Weir, death and date of death are recorded by the HRS in two ways: (1) the next-of-kin of non-responders are contacted for an exit interview to distinguish death from sample attrition, or (2) by query of the National Death Index (NDI+) at the end of each wave. In 2006, HRS also began collecting genotype data. We analyzed the second release of genetic data, which had been imputed with the 1000 Genomes Project phase 1 (version 3) reference panel.

### *3.1. Sample exclusions and mortality selection*

We restricted the sample to respondents who self-report to be "White/Caucasian" and "Not Hispanic", which we refer to as European ancestry[d]. The reason for this restriction is that the vast majority of existing GWAS have analyzed European-ancestry samples, which prohibitively limits their transferability to other ancestries. The genetic data includes 15,620 genotyped respondents, of which 10,596 are of European ancestry and passed our exclusion criteria to remove ancestral outliers (**Supplementary Information**).

Next, to exclude birth years with an inadequate number of complete observations, we restricted the study sample to individuals born between 1920 and 1955, which were between 51 and 90 years old at the subject-specific genotyping time point in 2006–2010. At this stage, we retained 9,272 respondents (hereafter referred to as the "genotyped sample"), of whom 2,332 (25.2%) were deceased by the most recent wave. Table 1 reports descriptive statistics for the genotyped sample in comparison with all European HRS respondents.

We investigated the possibility of mortality selection due to the delayed start of the genetic data collection and found that genotyped respondents are, on average, healthier and longer-lived than non-genotyped respondents (**Supplementary Fig. 1**). The presence of mortality selection should primarily increase the type II error (false negative) rate (Domingue et al., 2017). Therefore, our analyses

[d] We adhere to the standard definition of European ancestry applied in genetic epidemiology, which distinguishes "Hispanic/Latin American".

may understate the influence of polygenic scores for severe medical conditions that have an age of onset around or below the age of genotyping in the HRS (e.g., scores for cardiovascular disease traits).

### 3.2. Polygenic scores for common medical conditions and mortality risks

To generate polygenic scores, we searched the GWAS literature to identify the largest available studies, with at least $N > 100{,}000$, on a pre-registered set of common medical conditions and mortality risks. Following best practices from molecular genetics, a harmonized quality-control protocol was applied to the GWAS regression coefficients used as weights to create polygenic scores, and genetic principal components (PCs) were generated to be used as control variables for potential population stratification (Price et al., 2006) (**Supplementary Information**).

In notation, the $i$th respondent's polygenic score for the $k$th trait, $\hat{S}_{ik}$, was computed as

$$\hat{S}_{ik} = \sum_{j=1}^{M} \hat{\beta}_{jk} g_{ij} \tag{1}$$

where the respondent's genotype $g_{ij}$ at variant $j$ (coded as having 0, 1, or 2 effect-coded alleles[e]) was weighed by its corresponding trait-specific GWAS weight, $\hat{\beta}_{jk}$, and then summed across a large set of $M$ genetic variants (~1 million). Since GWAS are estimated as repeated regression applied individually to each genetic variant, we first applied the PRS-CS method that approximates multiple regression to derive adjusted weights (Ge et al., 2019). Such post-GWAS approaches are considered more computationally efficient compared to applying sophisticated techniques already at the GWAS stage. Importantly, this method allows for inclusion of correlated variants, leading to more comprehensive genome-wide coverage.

We generated polygenic scores for a total of 27 common medical conditions and mortality risks (**Supplementary Table 2**). Examples of common medical conditions are Alzheimer's disease, breast cancer, and coronary artery disease, and examples of mortality risks are cigarettes per day (i.e., smoking intensity), educational attainment, and cholesterol. By virtue of the central limit theorem and the large number of variants, the polygenic scores were approximately normally distributed. Because of estimation error in GWAS of finite sample size, polygenic scores will be noisy approximations of their "true" counterparts. This situation is akin to traditional measurement error. Polygenic scores and genetic PCs were standardized to mean zero and unit variance. We estimated customary genetic correlations and found that most of the 27 traits exhibit genetic overlap with parental lifespan, suggesting that they could be predictive of mortality risk (**Supplementary Table 3** and **Supplementary Fig. 2**).

## 4. Exploratory univariate survival analysis

Our empirical analyses begin with an exploratory univariate analysis of respondent, maternal, and paternal survival, by nonparametric estimation of stratified Kaplan-Meier functions (Kaplan and Meier, 1958). A benefit of analyzing parental survival is reduced censoring and mortality selection, but a disadvantage is that genetic data on offspring only partly capture parental genotypes. Another drawback is that mortality risk factors may change over time or between generations. We analyzed monthly survival in respondents, but could only observe yearly survival for the parents. Parental survival was only analyzed in the exploratory analysis.

To reduce researcher degrees of freedom, we had preregistered that we would evaluate each polygenic score by its top decile (i.e., highest genetic liability) versus either (a) its lower nine deciles, or (b) its bottom decile. Comparison (a) was deliberately chosen to mirror insurance underwriting, while comparison (b) is an extreme-groups approach frequently applied in the medical literature. We report the results of both comparisons but focus our presentation on (a) because it utilizes all observations. We believe that the decision to evaluate by the top decile is a reasonable trade-off between greater genetic liability against smaller group size (e.g., the top 5% includes less than 500 respondents). For most of the polygenic scores, we expect to observe a higher mortality rate in the top decile, but for a few positive lifespan correlates, like educational attainment or parental lifespan, we instead expect a lower mortality rate.

### 4.1. Results of the exploratory univariate survival analysis

Among respondents, for 15 of the 27 polygenic scores, comparison (a) finds significantly different survival functions at the conventional level ($P < 0.05$) (**Supplementary Fig. 3** and **Supplementary Table 4**). Each of these 15 scores distinguishes an appreciable difference in median lifespan, defined here as ≥1 year. The detected differences were in the anticipated direction, e.g., the median lifespan in the top decile of the score for parental lifespan was 89.9 years ($CI_{95\%}$ = [89.4–91.1]) versus 88 years ($CI_{95\%}$ = [87.7–88.3]) in the lower nine deciles, which is a difference of 1.9 years. The largest difference was identified by the polygenic score for smoking initiation, followed by those for parental lifespan, cigarettes per day, and educational attainment. As expected, in comparison (b), the extreme-groups approach generally detected larger differences in median lifespan, but at the cost of wider confidence intervals due to fewer observations (**Supplementary Table 4**).

As for the parents, the score for parental lifespan detected the largest difference in median lifespan (**Supplementary Fig. 4–5**), and the pattern of association was largely similar to that of respondent survival (**Supplementary Table 4**). A notable dissimilarity is

[e] In the standard model of additive genetic effects, the choice of effect-coded allele is arbitrary and can be changed by simply reversing the sign of the GWAS coefficient (**Supplementary Information**).

that the scores for several cardiovascular disease traits; like atrial fibrillation, coronary artery disease, and HDL cholesterol; detected significant differences among parents but not respondents. We believe that mortality selection in our genotyped sample explains this dissimilarity. The following six polygenic scores did not show an apparent influence on mortality in any group: chronic kidney disease, drinks per week, LDL cholesterol, prostate cancer, schizophrenia, and total cholesterol. Overall, the results of the univariate analysis suggest that many of the polygenic scores contain some information on mortality risk, motivating their inclusion in the multiple regressions, described next.

## 5. Cox proportional hazards regression of survival

Our main analysis is a series of multiple regressions of respondent survival, consisting of four nested Cox proportional hazards models (Cox, 1972). The dependent variable in Cox regression is the hazard function, $\lambda$, defined as the instantaneous probability of event (death), centered on a time point in the observation period. Here, we model the probability of dying in a particular month of observation, letting the subject-specific genotyping time point in 2006–2010 define the model time origin ($t = 0$). The observation period is right-censored by the most recent wave of data collection that ended in 2018, providing up to 136 months of observation.

Formally, a general specification of the Cox regression model with time-invariant coefficients can be written as

$$\lambda(t;\ Z) = \lambda_0(t) \times \exp(Z\alpha) \tag{2}$$

where the hazard $\lambda$ is defined as a function of time $t$ and $Z$, an $N \times P$ matrix with regressor values for $N$ individuals and $P$ regressors. The $P \times 1$ vector $\alpha$ denotes unknown coefficients to be estimated with maximum partial likelihood estimation. Here, we break with convention and denote these coefficients with $\alpha$, reserving $\beta$ for GWAS coefficients. In the general specification, the baseline hazard $\lambda_0$ is a function of time, derived non-parametrically from the data. We estimated robust standard errors (Lin and Wei, 1989), clustered at the household level.

