scientific reports

Check for updates

OPEN

Polygenic Health Index, General Health, and Pleiotropy: Sibling Analysis and Disease Risk Reduction

Erik Widen[1,2]✉, Louis Lello[1,2]✉, Timothy G. Raben[1], Laurent C. A. M. Tellier[1,2] & Stephen D. H. Hsu[1,2]

We construct a polygenic health index as a weighted sum of polygenic risk scores for 20 major disease conditions, including, e.g., coronary artery disease, type 1 and 2 diabetes, schizophrenia, etc. Individual weights are determined by population-level estimates of impact on life expectancy. We validate this index in odds ratios and selection experiments using unrelated individuals and siblings (pairs and trios) from the UK Biobank. Individuals with higher index scores have decreased disease risk across almost all 20 diseases (no significant risk increases), and longer calculated life expectancy. When estimated Disability Adjusted Life Years (DALYs) are used as the performance metric, the gain from selection among ten individuals (highest index score vs average) is found to be roughly 4 DALYs. We find no statistical evidence for antagonistic trade-offs in risk reduction across these diseases. Correlations between genetic disease risks are found to be mostly positive and generally mild. These results have important implications for public health and also for fundamental issues such as pleiotropy and genetic architecture of human disease conditions.

Interest in polygenic risk scores (PRS) and the ability to estimate disease risks from genotypes has increased steadily over the past decade. A polygenic risk score maps an individual genotype to a score that reflects genetic risk for a particular disease; most PRS depend on hundreds or thousands of individual loci in the genome. As biobank data sets have grown larger, so have the performances and applicability of PRS. There are now a multitude of predictors that can assign estimated disease risks with an accuracy that has reached clinical utility. Disease conditions as diverse as coronary artery disease, breast cancer, and schizophrenia can be predicted with a useful accuracy from genetic information alone[1–21]. Typically, PRS are trained on and applied to a single disease but with many such risk predictions available it is natural to ask whether they could be combined into a general health index—a single number to describe the overall health of an individual. This question has already been explored in[22], where the authors created a composite PRS using a cox-hazard model, utilizing diseased participants of the UK Biobank (UKB). This composite PRS was found to predict longevity. The impact on longevity and individual disease burdens from individual variants has also been studied, using the Finish databank FinnGen[23].

In this paper, we construct a form of general health index by combining PRS for 20 diseases (Table 1), choosing the individual disease weights in an attempt to minimize the number of life years lost due to illness. The choice of conditions to include in the index was partly idiosyncratic—determined by the set of well-performing PRS available, prioritized by overall burden (life expectancy impact times population prevalence). The list is not exhaustive and future extensions of this work are planned. We evaluate whether a single number index score is a useful reflection of an individual's various disease risks and their combined effect on estimated life years. If true, health indices could be a valuable tool for clinicians and patients to assess combined risks and genetic health predisposition. For a wide range of reasons, interpreting clinical risk based on genetic data can be difficult for both patients[24–29] and clinicians[30–32]. Combining PRS into a single metric can greatly simplify the process of evaluating genetic risk reports.

Another prominent application of a general health index is to inform embryo selection in IVF cycles (in vitro fertilization). Embryos are routinely biopsied for aneuploidy and monogenetic disease tests. For cycles resulting in more than one euploid embryo (without any of the monogenetic disease variants), clinicians and prospective

[1]Department of Physics and Astronomy, Michigan State University, 567 Wilson Rd, East Lansing, MI 48824, USA. [2]Genomic Prediction, Inc., 671 US Highway One, North Brunswick, NJ 08902, USA. ✉email: wideneri@msu.edu; lellolou@msu.edu

Abbr.	Disease	AUC	Abbr.	Disease	AUC	Abbr.	Disease	AUC
AD	Alzheimer's disease	$0.686_{(0.004)}$	HA	Heart attack	$0.580_{(0.008)}$	Obes	Obesity	$0.669_{(0.002)}$
AFib	Atrial fibrillation	$0.623_{(0.004)}$	HCL	Hypercholesterolemia	$0.616_{(0.003)}$	PC	Prostate cancer	$0.64_{(0.02)}$
ASA	Asthma	$0.626_{(0.004)}$	HTN	Hypertension	$0.635_{(0.003)}$	SCZ	Schizophrenia	$0.67_{(0.03)}$
BC	Breast Cancer	$0.594_{(0.008)}$	IBD	Inflammatory bowel disease	$0.647_{(0.003)}$	T1D	Type I diabetes	$0.63_{(0.02)}$
BCC	Basal cell carcinoma	$0.62_{(0.01)}$	IS	Ischemic stroke	$0.541_{(0.002)}$	T2D	Type II diabetes	$0.616_{(0.004)}$
CAD	Coronary artery disease	$0.616_{(0.005)}$	MDD	Major depressive disorder	$0.534_{(0.001)}$	TC	Testicular cancer	$0.61_{(0.04)}$
Gout	Gout	$0.65_{(0.01)}$	MM	Malignant melanoma	$0.57_{(0.02)}$			

Table 1. Disease abbreviations and predictor AUCs. AUCs are listed with the standard deviations in parentheses, as computed by 30 bootstrap runs (see Supplementary Information for more details and case/control counts).

parents typically select which embryo to implant based on visually assigned embryo grades. With the advent of preimplantation polygenetic testing, a general health index could *additionally* be used to guide this choice and reduce the overall disease risk for the baby.