### 5.1. Covariate selection to model a personal health declaration

In our insurance scenario, a personal health declaration was modelled using retrospective questions to infer a cross section that could realistically be observed at age 35, which is about the modal age of applicants (Section 2). Covariates were selected to closely reflect the customary information collected in a health declaration while maximizing the sample size. Unfortunately, some retrospective questions had only been surveyed among smaller subsets of respondents, motivating a couple of proxies. We consider these proxies to be reasonable approximations of past values that preserve the sample size. E.g., BMI in the first wave correlates at ~0.74 with retrospective BMI at age 45, while the latter was only available for ~700 respondents. Unless stated otherwise, retrospective questions with repeated observations in multiple waves were defined as the most frequent value, while proxies were defined using the first wave with nonmissing data to be as close in time as possible to our inferred cross section. Variables on "childhood/growing up" refer to the period from birth to age 16.

We aimed to model as many medical conditions as possible, while acknowledging the difficulty of exhaustively capturing medical history without observing actual medical records. Among the survey questions for medical conditions in adulthood, diabetes was the only condition for which we observed respondents with a diagnosis by age 35. Nonetheless, our selection of covariates is similar to previous studies in insurance economics that analyzed HRS data (Meyricke and Sherris, 2013). As in previous works, we did not consider lifestyle factors beyond drinking, smoking, and BMI because of the difficulty for insurers to accurately measure and rate such factors. Due to the arguably narrow set of variables collected for medical conditions and family history, we mostly consider the insurance scenario to generalize to healthy applicants with no alarming family or medical history. However, we believe that insurers are particularly interested in observing predictive genetic information for this group.

#### 5.1.1. Life table parameters and demographic underwriting factors

As insurance companies evaluate the expected relative mortality for a given age by using a sex and birth-year specific life table, we categorized covariates for sex and birth year as "life table parameters", and we consider additional controls for birth month and age (at the time of genotyping) to belong to this category. Below we evaluate the relative performance of this category separately from the demographic and medical underwriting factors, described next.

We defined the following demographic underwriting factors: marital status (ever/never), years of education, job code for longest tenure, ever active military service, heavy drinking ($\geq$ 5 drinks when drinking), former or current smoking, BMI, and region of residence. Due to limited retrospective data, marital status was proxied with the first non-retrospective observation; years of education with the highest reported education; heavy drinking, smoking status, and BMI with the first non-retrospective observation; and region of residence with the first observed census division (i.e., nine regions defined by the U.S. Census Bureau). These variables capture with varying precision most standard nonmedical underwriting factors, except drug use, hazardous activities, and motor vehicle records, which were not available.

#### 5.1.2. Medical underwriting factors

We defined the following medical underwriting factors: self-rated health growing up on a five-point ordinal scale; two binary indicators for maternal or paternal mortality by age 65 (observed by most respondents at age 35); two binary indicators for having

had a medical condition[f] in childhood that led to either (a) a disability for six months or more, or (b) an important or serious health problem; being diagnosed with diabetes by age 35 (we merged measures of diabetes in childhood and adulthood); and finally, binary indicators for having had either of the following 11 medical conditions in childhood: allergic condition, asthma, chicken pox, depression, epilepsy or seizures, severe headaches or migraines, heart problems, measles, mumps, respiratory disorder, or stomach problems. The latter 11 childhood conditions had only been surveyed among two thirds of respondents. To preserve the sample size, these covariates were only entered in the fourth and last model specification below. The following childhood medical conditions were considered but eventually excluded because of extensive missingness: drugs or alcohol problems, high blood pressure, and other psychiatric problems.

### 5.2. Nested Cox model specifications

Here we detail four nested model specifications represented by $Z$ in the general Cox regression model (Eq. (2)). The regressors were categorized as belonging to either of four non-overlapping sets: (1) the polygenic scores; (2) the genetic PCs, (3) the demographic and medical underwriting factors, and (4) the life table parameters. Every Cox model included the life table parameters (sex, birth year and month, and age at genotyping) to control for such differences, as well as for any potential time effects from the subject-specific age at genotyping, with the intention of making respondents comparable in this regard. All models, except the covariate-first model (2), included 10 genetic PCs as control variables for possible subtle differences in genetic ancestry (see Section 3.3). Specifically, the four Cox models included the following regressors:

Model 1. genetics-first model, which included the polygenic scores (and the genetic PCs);

Model 2. covariate-first model (without genetic PCs), which included marital status, years of education, job with longest tenure, ever active military service, heavy drinking, smoking, BMI, region of residence, maternal and paternal mortality at age 65, self-rated health growing up, childhood disability for six months or more, important or serious health problem in childhood, and diabetes diagnosis by age 35;

Model 3. models (1) and (2) combined; and

Model 4. model (3) together with 11 childhood medical conditions: allergic condition, asthma, chicken pox, depression, epilepsy or seizures, severe headaches or migraines, heart problems, measles, mumps, respiratory disorder, or stomach problems.

The purpose of model (1) is to allow for interpretation of the polygenic scores once conditioned on each other and the genetic PCs; and to interpret the stability of their coefficients once conditioned on the demographic and medical underwriting factors in models (3) and (4). Our main interest is in evaluating the difference in mortality risk classification by going from model (2) to (3), which adds the polygenic scores to the covariate-first model (2). As it maximizes sample size, we consider model (3) to be our preferred model for this evaluation, while the extended model (4) provides a robustness check of whether the scores could add any information over the childhood medical conditions that were entered last due to missing data.

### 5.3. Results of the Cox proportional hazards regressions

Table 2 reports a selection of the estimated coefficients on the log hazard scale ($\hat{\alpha}$) for models (1) to (3). See **Supplementary Table 5** for all coefficients, hazard ratios (i.e., $\exp[\hat{\alpha}]$), and variable-specific tests of the proportional-hazards assumption, for each model. A positive coefficient ($\hat{\alpha}$) means that a regressor is associated with greater mortality risk. In the text, we report some examples of relative mortality (i.e., the percentage difference in the hazard rate) for a unit change in a regressor, computed as $\exp[\hat{\alpha}] - 1$ with corresponding confidence intervals $\exp[\hat{\alpha} \pm 1.96 \times SE] - 1$. In the models, age and sex were always strongly associated ($P < 0.001$), and the mortality rate of women was consistently ~30–40% lower relative to men.

The genetics-first model (1) found nine polygenic scores associated at the conventional level ($P < 0.05$), all of which were also identified in the exploratory analysis. As the polygenic scores were standardized, reported coefficients on the log hazard are per unit of standard deviation. The polygenic score for parental lifespan had the largest coefficient, followed by Alzheimer's disease and smoking intensity (cigarettes per day). For example, a standard deviation increase in the polygenic score for parental lifespan was associated with 9.7% lower relative mortality ($CI_{95\%}$ = [5–14%]). Thus, the coefficient of any individual score could be considered small compared to, say, sex; and similar to or somewhat larger than the effect of an extra year of schooling or unit of BMI in the covariate-first model (2).

Some polygenic scores that showed relatively strong associations in the univariate analysis were not significant in model (1), for example, those for various measures of stroke or blood pressure. Their signal was maybe captured by other genetically correlated scores based on more well-powered GWAS. Two genetic PCs were nominally significant in model (1), but, as we discuss below, not in models (3) and (4) that control for region of residence. The global test of the proportional hazard (PH) assumption did not detect a violation for model (1). The fit statistics suggest that model (1) improved the model fit over a null model with only the baseline hazard (**Supplementary Table 5**). Overall, model (1) corroborates previous studies that showed that polygenic scores predict mortality rates when adjusted only for age and sex.

In the covariate-first model (2), sizable and significant associations were identified in the expected direction for many of the demographic and medical underwriting factors, including years of education, current and former smoking, BMI, paternal mortality,

[f] HRS categorized these childhood medical conditions in distinct categories, e.g., "Cancers and tumors"; or "Heart, circulatory, and blood conditions". But as most of these categories had cells with too few observations or events ($N < 10$), we decided to only include a summarizing binary covariate for having any such childhood condition.

**Table 2**
Cox proportional hazard models of survival.

| Regressors (polygenic scores shown in bold) | Genetics-first model (1) ($N$ = 9,272; 2,332 deaths) Coefficient (Robust SE) | Covariate-first model (2) ($N$ = 8,919; 2,063 deaths) Coefficient (Robust SE) | Combined model (3) ($N$ = 8,919; 2,063 deaths) Coefficient (Robust SE) |
|---|---|---|---|
| **Alzheimer's disease** | 0.097 (0.022)*** | | 0.063 (0.024)** |
| **Atrial fibrillation** | 0.033 (0.023) | | 0.044 (0.024)† |
| **Body Mass Index** | 0.051 (0.023)* | | 0.016 (0.027) |
| **Breast cancer** | 0.036 (0.022) | | 0.044 (0.024)† |
| **Cigarettes per day** | 0.067 (0.022)** | | 0.064 (0.023)** |
| **Depression** | 0.049 (0.024)* | | 0.048 (0.026)† |
| **Educational attainment** | –0.056 (0.023)* | | 0.012 (0.026) |
| **Height** | 0.062 (0.024)* | | 0.064 (0.026)* |
| **Parental lifespan** | –0.103 (0.024)*** | | –0.067 (0.026)** |
| **Smoking initiation** | 0.063 (0.024)** | | 0.014 (0.026) |
| **Type 2 diabetes** | 0.047 (0.023)* | | 0.044 (0.025)† |
| Years of education | | –0.050 (0.010)*** | –0.048 (0.010)*** |
| Ever married | | –0.206 (0.148) | –0.234 (0.148) |
| Ever military service | | 0.124 (0.067)† | 0.127 (0.068)† |
| Smoking (Base: Never) | | | |
| – Former smoker | | 0.209 (0.053)*** | 0.190 (0.054)*** |
| – Current smoker | | 1.069 (0.065)*** | 1.044 (0.065)*** |
| Body Mass Index | | 0.031 (0.005)*** | 0.028 (0.006)*** |
| Paternal mortality by age 65 | | 0.104 (0.051)* | 0.062 (0.052) |
| Health growing up (Base: 1. Excellent) | | | |
| – 2. Very good | | 0.157 (0.052)** | 0.134 (0.053)* |
| – 3. Good | | 0.061 (0.068) | 0.062 (0.068) |
| – 4. Fair | | 0.211 (0.120)† | 0.200 (0.120)† |
| – 5. Poor | | 0.421 (0.177)* | 0.368 (0.177)* |
| Diabetes | | 0.701 (0.206)*** | 0.696 (0.208)*** |
| Sex (1 = Female) | –0.472 (0.041)*** | –0.372 (0.070)*** | –0.398 (0.007)*** |
| Birth year and month, and age at genotyping | Yes | Yes | Yes |
| Genetic PCs | Yes | No | Yes |
| Global test of PH assumption | $\chi^2(85\ df) = 92.8$ | $\chi^2(88\ df) = 131.8^{**}$ | $\chi^2(125\ df) = 154.0^{*}$ |
| Cox-Snell $R^2$ | 0.185 | 0.211 | 0.219 |
| Harrel's $c$ statistic (SE) | 0.744 (0.005) | 0.775 (0.005) | 0.780 (0.005) |
| Gönen & Heller's $K$ statistic (SE) | 0.727 (0.006) | 0.747 (0.004) | 0.752 (0.004) |
| $R^2_D$ Polygenic scores | 2.1% (1.3–3.1%) | Not in (2) | 1.8% (1.1–2.7%) |
| $R^2_D$ Genetic PCs | 0.2% (0.0–0.6%) | Not in (2) | 0.2% (0.0–0.5%) |
| $R^2_D$ Demographic/medical factors | Not in (1) | 9.6% (7.8–11.4%) | 9.6% (7.9–11.5%) |
| $R^2_D$ Life table parameters | 32.7% (30.6–34.8%) | 32.4% (30.3–34.6%) | 32.3% (30.2–34.5%) |

*Notes:* † $P \leq 0.1$; * $P \leq 0.05$: ** $P \leq 0.01$; *** $P \leq 0.001$. Robust standard errors were clustered at the household level. $R^2_D$ is the Royston and Sauerbrei measure of variance explained on the log hazard scale, which we report for four non-overlapping sets of regressors, see Section 5 for details. Life table parameters refer to the covariates for sex, birth year and month, and age at genotyping (see Section 3 for details). Table 2 reports a selection of coefficients for models (1) to (3), see **Supplementary Table 5** for the complete results.

worse childhood health, and diabetes. In particular, current smoking had the largest coefficient among these factors, corresponding to a higher relative mortality of 191%, followed by diabetes with 101%, although the latter estimate is less precise because only 0.9% of the respondents had diabetes by age 35. Virtually none of the job codes were significant in any model, which aligns with the documented trend that occupation is becoming a less relevant underwriting factor in modern economies (Brackenridge et al., 2006). For model (2), we detected a significant violation of the global test of the proportional hazard assumption, which seemed to be localized to a handful of regressors. Model (2) substantially improved the fit over the null model (**Supplementary Table 5**).

Next, we estimated the combined model (3), our preferred model for evaluating whether the polygenic scores could improve the classification of mortality risk. In model (3), four polygenic scores were significant at the conventional level, and these were also significant in model (1). Four more scores had small $P$ values ($P < 0.1$) but were not significant at the conventional level, of which two were significant in model (1). The coefficients of most of the underwriting factors did not change from model (2) to (3). An exception is the coefficient of paternal mortality by age 65, which was now about half. The likelihood ratio test between models (2) and (3) suggested that the addition of polygenic scores significantly improved the fit of the model ($P = 3.42 \times 10^{-7}$). None of the genetic PCs were significant in model (3), likely because this model controls for region of residence. A weak violation of the proportional-hazards assumption was detected for model (3), $P = 0.039$, but we do not think this marginally significant violation strongly limits the interpretability of our preferred model.

In the combined model (3), four observed traits from model (2) were now simultaneously modelled as polygenic scores: BMI, educational attainment, smoking initiation, and diabetes. As expected, these polygenic scores were no longer significant in model (3), while their observed counterparts were all strongly associated ($P < 0.001$). This result exemplifies that polygenic scores can proxy for missing observables, but also serves as a reminder that there is usually much more signal in an observed trait compared to its polygenic score (**Supplementary Information**). However, the polygenic score for type 2 diabetes, which was significant in model

(1), had a similar coefficient and small $P$ value ($P = 0.082$) but was no longer significant in model (3). While not significant, this results suggest that this score could still contribute some additional information on lifetime diabetes that was not captured by diabetes by age 35.

Further, although model (3) was adjusted for observed smoking, the polygenic score for cigarettes per day had a stable and significant coefficient. This may be because this score, similar to diabetes, could contribute information on lifetime exposure that was not fully captured by the discrete covariate. Thus, polygenic scores can also act as complements to variables measured with error or that have a functional form not captured by a discrete covariate. Unfortunately, detailed information on smoking intensity and cessation was not available in an adequately-sized sample. Alternatively, this association might be explained by genetic correlations with other traits; or that never smokers could be influenced via indirect pathways that correlate with genotype, such as parental smoking during pregnancy (Pingault et al., 2018). Following a referee's suggestion, we ad hoc re-estimated model (3) to test for an interaction effect between ever being a smoker with the polygenic score for cigarettes per day. We found that the coefficient of the polygenic score was attenuated by a third among never smokers, but while the interaction term was sizable it had a wide confidence interval. Because of the imprecise estimate, the interaction term was not kept in any of the main models.

Finally, as a robustness check, we estimated the extended model (4). Here, only the polygenic scores for parental lifespan and insomnia were significant, and the scores for cigarettes per day and type 2 diabetes, which were both significant in model (3), had small $P$ values ($P < 0.1$) but were not significant at the conventional level (**Supplementary Table 5**). From the 11 childhood medical conditions added to this model, only severe headaches or migraine was significant ($P = 0.014$), and epilepsy or seizures had a sizable effect but also large standard errors due to few cases ($P = 0.093$). By estimating model (4), we could show that the polygenic scores were able to contribute information also to this extended set of childhood medical conditions, but because of the smaller sample size, this model was not used below to evaluate the improvement in mortality risk classification.