A priori, it is not given that such a health index would be useful. A common preliminary objection is that an index or single PRS, while reducing the risk for one disease, could inadvertently increase the risk for another[33,34]. However, it has long been known that several pairs of diseases, often grouped into categories, in reality tend to co-occur[35–46]. It seems possible that there are genetically influenced large systems (circulatory, digestive, metabolic, etc.) that vary across individuals in robustness, and which affect disease risks across multiple conditions. This could, at least for some broad categories of diseases, allow for useful indices. The specific concern raised for polygenic health indices has been the possibility of antagonistic pleiotropy, i.e., that a single gene may affect more than one disease risk simultaneously and in such a way that it decreases one disease risk while increasing another. If such pleiotropy were very common, there would not be much point of a genetically based health index.

In this paper, we examine both underlying phenotypic comorbidities and genetic pleiotropy to answer whether the notion of genetic general health can be meaningful and—if so—if the proposed health index is indicative of health outcomes and can be used to reduce several disease risks without risking significant trade-offs. We find that the 20 studied diseases frequently occur together, sometimes with strong positive phenotypic correlation, while the genetic pleiotropy is usually small and slightly *positive*, or negligible. More importantly, we show in practice, using real genetic and health data, that the proposed health index can identify individuals at high or low risk for almost all the 20 diseases simultaneously. We observed individual disease risk reductions even beyond 40% (CAD, heart attack, diabetes type II) when selecting the highest index among five individuals, as compared to the general population. We further see no statistically significant evidence for inadvertent risk increments among any of the 20 diseases, nor among any of 11 additionally analyzed common diseases that did not have predictors included in the index.

These conclusions are drawn from several experiments. We apply the constructed index to about 40,000 late-life individuals of European ancestry for whom both genotypes and medical history are known, using the UK biobank (UKB). Odds (prevalence) plots are shown for the most common diseases but the majority of the results are in form of selection experiments. The test data samples are grouped, using different group sizes in different experiments, and the sample with the highest health index is selected from each group. The selected individuals are then compared to the total test set to see the health differences in the medical history data, computing metrics like Relative Risk Reduction (RRR) and estimated gained life years. These experiments are repeated and confirmed with a very strong test of the genetic signal: selection among pairs (21,539) and trios (969) of genetic siblings. Siblings have both less genetic variation and typically share similar family environments, thus constituting an excellent test set. Finally, the underlying phenotypic and PRS dependencies among the 20 diseases in the index are analyzed, as well as the index relations (t-tests and correlations) to 11 common diseases not in the index, 5 addiction phenotypes, and 5 continuous phenotypes.

It is well-established that PRS are more accurate within a population ancestrally homogeneous and similar to the training population—–however, generally a positive effect in one ancestry will persist in more distant ancestries. Research on this topic is ongoing and of high interest[7,47–50]. The primary motivation for this paper is to investigate whether a composite genetic health index is reflective of general health in principle and we therefore focused on a single ancestry with maximum amount of data.

Only the listed 20 diseases in the index, and an additional 11 conditions, were analyzed in this paper. Although studies of general health will never exhaust the list of everything that may be relevant, it is important to stress the limited scope of this first analysis of the genetic health index. There are many diseases with significant mortalities and disability burdens whose impact and dependencies on the index are not taken into account in this presentation . Also, non-pathological traits, such as grip strength, reaction time, and cognitive metrics etc., may correlate with the index. This paper only examined five such phenotypes. Follow-up studies expanding the scope of the analysis—both in terms of more diseases and other traits—are already ongoing. For this publication however, we emphasize again that the results presented refer to general health in terms of the listed 20 diseases only, and when indicated the additional 11 conditions.

All analyses, except where otherwise specified, are performed on self-reported white samples from the full UKB release (2021-04); these are almost exclusively of European ancestry. We set aside 39,913 samples (containing a large number of genetic siblings) as a pure test set, withheld from all predictor training and hyperparameter tuning (see the Supplementary Information for details on the test set). The PRS are constructed through a previously published LASSO-algorithm[7] trained on $\sim$ 200k-400k samples from the training portion of the same UKB data, except for the predictors for AD, IBD, IS, MDD, and SCZ (predictors leveraging other specialized datasets performed better for these traits). More details on the predictors can be found in the Supplementary Information.