## 6. Benchmark of a polygenic hazard index

In this section, we combine the effects of multiple regressors into indexes in order to evaluate their respective share of variance explained, and to evaluate the combined capacity of the polygenic scores to classify mortality risk compared to that of the demographic and medical underwriting factors. To combine the total genetic liability of the polygenic scores into a weighted index on the log hazard scale, we used the covariate-adjusted coefficients from a particular Cox model as weights, following an approach proposed by Royston and Altman (2013). Formally, the $i$th respondent's polygenic "hazard index" was computed as:

$$\widehat{HI}_{PGS,i} = \sum_{k=1}^{K} \hat{\alpha}_k \hat{S}_{ik} \tag{3}$$

where $\widehat{HI}_{PGS,i}$ is the weighted sum of the respondent's polygenic score values, $\hat{S}_{ik}$, weighed by each score's coefficient ($\hat{\alpha}_k$) as estimated in, e.g., the genetics-first model (1). We hereafter omit the subject index and the circumflex symbol for readability. The HIs for the three remaining non-overlapping regressor sets—the demographic and medical underwriting factors ($HI_{COVAR}$), the genetic PCs ($HI_{PC}$), and the life table parameters ($HI_{LIFE-TABLE}$)—were computed analogously.

Arguments can be made that the polygenic scores should be evaluated together with the genetic PCs because insurance companies are, in theory, incentivized to use any type of predictive genetic information. However, we decided to evaluate the polygenic scores separately because (1) this is the convention in the epidemiological literature, (2) the pre-registered purpose of the study is to evaluate the predictive capacity of polygenic scores and not genetic data per se, and (3) the variance explained by the PCs was close to zero anyways (Table 2). Similarly, the life table parameters were evaluated separately from the demographic and medical underwriting factors because underwriters first determine the total risk conferred by these factors before looking up said risk for a given age in a sex and birth-year specific life table.

### 6.1. Evaluation of the variance explained on the log hazard scale

For each Cox model, we used the HIs to estimate the Royston & Sauerbrei $R^2_D$ measure of variance explained on the log hazard scale (Royston and Sauerbrei, 2004). This method projects the respondents onto a normal distribution to attain rankits, i.e., expected $Z$-scores based on the number of individuals and their HI rank order. Then, an auxiliary Cox regression is performed on the rankits alone, and the resulting regression coefficient can be transformed into a measure of variance explained. Thus, $R^2_D$ is more similar to the traditional coefficient of determination ($R^2$) of linear regression than to various pseudo-$R^2$ based on log-likelihoods.

Out of the four regressor sets, the life table parameters ($HI_{LIFE-TABLE}$) consistently explained the largest share of variance: $R^2_D \sim 32$–$33\%$ (see Table 2 for confidence intervals). The second largest share was explained by demographic and medical underwriting factors ($HI_{COVAR}$), e.g., $R^2_D = 9.6\%$ in models (2) and (3). The polygenic scores ($HI_{PGS}$) came in third with $R^2_D = 2.1\%$ in model (1), and 1.8% in model (3). Thus, the demographic and medical underwriting factors explained about five times more variation than the polygenic scores did. The variance explained by the genetic PCs was always close to zero. The $R^2_D$ for the demographic and medical underwriting factors was basically the same in models (2) and (3), suggesting that the polygenic scores did not reduce the signal of the underwriting factors. On the other hand, the variance explained by the polygenic scores decreased somewhat going to model (3), suggesting that part of their signal was adjusted away by the underwriting factors.

**Table 3**
Mortality risk classification with hazard indexes (HIs).

| Panel A. Polygenic hazard index based on the coefficients of the genetics-first model (1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Stratum $HI_{PGS}$ | Stratum $HI_{COVAR}$ | *N* | Events | Median (years) | 95% lower CL | 95% upper CL | Difference in median | Log-rank *P* |
| Lower nine deciles | – | 8,345 | 2,009 | 88.6 | 88.2 | 88.9 | 4.6 | **0.000** |
| Top decile | – | 927 | 323 | 84.0 | 82.8 | 85.2 | | |
| Panel B. Demographic and medical hazard index based on the coefficients of the covariate-first model (2) | | | | | | | | |
| Stratum $HI_{PGS}$ | Stratum $HI_{COVAR}$ | *N* | Events | Median (years) | 95% lower CL | 95% upper CL | Difference in median | Log-rank *P* |
| – | Lower nine deciles | 8,028 | 1,728 | 89.2 | 88.8 | 89.7 | 8.1 | **0.000** |
| – | Top decile | 891 | 335 | 81.1 | 80.3 | 82.0 | | |
| Panel C. Both hazard indexes based on the coefficients of the combined model (3) | | | | | | | | |
| Stratum $HI_{PGS}$ | Stratum $HI_{COVAR}$ | *N* | Events | Median (years) | 95% lower CL | 95% upper CL | Difference in median | Log-rank *P* |
| Lower nine deciles | Lower nine deciles | 7,259 | 1,510 | 89.6 | 89.2 | 90.0 | | **0.000** |
| Top decile | Lower nine deciles | 769 | 213 | 87.0 | 85.3 | 88.0 | 2.6 | |
| Lower nine deciles | Top deciles | 777 | 291 | 81.5 | 80.8 | 83.1 | 8.1 | |
| Top decile | Top decile | 114 | 49 | 78.9 | 77.2 | 80.2 | 10.7 | |
| Panel D. Robustness analysis with demographic and medical hazard index unadjusted for the polygenic scores | | | | | | | | |
| Stratum $HI_{PGS}$ | Stratum $HI_{COVAR}$ | *N* | Events | Median (years) | 95% lower CL | 95% upper CL | Difference in median | Log-rank *P* |
| Lower nine deciles | Lower nine deciles | 7,263 | 1516 | 89.6 | 89.2 | 90.0 | | **0.000** |
| Top decile | Lower nine deciles | 765 | 212 | 87.0 | 85.3 | 88.0 | 2.6 | |
| Lower nine deciles | Top deciles | 773 | 285 | 81.4 | 80.6 | 83.0 | 8.2 | |
| Top decile | Top decile | 118 | 50 | 78.9 | 77.2 | 80.8 | 10.7 | |

*Notes:* Median (years) refers to the observed median lifespan in each stratum. 95% lower and upper CL refers to the confidence limits of the median lifespan. Difference in median is the difference in median lifespan compared with the top stratum in each panel. Log-rank *P* tests the null hypothesis of no difference in survival functions.

### 6.2. Mortality risk classification with polygenic hazard index

The distributions of the two hazard indexes of special interest, $HI_{PGS}$ and $HI_{COVAR}$, were used to categorize respondents into low- and high-risk groups for mortality risk classification. Specifically, following the preregistration, respondents in the lower nine deciles were classified as low risk, and those in the top decile as high risk. Thus, for the genetics-first model (1) and the covariate-first model (2), we categorized respondents by the hazard index relevant to each respective model. Then, for the combined model (3), we instead classified four groups resulting from the two-way combination of low/high $HI_{PGS}$ and/or $HI_{COVAR}$. Here, a total of 114 respondents were categorized as high risk by both indexes, while 769 and 777 were denoted as high risk by $HI_{PGS}$ and $HI_{COVAR}$, respectively. The remaining 7,259 respondents were classified as low risk. We estimated stratified survival functions and compared the median lifespan of each group, which is a meaningful actuarial measure (Brackenridge et al., 2006). The results are reported in Table 3. In each of the following comparisons, the 95% confidence limits never overlapped, suggesting satisfactory precision.

When $HI_{PGS}$ was based on the coefficients of the genetics-first model (1), the median lifespan of low and high $HI_{PGS}$ was 88.6 and 84.0 years, respectively; a 4.6-year difference (Table 3A). Thus, by combining a large number of polygenic scores into a hazard index we were able to improve the performance over previous studies, such as Timmers et al. (2019) that found a 3.5-year difference between the top and bottom decile of a multi-trait enhanced polygenic score for parental lifespan. Next, when $HI_{COVAR}$ was based on the covariate-first model (2), respondents with low and high $HI_{COVAR}$ had median lifespans of 89.3 and 81.2 years, respectively; an 8.1-year difference (Table 3B). This difference is shown by the horizontal distance between **a** and **b** in Fig. 1A. Thus, the demographic and medical underwriting factors could in concert detect about twice the difference in lifespan compared to the polygenic scores, prior to adjustment for each other model (3).