Results

Overview of methods.   *Polygenic health index.*   There are many ways to construct a polygenic health index from multiple PRS. Here we investigate the performance of a single linear combination of risk estimates, attempting to reduce lost life years. Let l_d be the estimated reduction in life expectancy for an individual having a disease d as compared to the general population, and let ρ_d be the lifetime risk in the general population of getting the disease. For the predicted risks r_d, we define the health index to be

$$I = \sum_{d \in \mathcal{D}} l_d(\rho_d - r_d),$$
(1)

for a selected set of diseases $\mathcal{D}$ (this paper consistently uses the 20 diseases in Table 1). As such, a higher I should correspond to a healthier individual. As a proxy for ground truth in our test data set, we also define a case/control-based version, I^c, which instead of the risk r_d uses the recorded case/control status c. (Since there is a very large overlap between the case definitions we used for CAD and HA, we choose to exclude HA from the case/control based index I^c. Otherwise HA would practically be double-counted in the performance evaluation.) We use this quantity as measure of the real world outcome value of the index. We note that the majority of our UKB test set is still alive (age $\mu = 70, \sigma = 7$ years) making I^c an imperfect measure of lifetime outcomes and skewed towards diseases with early onset. Still, since the mean age is not more than about one standard deviation (SD) from the average lifespan and the incomplete data masks cases as controls, rather than vice versa, we expect that a health index validated on an I^c using complete data (with perfect lifetime medical records and age of death) would have a better performance than what is measured in the UKB data. (The Supplementary Information contains more characterization of the test data.)

The index parameters l_d and ρ_d were taken from literature studies, using the average values if more than one source was used (see Supplementary Information).

Gaussian risk model.   The health index definition Eq. (1) requires an estimated absolute (lifetime) risk r for each disease, modeled from the PRS as input. Depending on disease and predictor specifics, there are different possible choices for this modeling. A fairly general model, which works very well for sufficiently polygenic PRS (i.e., such that the Central Limit Theorem can be applied), models the PRS as drawn from a sum of two normal distributions with case/control status dependent means (μ_1/μ_0) and joint variance. The PRS probability distribution can then be written as

$$\phi(\text{PRS}) = (1 - \pi)\mathcal{N}(\mu_0, \sigma) + \pi\mathcal{N}(\mu_1, \sigma),$$
(2)

where π is the population prevalence and $\mathcal{N}$ is the normal distribution. This leads to the Gaussian risk model

$$r(\text{PRS}) = \cfrac{1}{1 + \frac{1-\pi}{\pi} \exp\left[\frac{1}{2\sigma^2}\left((\text{PRS} - \mu_1)^2 - (\text{PRS} - \mu_0)^2\right)\right]}.$$
(3)

The case and control variances do not need to be equal in principle (unequal variances can lead to unrealistic behavior in the tails) but in practice tend to be close in value (see Supplementary Information). We use estimates of μ_0, μ_1, and σ based on the PRS in test set controls and cases.

Selection experiment from groups of unrelated individuals.   To evaluate the performance of the health index, we created sets of groups and carried out selection experiments, i.e., we grouped together random individuals in the test set into groups of a specific size and than picked one individual from each group. In index selection experiments, we selected the individual with the highest index value. In PRS selection experiments we selected the individual with the lowest PRS (lowest risk) for a specific disease.

We created 40k random groups from the samples belonging to the intersection of all predictor test sets, such that no sample was used in any type of training nor hyperparameter tuning. Each sample was scored and assigned a raw and a sex-adjusted (see Supplementary Information) health index, as in Eq. (1). For each selection outcome, we calculated the relative risk reduction (RRR) for each individual disease and the index gain as measured in the case/control-based index I^c, as compared to a completely random selection (i.e., the general population statistics):

$$\text{RRR} = \frac{\pi_{\text{rand}} - \pi_{\text{sel}}}{\pi_{\text{rand}}}; \quad \Delta I^c = \frac{1}{N_{\text{group}}} \sum_{g \in \text{groups}} \left(I^c_{g_{\text{sel}}} - \langle I^c \rangle_g\right) \overset{*}{=} \langle I^c \rangle_{\text{sel}} - \langle I^c \rangle.$$
(4)

Here g sums over all N_{group} groups, $I^c_{g_{\text{sel}}}$ is the health index for the selected individual in group g, and $\langle \cdot \rangle$ denotes the sample means, i.e., $\langle I^c \rangle_g$ is the average health index value in group g, $\langle I^c \rangle_{\text{sel}}$ is the average among all selected individuals, and $\langle I^c \rangle$ is the average in the total test set. The index gain ΔI^c can be viewed either as the average
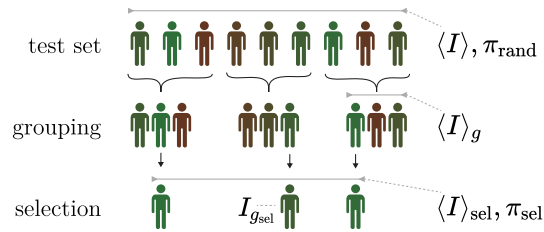
Figure 1. The selection experiments. The test set is scored with the health index I or a single PRS and is randomly divided into groups of equal size. The individual with the best score in each group is selected and the health status among the selected are then compared with the general test set. The symbols in Eq. (4) refer to indicated subsets.

index difference between the selected individual and its group average or as the difference between the average selected index and the general population average (∗ holds for constant group size). Note here that we are using the case/control status based index, I^c, as evaluation metric which does not use any genetic information but only individual lifetime disease status (see Supplementary Information for details), together with the population based lifespan impact and lifetime risk estimates. The full selection experiment procedure is illustrated in Fig. 1.

We repeated all selection experiments 25 times to get a bootstrap estimate of the errors, reusing the same samples but assigning them into different groups. Thus, these are underestimates neglecting the additional variance that would come from also using other samples, while the groupings are practically unique.

For the three sex specific diseases (breast, prostate and testicular cancer), we compared only the subsets with the relevant sex of the selected and random sets when calculating the RRR and index gain.

Genetic sibling selection. The selection experiments on unrelated individuals provide good metrics for how the health index performs in the general population. A much stronger test, that is also more relevant to the application of embryo selection, is to repeat the same experiments using real world siblings, sharing half their genetic material. Accurate prediction within siblings is challenged both by this reduced genetic variance and by more similar environments; it is thus a rigorous test of genetic prediction performance.

We repeated the selection experiments for 21,539 pairs and 969 trios of genetic siblings. Since the sibling data cannot be re-grouped as in the unrelated selection experiments, we opted to not use bootstrap errors but instead calculate the theoretical 95% confidence interval for the prevalence among the selected siblings, based on the Wilson score interval. It was translated to the RRR metric through Eq. (4), keeping the population prevalence π_{rand} fixed. We did not estimate the errors for the index gain metric when selecting among genetic siblings.

Additional diseases and phenotypes. The health index probes 20 diseases directly. Although that corresponds to a sizable subspace of the most common and impactful diseases, it is still far from a complete coverage of "general health". To make an initial probe of diseases and phenotypes not directly included in the index, we examined the genetic health index distributions among cases and controls for 11 additional diseases: bipolar disorder, chronic kidney disease, chronic obstructive pulmonary disease, colorectal cancer, leukemia, lung cancer, lupus, lymphoma, osteoporosis, rheumatoid arthritis, and stomach cancer. In addition, we looked at five self-reported survey questions about addiction history for which we did the same binary trait analysis.

We also examined the correlations between the genetic health index and five continuous phenotypes: lung capacity (forced expiratory volume and forced vital capacity), fluid intelligence, grip strength, and height. Lastly, we performed a linear regression using all the (L2-normalized) additional phenotypes to see whether they were predictive of the health index. Since the health index is systematically different for males and females, we conducted all these additional analyses separately for the two sexes (see the Supplementary Information for a sex neutral version of the index).