When the hazard indexes were based on the combined model (3), the difference detected by $HI_{COVAR}$ remained the same, independent of $HI_{PGS}$ (Table 3C). In Fig. 1B, this 8.1-year difference is shown by the horizontal distance between **a** and **b** for low $HI_{PGS}$, and **a'** and **b'** for high $HI_{PGS}$. Conversely, independent of $HI_{COVAR}$, the difference detected by $HI_{PGS}$ was 2.6 years, which is the horizontal distance between **a** and **a'** for low $HI_{COVAR}$, and **b** and **b'** for high $HI_{COVAR}$. Thus, the difference detected by the polygenic hazard index was reduced from 4.6 to 2.6 years in the combined model (3), which we nonetheless consider to be appreciable. Lastly, since it can be argued that it is more relevant to conduct this comparison with $HI_{COVAR}$ based on model (2) (i.e., prior to adjustment for the polygenic scores), we performed a robustness check with that combination of hazard indexes and found basically identical results (Table 3D). In summary, once the polygenic scores were adjusted for the underwriting factors in our insurance scenario, they could together detect a 2.6-year shorter median lifespan in the top decile of total genetic liability.

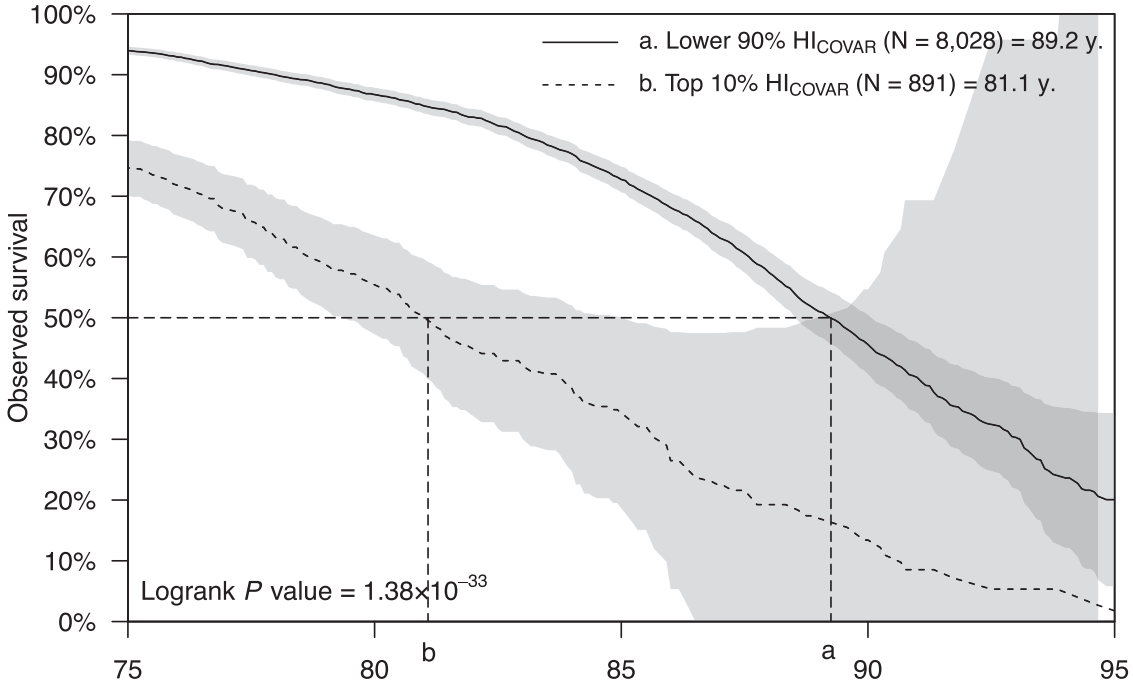


**Fig. 1. Mortality risk classification with hazard indexes.** Panel (**A**) displays Kaplan-Meier survival functions stratified by the lower nine (low) versus the top (high) decile of the hazard index for the demographic and medical underwriting factors ($HI_{COVAR}$), based on the coefficients of the covariate-first model (2). The horizontal dashed line at 50% shows the median lifespan in each group, and the distance between **a** and **b** shows a difference of 8.1 years between low and high $HI_{COVAR}$ (Table 3B). Panel (**B**) displays Kaplan-Meier survival functions stratified by the lower nine (low) and top (high) deciles of both $HI_{COVAR}$ and the polygenic hazard index ($HI_{PGS}$), based on the combined model (3). Here, $HI_{COVAR}$ detected the same 8.1-year difference (the distance between **a** and **b**, or **a'** and **b'**), while the distance between **a** and **a'** (or **b** and **b'**) shows the 2.6-year shorter median lifespan detected by high $HI_{PGS}$, independent of $HI_{COVAR}$ (Table 3C). See Table 3 for omitted confidence intervals.

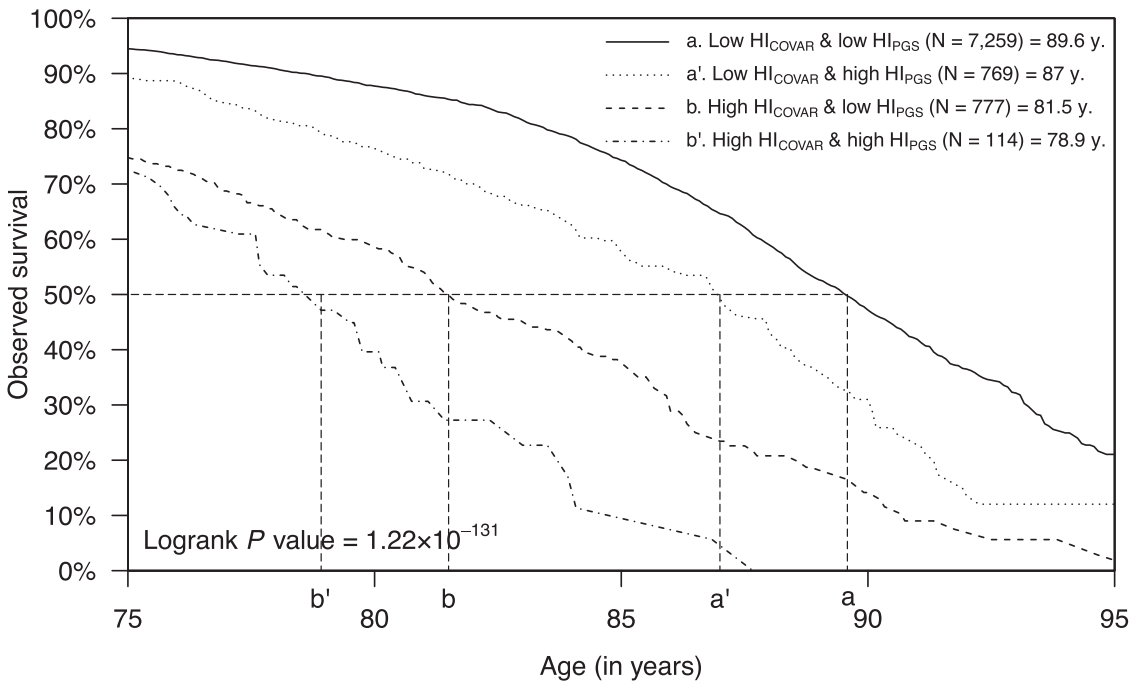


**Fig. 1.** Continued

*6.2.1. Cumulative mortality rates and comparison with selected underwriting factors*

We proceeded by computing empirical 10-year cumulative mortality rates for a life turning 60, 70, or 80. The $t$-year cumulative mortality rate is defined as $1 - \prod_{i}^{i+(t-1)} p_i$, where $p_i$ is the survival rate of respondents in a given group between that start of age $i$ and $i$+1 (Brackenridge et al., 2006). Relative to the average of the genotyped sample, having high $HI_{PGS}$ was associated with ~35–67% higher relative mortality, and high $HI_{COVAR}$ with an extra ~52–185% (**Supplementary Table 6**). We did not compute mortality rates for the 114 respondents that were high in both indexes as that would require interpolation.

As a final benchmark, we stratified empirical survival functions by each of the following four underwriting factors considered on their own: current or former smoking; top decile of BMI; top decile of years of schooling; and sex. We found that the 2.6-year difference in median lifespan detected by $HI_{PGS}$ in Section 6.2 was comparable to that of former smoking (2.5 y), slightly smaller than that of sex (2.8 y), and about twice that of being in the top decile of years of schooling (1.3 y) (**Supplementary Table 7** and **Supplementary Fig. 6**). However, the difference detected by $HI_{PGS}$ was only about half compared to the top decile of BMI (4.4 y), and about a quarter of being a current smoker (9.9 y). Thus, when already-existing polygenic scores are considered together, their capacity to classify mortality risk is by now on par with some established underwriting factors, like sex or former smoking.

*6.3. Cross-validation and investigation of sample attrition*

To maximize sample size, we evaluated the polygenic hazard index in the same sample as was used to estimate the Cox coefficients. To investigate for overfitting, we performed a cross-validation with 1000 iterations for the combined model (3). We trained the Cox coefficients in a random sample of 65% of the households, while holding out the remaining households as a validation sample. Some initial decrease in performance is expected due to the reduced size of the training sample. We evaluated the so-called "calibration slope" by testing whether the coefficient of the newly trained $HI_{PGS}$ is less than unity (i.e., worse calibration) when fitted in the validation sample, by taking the median $Z$ statistic of this test across the iterations.

In the cross-validation, the median $Z$ statistic of the test of the calibration slope was marginally significantly less than unity ($P = 0.013$), suggesting some attenuation in performance. However, we believe this relatively weak attenuation is mostly explained by the reduction in training sample size rather than overfitting. Reassuringly, the difference in median lifespan detected by high $HI_{PGS}$ was similar in the cross-validation, now 2.3 instead of 2.6 years. Thus, the cross-validation exercise suggests that our main findings are largely robust to overfitting.

As another robustness check, we investigated non-death sample attrition with logistic regression. In total, 140 respondents (~1.5%) withdrew from the genotyped sample during the follow-up period. Two polygenic scores were found associated with attrition at the conventional level ($P < 0.05$): breast cancer and LDL cholesterol. Reassuringly, neither of these two polygenic scores were among the most strongly associated scores in the Cox models, and none of the most strongly associated scores were significant in this robustness check. Only a single underwriting factor was significantly associated with attrition: poor childhood health. The small number of withdrawn respondents and the absence of associations with the best-performing polygenic scores suggest that the results are mostly robust to non-death attrition.

## 7. Discussion

We investigated whether a battery of polygenic scores for common medical conditions and mortality risks could improve mortality risk classification in life insurance, and then compared their performance in this regard to that of conventional underwriting factors. As a relevant insurance scenario, covariates were selected to reflect a personal health declaration. Our main finding is that the polygenic scores could together detect a 2.6-year difference in median lifespan, independent of the covariates, which is substantial but less than the 8.1 years detected by the demographic and medical underwriting factors. However, an obvious advantage of predictive genetic testing is that genetic risks could be revealed much earlier in life than the manifestation of many underwriting factors. When the performance of the polygenic scores was compared to each of some established factors on their own, we found that it was similar to that of sex or former smoking, but less than obesity or current smoking. We emphasize that our estimates represent the present lower bound and that polygenic score accuracy will increase considerably in the near future as larger GWAS become available.

By estimating cumulative mortality rates, we found that respondents in the top decile of total genetic liability have an extra relative mortality rate of 35–67%, depending on the age group. This extra mortality is hovering around the threshold that is commonly used to define the first substandard risk class (i.e., typically +25–75%). Therefore, in theory, insurers could already have sufficient motivation to rate an applicant that is otherwise of average risk as substandard based on high polygenic load. Our estimates serve as a proof-of-concept that polygenic scores will at some point be predictive enough to be considered for integration in life underwriting, while much research is needed to determine exactly when polygenic scores for specific diseases or disorders can be considered actuarially sound.

Our results can contribute to theoretical studies on adverse selection or consumer welfare by providing a reasonable parameter space for the population distribution of risk classes. This information could help settle the disagreement in the economic literature on whether or not predictive genetic testing constitute a major problem for adverse selection. For example, with the results presented here, a future study could model the likelihood of adverse selection for a predictive genetic tests that can discern up 2.6-year difference in lifespan for a risk type affecting 10 percent of the population. This predictive capacity is maybe not enough for insurance market failure. Further, an interesting topic beyond the scope of this study is to determine whether relatively inexpensive genetic tests may eventually be able to replace more costly medical examination procedures, which could reduce the market price of life insurance. For

example, we speculate that predictive genetic tests may eventually be able to replace certain laboratory analyses, e.g., of blood lipids, or could help prioritize at-risk applicants that are more worthwhile for insurers to subject to more expensive examination procedures.

It has been argued that genetic information bans could eventually threaten the affordability and viability of private life insurance markets (Rothstein, 2004). In the circumstance that existing bans on insurer's use of genetic information remain in place, our findings lend some support to the notion that adverse selection could intensify as more people acquire private genetic information. Therefore, we agree with Hoy and Ruse (2005) and others who have argued that existing bans must be evaluated periodically and that genetic information in life insurance underwriting may eventually have to be made permissible. Also, life insurance is often considered less essential than health insurance, which is a reason why genetic information may be deemed more permissible for this insurance type (Rothstein, 2004).

At the same time, there is a real risk of genetic discrimination and that coverage may be denied to those who received a burdening outcome in the "genetic lottery", and, as such, are in greater need of solidarity (Harden, 2021). Concerns have been raised about disincentivizing genetic testing for medical or research purposes (Keogh et al., 2017; Rothstein, 2018). Thus, strong arguments can be made that actuarial discrimination of any kind based on genetic factors should be restricted (Rothstein, 2018; Newson et al., 2018). Additionally, the heritability of common diseases is not necessarily deterministic, and some genetic effects act via non-biological pathways, such as lifestyle choices, that are yet uncharted or that may be modifiable. This caveat is particularly important to determine whether it is justifiable to rate, say, never smokers based on their polygenic score for cigarettes per day. These complications question the general ethicality of classifying mortality risk based on predictive genetic tests. We encourage policymakers, the insurance industry, and stakeholders to monitor this development closely and to have a scientifically informed discussion about how to best balance competing interests among stakeholders.

The study should be considered in light of a few limitations. First, finite GWAS *N*, and likely also mortality selection, has underestimated the true performance of polygenic scores. Mortality selection should mostly have affected scores for conditions that manifest at middle age, which could explain why, in contrast to our expectations, we did not find an association with the score for coronary artery disease (a condition of great relevance to insurers).

A second limitation is that we only studied individuals of European ancestry. Thus, we do not know whether our findings generalize to other ancestries and it will take many years before there is enough data to thoroughly answer this question. Lastly, we acknowledge that the medical conditions ascertained in the HRS are based on self-reports and lack specificity, and it could be that their true impact on survival is understated in our analyses. Future studies with access to richer medical data will have to determine more precisely the amount of information polygenic scores can add in addition to acquired medical conditions or other disease markers.

Our results bear on the ongoing debate on whether regulation is needed to control the commercialization of predictive genetic testing. Many companies market these tests to the public while providing limited guidance and lengthy disclaimers, which could be considered a questionable practice. In particular, overweighting the importance of genetic tests with low accuracy could be dangerous. For example, it could be financially detrimental for the beneficiaries of a life insurance policy if the policy holder terminates the policy due to false belief of low genetic risk based on an inaccurate genetic test. Therefore, we agree with other experts who call for more consumer protection in the consumer genetics market (Schleit et al., 2019). Overall, more research is needed to determine how vulnerable types of consumers, for example, those with particularly high genetic risk or those with weak genetic literacy, react after exploring their DNA for health information. For now, we think that commercialization of genetic tests that predict disease or mortality risk is ethically problematic.