Selection experiment using groups of unrelated individuals. We report the overall index gain (ΔI^c from Eq. 4) from the selection experiments on unrelated individuals in Fig. 2. It documents a well-established and consistent gain that increases with group size, maintaining a positive increment even when selecting among more than ten people. The health index distribution is non-Gaussian with standard deviation (SD) of 1.56 estimated life years and with a skewness of -0.49. The difference between the mean health index values for the top and bottom 5% of the index I was 5.10 predicted life years. The corresponding difference between these groups was 3.49 years when measured with the case/control based index I^c (a smaller difference is to be expected due to the incomplete case/control data). Despite different methods and disease sets, we note the connection to[22] which reported similar values in lost life years per SD and difference between top and bottom 5% of composite PRS. In Fig. 3, the selection experiment result at the group size of five is broken down into the RRR and the component-wise index gain for each disease, allowing a more fine-grained view of the performance. Strikingly, the RRR graph is overwhelmingly positive thus demonstrating compelling evidence that selected individuals with higher health index score have lower incidence for almost all diseases at the same time. 15 out of the 20 disease have statistically significant positive RRR, reaching over 40% for the most reduced disease risks (CAD, HA, T2D), whereas none is significantly negative or even has a negative central value. It is important to note that although the weights l_d matter for how the index is constructed and thus for whom is selected, they have no direct impact
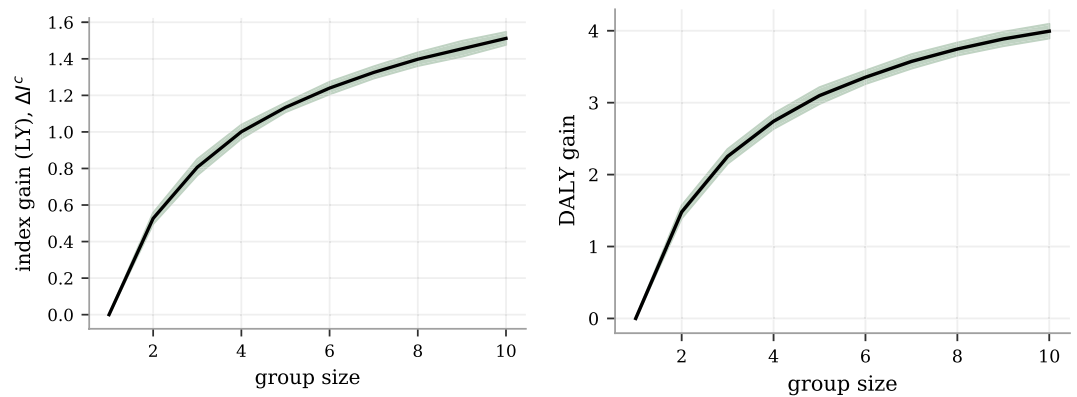
Figure 2. The estimated gain from index selection is a clearly positive function of group size, both using disease weights as defined by population estimates of lost life years or using disease weights based on disabilty-adjusted life years (DALY). Left: the index gain, as measured as the average health index difference between selected and random individuals (ΔI^c in Eq. (4)), is growing monotonically with group size and with a continued clear positive derivative at group sizes of 10. Notably, there is a strongly significant gain for all group sizes, even at a group size of 2. The error band is a 95% CI as computed by 25 experiments with independent selection groupings. Right: while still selecting on the same index Eq. (1), we evaluated it on a case/control status metric using DALY-weights, taking quality of life into account. Again, there is a clear and steady gain, with the gain at a group size at 10 reaching about 4 years. The error band is a 95% CI as computed by 25 experiments with independent selection groupings.
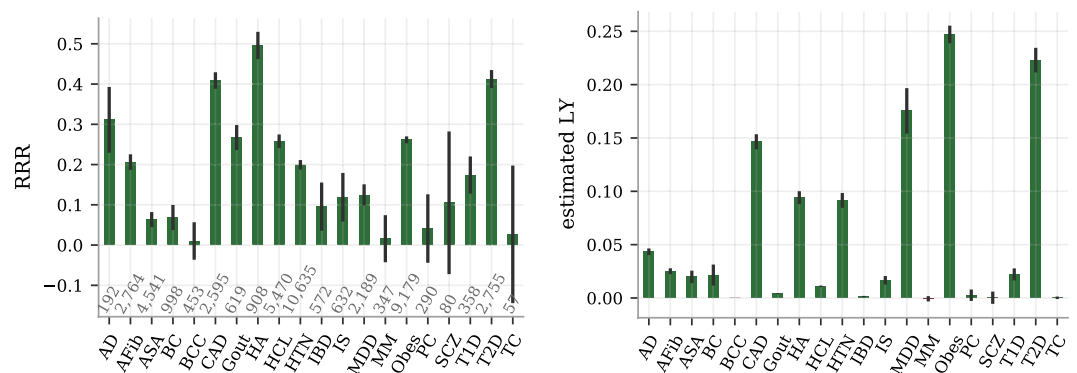
Figure 3. Selecting on health index among five randomly grouped individuals reduces simultaneously the risk of almost all the studied diseases. Left: the RRR among the selected individuals as compared to random selection is dominantly positive, ranging from a few risks reductions statistically consistent with zero up to more than 40%. No disease risk is demonstrably increased. The case numbers for each disease are printed just above the x-axis and the error bars are 95% CI estimates from 25 repeated experiments with different selection groupings. Right: the estimated index gain for each of the index components (diseases), i.e., the disease component breakdown of Eq. (4), also shows non-negative gains across the board with most component gains being statistically significant. The unit on the y-axis is estimated life years (LY), as is the unit of I^c. This index is primarily driven by CAD, heart attack, hypertension, major depressive disorder, obesity and type II diabetes, due to their combinations of strong impacts l_d and high population prevalence.

on the RRR metric itself - only the actual disease status is measured. As such, the RRR plot is a true measurement of the reduced disease incidence. In contrast, the right plot in Fig. 3 of the index gain ΔI^c involves the weights both in selection and in evaluation. Using the weights based on estimated lost life years, we get a disease-by-disease breakdown of the index gain. Again, there is a statistically significant positive contribution from almost all diseases with obesity, type II diabetes, major depressive disorder and CAD as the strongest contributors.

The average component gains in Fig. 3 depend both on the quality of the individual PRS, the weights l_d and the test set prevalences. For example, the AD predictor has a much stronger individual performance than MDD (AUC $\sim$ .69 vs $\sim$ .53) while MDD has stronger weights than AD in the index ($l_{MDD}/l_{AD} \approx 1.6$). The index achieves a RRR of about 31% for AD and 12% for MDD, with the individual PRS-performance having a larger impact on the RRR metric. Meanwhile, MDD has about four times the AD contribution to the index gain, largely due to it being about ten times more prevalent in the test set. Naturally, common diseases contribute more to the average index difference than rare ones. Both AD and MDD have some strong comorbidities and milder
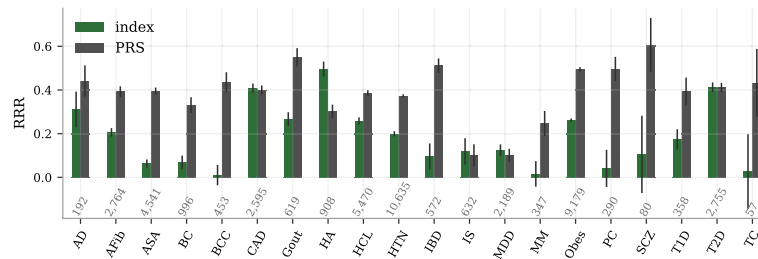
Figure 4. RRR comparison between selection on index and selecting on individual disease PRS. The individual disease RRR obtained by index selection contrasted with selection directly on the individual PRS, using a group size of 5. The case numbers in the test set for each disease are shown above the x-axis and the error bars are 95% CI as computed by 25 independent experiment runs.