Therefore, as a policy recommendation, we encourage policymakers to consider predictive genetic testing of medical conditions to be a form of genetic counseling. As such, we would consider it reasonable to limit the practice to licensed or accredited institutions, be they public or private. However, international standards for genetic counseling are still maturing globally, and until regulatory measures are taken and an industry standard has been established, it is likely that appropriate consumer protection will lag behind the rapid technological developments in this area. At the same time, since many people appear eager to purchase genetic health information, we consider it undesirable and practically impossible to restrict an individual persons' right to explore their own DNA, with or without the assistance of a certified counselor. Additionally, the borderless nature of genetic testing, where consumers can send their genetic data to services located in different jurisdictions, makes it practically impossible to effectively regulate this market only at the national level. Thus, more research is needed to develop appropriate policies to regulate consumer genetics, and such regulation may require international agreements to be effective.

In conclusion, the present study has demonstrated the relevance of polygenic scores in the context of mortality risk classification in life insurance. However, more research is required to determine what specific polygenic scores could be deemed actuarially justified and to develop suitable risk classes for insurance underwriting. Ultimately, governments and policymakers will have to strike a difficult balance between adverse selection and genetic discrimination in order to keep private life insurance fair and viable, while ensuring adequate consumer protection and access to coverage. In the meantime, as long as insurers remain prohibited from using predictive genetic tests for insurance decisions, we see a tangible risk of growing information asymmetries that could lead to adverse selection.

## Authorship statement

Manuscript title: Genetic risk scores in life insurance underwriting

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, writing, or revision of the

manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in *the Journal of Health Economics*.

## Authorship contributions

This study was conceived by R.K.L., and designed by R.K.L with critical input by P.D.K. P.D.K oversaw and supervised the study. R.K.L performed the analyses. R.K.L. and P.D.K interpreted the results. R.K.L drafted and revised the manuscript with critical input from P.D.K.

## Acknowledgements

All persons who have made substantial contributions to the work are reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements.

## This statement is signed by all the authors

Dr. Richard Karlsson Linnér*

Prof. Dr. Philipp Koellinger

Vrije Universiteit Amsterdam
School of Business and Economics
Department of Economics
De Boelelaan 1105
1081HV Amsterdam
The Netherlands
* Correspondence should be addressed to Dr. Karlsson Linnér: r.karlssonlinner@vu.nl or r.karlsson.linner@law.leidenuniv.nl

## CRediT authorship contribution statement

**Richard Karlsson Linnér:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Philipp D. Koellinger:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgments

The research was approved by the Research Ethics Review Board (RERB) at the School of Business and Economics (SBE) of Vrije Universiteit Amsterdam (20180927.1.pkr730). We gratefully acknowledge Netspar for financially supporting this research project with a topicality grant (BD2018.01) and thank Bas Werker and Anja de Waegenaere for their valuable feedback on the research questions, study design, and draft manuscript. The study was financially supported by an ERC Consolidator Grant to Philipp Koellinger (647648 EdGe). This work was carried out on the Dutch national e-infrastructure (grant 17148) with support of the SURF Cooperative. We are thankful to Casper Burik for constructing the genetic PCs and to Aysu Okbay for a providing a meta-analysis of educational attainment that excluded the Health and Retirement Study (HRS). We thank Jonathan Beauchamp for valuable input on the draft manuscript, and the Editor and anonymous referees for their helpful and constructive feedback.

HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University in Baltimore, Maryland. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington in Seattle, Washington. Genotype data can be accessed via the database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap, accession number phs000428.v2.p2). Researchers who wish to link genetic data with other HRS measures that are not in dbGaP, such as educational attainment, must apply to the HRS for permission. See the HRS website (http://hrsonline.isr.umich.edu/gwas) for details.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jhealeco.2021.102556.

## References

Balia, S., Jones, A.M., 2008. Mortality, lifestyle and socio-economic status. J. Health Econ. 27 (1), 1–26. doi:10.1016/j.jhealeco.2007.03.001.

Bardey, D., De Donder, P., Mantilla, C., 2019. How is the trade-off between adverse selection and discrimination risk affected by genetic testing? Theory and experiment. J. Health Econ. 68, 102223. doi:10.1016/j.jhealeco.2019.102223.