PRS-correlations with other diseases; this is discussed further in "Characterization of phenotypic and genetic dependencies" . See also the Supplementary Information for a deeper discussion of the test set prevalences and their influence on the quantitative results.

The RRR and index gain metrics offer complementary information of the potential benefits: the RRR captures how much the risk can be reduced simultaneously, while the index gain translates this into estimates of the corresponding life years gained on average. All selection experiments selected on the index in equation (1), using lost life years l_d as weights. A common alternative for assigning relative importance to diseases is the unit Disability Adjusted Life Years (DALY). While still *selecting on our index* (1), we make contact to the existing DALY-literature by *evaluating the index gain* using a DALY-scale to the right in Fig. 2. The weights in the evaluating index difference ΔI^c were computed as population level DALY-coefficients $l_d + q_d \Delta y_d$, where q_d is a disability factor between 0 and 1 and Δy is the number of years between average age of onset and average age of death. As for the lost-life-year-based index, we only included contributions from the 20 listed diseases. The individuals selected from groups of size 10 had an increase of 4 DALY as compared to randomly selected individuals. This magnitude scale comports with previous studies[23].

The index tries to minimize the risk for several diseases simultaneously. In Fig. 4 we demonstrate how all the RRR from index selection compare to the RRR when selecting directly on the individual disease PRS, i.e., how much the index retains of the maximal risk reduction you would achieve if you focused on reducing a single disease. The direct PRS-selection tend — as naively expected — to reduce the specific diseases risk more than the index, especially for those diseases with very small weights (BCC, IBD). Yet, there are several examples where the index actually matches or even surpasses the direct PRS performance, most notably HA (probably because the strong/large comorbidity with CAD, HTN and obesity).

The PRS-comparison in Fig. 4 is a cross-section of the results at a group size of 5. The patterns are however consistent across all tested sizes, as seen in Fig. 5. The index reduces the risk of both T2D and CAD by about 50% at group size 10, consistently matching both the individual PRS-performances simultaneously. The consistent difference between PRS and index selection are also shown for Alzheimer's disease and obesity.

For the most prevalent diseases (ASA, HCL, HTN and obesity), we also provide prevalence-per-index quantile plots (odds ratio plots if divided by the general prevalence) in Fig. 6; the less prevalent diseases did not have enough cases for such high resolution. The top 4 percentiles have about half the risk of the bottom 4 percentiles to have either of hypercholesterolemia, hypertension, and obesity, while the risk reducing trend for asthma is less dramatic.

Genetic sibling pairs and trios. The primary results for the selection experiment on pairs of siblings is shown in Fig. 7, broken down into RRR and component index gain for each disease. The same graphs also include as reference the results from the selection among unrelated samples at group size 2. The sibling with the largest health index was selected from each of the 21,539 sibling pairs; no bootstrap was carried out. Instead the RRR error bars for the genetic siblings are theoretical 95% confidence intervals using the Wilson score interval for the prevalences among the selected siblings. They are generally larger than the corresponding error bars for the group size 2 bootstrap experiment. The limited data, for the rarest diseases in particular, decrease the certainty and result in the large error bars. Yet, we conclude from Fig. 7 that even in the most challenging task of minimizing the disease risk among only two genetic siblings the index provides a simultaneous and verifiable reduction of many diseases, while others are left inconclusive in this data set. Among the 20 studied diseases, there is no example of verified increased disease risk. Similarly, the estimated index gain is non-negative for all disease components and sum up to a significant gain also among pairs of genetic siblings. (The mean values for BCC and Gout are negative but much smaller in magnitude than the uncertainty.).