Barth, D., Papageorge, N.W., Thom, K., 2020. Genetic endowments and wealth inequality. J. Polit. Econ. 128 (4), 1474–1522. doi:10.1086/705415.
Bélisle-Pipon, J.-C., Vayena, E., Green, R.C., Glenn Cohen, I., 2019. Genetic testing, insurance discrimination and medical research: what the United States can learn from peer countries. Nat. Med. 25 (8), 1198–1204. doi:10.1038/s41591-019-0534-z.
Brackenridge, R.D.C., Croxson, R.S., Mackenzie, R., 2006. Medical Selection of Life Risks, 5th ed. Palgrave Macmillan, New York Edited by Brackenridge, R.D.C., Croxson, R.S., and Mackenzie, R. doi:10.1007/978-1-349-56632-7.
Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. Ser. B (Methodol.) 34 (2), 187–220. http://www.jstor.org/stable/2985181.
Cummins, J.D., Smith, B.D., Neil Vance, R., VanDerhei, J.L., 1983. Risk Classification in Life Insurance, 1st ed. Springer Science+Business Media, New York Edited by Cummins, J.D..
Dionne, G., Rothschild, C., 2014. Economic effects of risk classification bans. Geneva Risk Insur. Rev. 39 (2), 184–221. https://www.jstor.org/stable/24712688.
Doherty, N.A., Thistle, P.D., 1996. Adverse selection with endogenous information in insurance markets. J. Public Econ. 63 (1), 83–102. doi:10.1016/0047-2727(95)01568-X.
Domingue, B.W., Belsky, D.W., Harrati, A., Conley, D., Weir, D.R., Boardman, J.D., 2017. Mortality selection in a genetic sample and implications for association studies. Int. J. Epidemiol. 46 (4), 1285–1294. doi:10.1093/ije/dyx041.
Fiorentino, A., 2021. New genetic test for heart attack risk launched for patients at mass general. Broad Institute 2021. https://www.broadinstitute.org/news/new-genetic-test-heart-attack-risk-launched-patients-mass-general.
Ganna, A., Rivadeneira, F., Hofman, A., Uitterlinden, AG., Magnusson, PK.E., Pedersen, NL., Ingelsson, E., Tiemeier, H., 2013. Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality? Hum. Genet. 132 (5), 553–561. doi:10.1007/s00439-013-1267-6.
Ge, T., Chen, C-Y., Ni, Y., Feng, Y-C.A., Smoller, J.W., 2019. Polygenic prediction via bayesian regression and continuous shrinkage priors. Nat. Commun. 10 (1), 1776. doi:10.1038/s41467-019-09718-5.
Gerdtham, U-G., Johannesson, M., 2004. Absolute income, relative income, income inequality, and mortality. J. Hum. Resour. 39 (1), 228–247. doi:10.2307/3559011. http://10.0.9.3/3559011.
Harden, K.P., 2021. Genetic Lottery: Why DNA Matters for Social Equality. Princeton University Press, Princeton.
Hoel, M., Iversen, T., Nilssen, T., Vislie, J., 2006. Genetic testing in competitive insurance markets with repulsion from chance: a welfare analysis. J. Health Econ. 25 (5), 847–860. doi:10.1016/j.jhealeco.2005.12.003.
Hoy, M., Lambert, P., 2000. Genetic screening and price discrimination in insurance markets. Geneva Pap. Risk Insur. Theory 25 (2), 103–130. https://www.jstor.org/stable/41953388.
Hoy, M., Polborn, M., 2000. The value of genetic information in the life insurance market. J. Public Econ. 78 (3), 235–252. doi:10.1016/S0047-2727(99)00118-8.
Hoy, M., Ruse, M., 2005. Regulating genetic information in insurance markets. Risk Manag. Insur. Rev. 8 (2), 211–237. doi:10.1111/j.1540-6296.2005.00057.x.
Hoy, M., Witt, J., 2007. Welfare effects of banning genetic information in the life insurance market: the case of BRCA1/2 genes. J. Risk Insur. 74 (3), 523–546. doi:10.1111/j.1539-6975.2007.00223.x, 10.1111/j.1539-6975.2007.00223.x.
Jan, S., Laba, T-L., Essue, B.M., Gheorghe, A., Muhunthan, J., Engelgau, M., Mahal, A., et al., 2018. Action to address the household economic burden of non-communicable diseases. Lancet North Am. Ed. 391 (10134), 2047–2058. doi:10.1016/S0140-6736(18)30323-4.
Joly, Y., Burton, H., Knoppers, B.M., Feze, I.N., Dent, T., Pashayan, N., Chowdhury, S., et al., 2014. Life Insurance: genomic stratification and risk classification. Eur. J. Hum. Genet. 22 (5), 575–579. doi:10.1038/ejhg.2013.228.
Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Am. Statist. Assoc. 53 (282), 457–481. doi:10.2307/2281868.
Keogh, L., Niven, H., Rutstein, A., Flander, L., Gaff, C., Jenkins, M., 2017. Choosing not to undergo predictive genetic testing for hereditary colorectal cancer syndromes: expanding our understanding of decliners and declining. J. Behav. Med. 40 (4), 583–594. http://10.0.3.239/s10865-016-9820-0.
Lambert, S.A., Abraham, G., Inouye, M., 2019. Towards clinical utility of polygenic risk scores. Hum. Mol. Genet. 28 (R2), R133–R142. doi:10.1093/hmg/ddz187.
Lefebre, J.J., Salgaonkar, S., Willwerth-Pascutiu, G., Zimmerman, D.D., 2019. Genetics and insurance: challenges and opportunities II. RGA Reinsurance Company, Chesterfield.
Lin, D.Y., Wei, L.J., 1989. The robust inference for the cox proportional hazards model. J. Am. Statist. Assoc. 84 (408), 1074–1078. doi:10.2307/2290085.
Macdonald, A., Tapadar, P., 2010. Multifactorial genetic disorders and adverse selection: epidemiology meets economics. J. Risk Insur. 77 (1), 155–182. doi:10.0487/j.1539-6975.2009.01342.x.
Marioni, R. E., Ritchie, S.J., Joshi, P.K., Hagenaars, S.P., Okbay, A., Fischer, K., Adams, M.J., et al., 2016. Genetic variants linked to education predict longevity. Proc. Natl. Acad. Sci. U.S.A. 113 (47), 13366–13371. doi:10.1073/pnas.1605334113.
Maxwell, J.M., Russell, R.A., Wu, H.M., Sharapova, N., Banthorpe, P., O'Reilly, P.F., Lewis, C.M., 2020. Multifactorial disorders and polygenic risk scores: predicting common diseases and the possibility of adverse selection in life and protection insurance. Ann. Actuar. Sci. 1–16. doi:10.1017/S1748499520000226.
McCrea, M., Farrell, M., 2018. A conceptual model for pricing health and life insurance using wearable technology. Risk Manag. Insur. Rev. 21 (3), 389–411. doi:10.1111/rmir.12112.
McDaid, A.F., Joshi, P.K., Porcu, E., Komljenovic, A., Li, H., Sorrentino, V., Litovchenko, M., et al., 2017. Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. Nat. Commun. 8 (July), 15842. doi:10.1038/ncomms15842.
Meyricke, R., Sherris, M., 2013. The determinants of mortality heterogeneity and implications for pricing annuities. Insur. Math. Econ. 53 (2), 379–387. doi:10.1016/j.insmatheco.2013.06.002.
Nabholz, C., Somerville, K., 2011. Fair Risk Assessment in Life & Health Insurance. Swiss Re, Zurich.
Newson, A.J., Ayres, S., Boyle, J., Gabbett, M.T., Nisselle, A., 2018. Human genetics society of Australasia position statement: genetic testing and personal insurance products in Australia. Twin Res. Hum. Genet. 21 (6), 533–537. doi:10.1017/thg.2018.60.
Palme, M., Sandgren, S., 2008. Parental income, lifetime income, and mortality. J. Eur. Econ. Assoc. 6 (4), 890–911. https://www.jstor.org/stable/40282688.
Peter, R., Richter, A., Thistle, P., 2017. Endogenous information, adverse selection, and prevention: implications for genetic testing policy. J. Health Econ. 55, 95–107. doi:10.1016/j.jhealeco.2017.06.010.
Pilling, L.C., Kuo, C.L., Sicinski, K., Tamosauskaite, J., Kuchel, G.A., Harries, L.W., Herd, P., Wallace, R., Ferrucci, L., Melzer, D., 2017. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging 9 (12), 2504–2520. doi:10.18632/aging.101334.
Pingault, J-.B., O'Reilly, P.F., Schoeler, T., Ploubidis, G.B., Rijsdijk, F., Dudbridge, F., 2018. Using genetic data to strengthen causal inference in observational research. Nat. Rev. Genet. 19, 566–580. doi:10.1038/s41576-018-0020-3.
Polborn, M.K., Hoy, M., Sadanand, A., 2006. Advantageous effects of regulatory adverse selection in the life insurance market. Econ. J. 116 (508), 327–354. doi:10.1111/j.1468-0297.2006.01059.x. http://www.jstor.org.vu-nl.idm.oclc.org/stable/3590347.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38 (8), 904–909. doi:10.1038/ng1847.
Prince, A.E.R., 2019. Comparative perspectives: regulating insurer use of genetic information. Eur. J. Hum. Genet. 27 (3), 340–348. doi:10.1038/s41431-018-0293-1.
Rechfeld, F., Rischatsch M., Nemoto T., and Nabholz C.. 2019. "Can Life Insurance Pass the Genetic Test?" Zurich: Swiss Re.
Rothstein, M., 2004. Genetics and Life Insurance: Medical Underwriting and Social Policy. MIT Press, Cambridge NV-1 online.
Rothstein, MA., 2018. Time to end the use of genetic test results in life insurance underwriting. J. Law Med. Ethics 46 (3), 794–801. doi:10.1177/1073110518804243.
Royston, P., Altman, D.G., 2013. External validation of a cox prognostic model: principles and methods. BMC Med. Res. Method. 13 (March), 33. doi:10.1186/1471-2288-13-33.
Royston, P., Sauerbrei, W., 2004. A new measure of prognostic separation in survival data. Stat. Med. 23 (5), 723–748. doi:10.1002/sim.1621.
Sari, D.J., Lestari, D., Devila, S., 2019. Pricing life insurance premiums using Cox regression model. AIP Conf. Proc. (November) 2168. doi:10.1063/1.5132461.
Schleit, J., Naylor, L.V., Hisama, F.M., 2019. First, do no harm: direct-to-consumer genetic testing. Genet. Med. 21 (2), 510–511. doi:10.1038/s41436-018-0071-z.
Sijbrands, E.J.G., Tornij, E., Homsma, S.J., 2009. Mortality risk prediction by an insurance company and long-term follow-up of 62,000 Men. PLoS One 4 (5), e5457. doi:10.1371/journal.pone.0005457, –e5457.

Sonnega, A., Faul, J.D., Ofstedal, M.B., Langa, K.M., Phillips, J.W.R., Weir, D.R., 2014. Cohort profile: the health and retirement study (HRS). Int. J. Epidemiol. 43 (2), 576–585. doi:10.1093/ije/dyu067.

Tabarrok, A., 1994. Genetic testing: an economic and contractarian analysis. J. Health Econ. 13 (1), 75–91. doi:10.1016/0167-6296(94)90005-1.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. doi:10.1038/s41576-019-0127-1.

Timmers, P.R.H.J., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., Bretherick, A.D., et al., 2019. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. ELife 8 (January), e39856. doi:10.7554/eLife.39856.

Torkamani, A., Wineinger, N.E., Topol, E.J., 2018. The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. 19 (September), 581–590. doi:10.1038/s41576-018-0018-x.

Visscher, P.M., Yengo, L., Cox, N.J., Wray, N.R., 2021. Discovery and implications of polygenicity of common diseases. Science 373 (6562), 1468–1473. doi:10.1126/science.abi8206.

Weir, D.R., 2016. Validating Mortality Ascertainment in the Health and Retirement Study. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan.

Widmer, U., 2019. Underwriting Insights Predicting Common Disease with Polygenic Risk Scores Could It Disrupt Life Insurance?. Swiss Re, Zurich.

Zeggini, E., Gloyn, A.L., Barton, A.C., Wain, L.V., 2019. Translational genomics and precision medicine: moving from the lab to the clinic. Science 365 (6460), 1409–1413. doi:10.1126/science.aax4588, LP –1413.