The index selection experiment result on the 969 trios had to the most part large uncertainties due the smallness of the data set and low case counts. Only two disease RRR reached statistical significance, according to the theoretical RRR confidence intervals. Hypercholesterolemia and obesity were confirmed with positive RRR, while hypertension and type II diabetes bordered to positive significance. No disease was confirmed to have negative RRR. The full RRR and index gain plots for trios are to be found in the Supplementary Information.
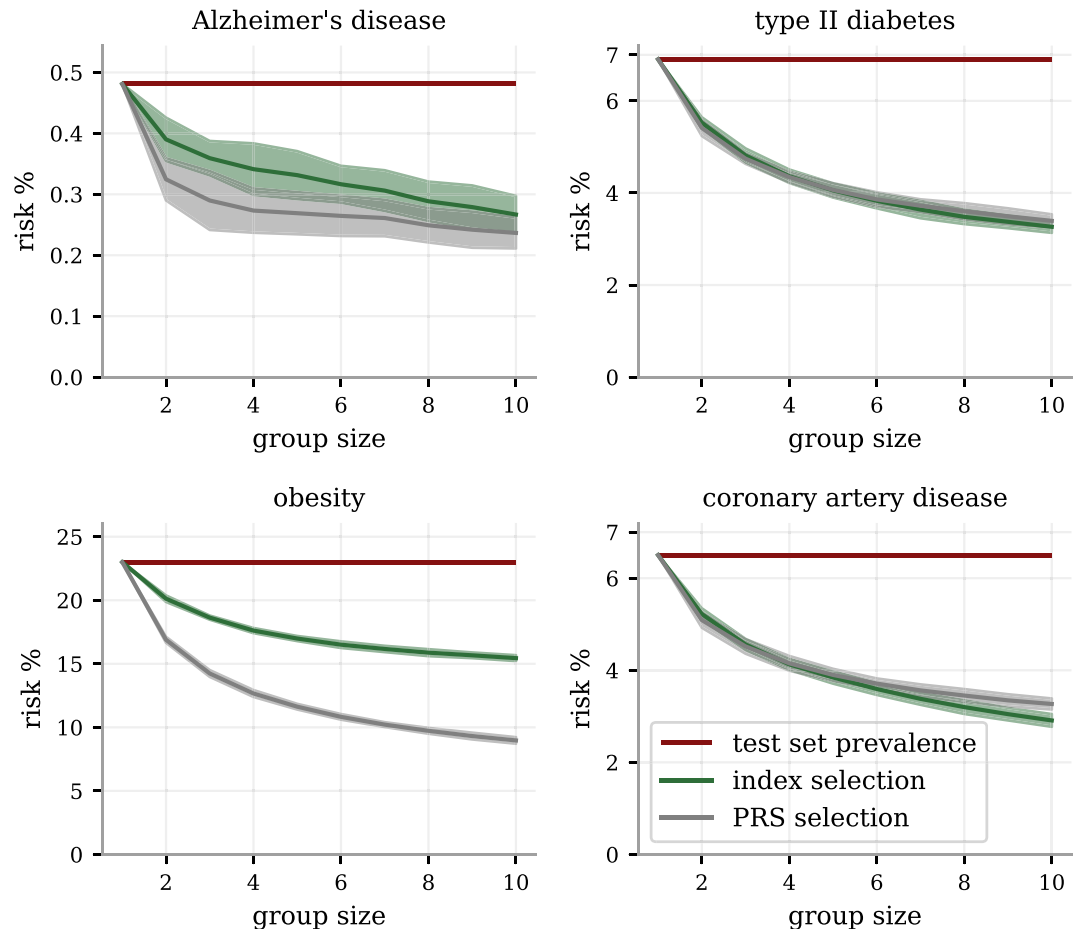
Figure 5. The disease risk reduction from index and PRS selection for different group sizes. The relative performance between index selection and PRS for individual diseases varies, as seen in Fig. 4. Here shown as functions of the group size, we see the strongest performance step between having no selection (group size 1) and selecting between between two and also the continued, but less dramatic, benefits with larger group sizes. Notably, for the chosen examples type II diabetes and CAD, the full health index consistently perform as well as selecting directly on the specific PRS, showing no reduced effects on these disease from taking all the other into account. The index performance for Alzheimer's disease and obesity, while not achieving the full risk reduction of their corresponding PRS, retain significant risk reductions for all group sizes. The error bars represent estimated 95% CI as computed by 25 selection experiments using different selection groupings.

Additional diseases and phenotypes. The t-tests for almost all the additional 11 diseases showed no statistical evidence for differences in mean for the health index distributions between cases and controls. That is, there is little to no relation between the non-significant diseases and the health index. Selecting on the health index would thus not affect these additional disease risks. Only in the cases of bipolar disorder and chronic obtrusive pulmonary disease (COPD) were there statistical significant differences between cases and controls among females. For males, only COPD and rheumatoid arthritis had significant differences. For all the mean differences of statistical significance, the health index is on average higher for the controls than for the cases. Note that no corrections for multiple testing was done and a Bonferroni correction (either with the number of diseases or number of sexes) would render the female bipolar result non-significant. Box plots, sample sizes and t-test p-values for all 11 diseases are presented in the Supplementary Information.

As with the 11 diseases, there were almost no significant deviations from equal health index means among the 5 addiction phenotypes. The statistical power was however much weaker due to the limited number of answering participants. Only male history of alcohol addiction had a significant mean difference between cases and controls, with cases having a slightly higher health index. Again, no correction for multiple testing was made and a Bonferroni correction (either per number of addictions or number of sexes) would leave all results non-significant. The box plots, sample sizes and t-test p-values for each addiction question are shown in the Supplementary Information.

The correlations with the additional continuous phenotypes were all weak but detectable. The strongest correlated trait was height at +0.06 for both males and females. While the correlations were small, the strong statistical power for these traits gave all linear regression slopes a non-zero value with high certainty. A table with correlations and p-values are presented in the Supplementary Information.
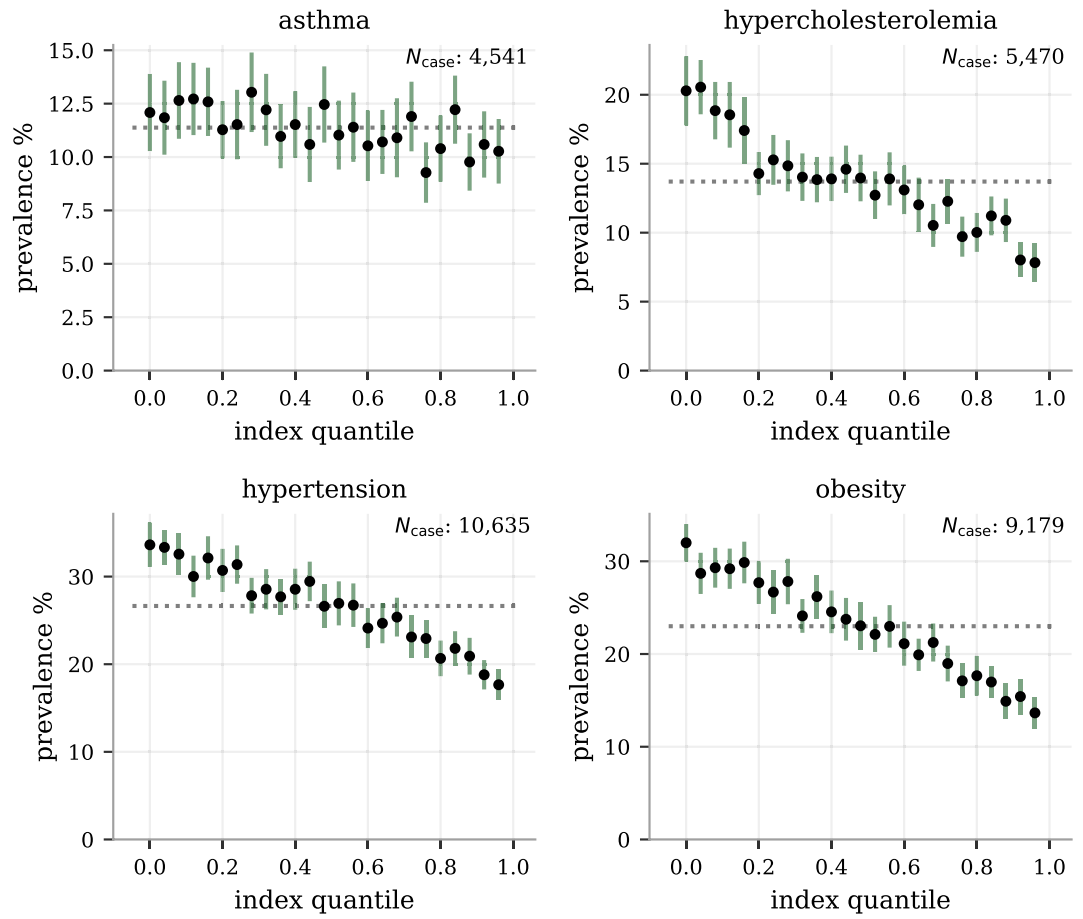
Figure 6. Prevalence in health index quantile bins for the most common diseases. We binned the test set according the health index into 25 equally distributed quantiles and plot the prevalence within each bin for the most prevalent diseases (allowing enough cases for the bin resolution to be meaningful). The general population prevalences are plotted as dotted reference lines (dividing with this number would give odds ratio plots) and the y-axis start at 0 to give a visual representation of the (odds) scales. For the intermediately risk reduced diseases (according to RRR Fig. 3) hypercholesterolemia, hypertension and obesity, there is a clear and systematic risk relationship across the entire range of the health index. For asthma, there is only a weak, detectable trend for the center values consistent with its existing but smaller RRR. The error bars are 95% CI estimates obtained through 100-fold bootstrap calculations of the prevalence within each bin (no re-binning was done).

Lastly, the multivariate linear regression using all additional phenotypes to predict the genetic health index did not explain any of the variance. The R^2 was 0.003 (std 0.009) for females and 0.005 (std 0.011) for males. We concluded that none of the additional $11 + 5 + 5$ phenotypes were linearly predictive of the genetic health index.

Characterization of phenotypic and genetic dependencies. The simultaneous disease risk reduction demonstrated for the index selection is bounded by potential disease dependencies, i.e., if two or more diseases tend to occur together (comorbidity) or are mutually exclusive. A commonly raised concern for PRS, and even more so for a composite health index, is the risk of antagonistic pleiotropy, i.e., that the same gene simultaneously increases the risk for one disease while decreasing the risk for another. Such a situation (or any cause of negatively correlated disease incidence) would impede simultaneous risk reduction. We examined this question for the 20 chosen diseases within our test set both on a genetic and phenotypic level. The result is presented in Fig. 8 through three quantities for each pair of diseases: the correlation between the PRS, the ratio between observed and expected comorbidity (called the χ^2 ratio), and the p-value of a χ^2 independence test (see figure caption for the details of the quantity visualization). The high information density in the plot requires some explanation but allows for quick comparison between all three quantities, both for individual pairs and for the disease set as a whole.

Contrary to the concern about strong impacts of antagonistic pleiotropy, we find that the disease incidences typically are pairwise dependent and overwhelmingly occur together. The predominantly solid green squares above the diagonal confirm that most of the disease pairs have comorbitities of statistical significance, in line with longstanding results such as coincidence of CAD and hypercholesterolemia. This makes a health index not only possible but an almost natural concept. The χ^2 ratio, lower green triangle—triangles below the diagonal,
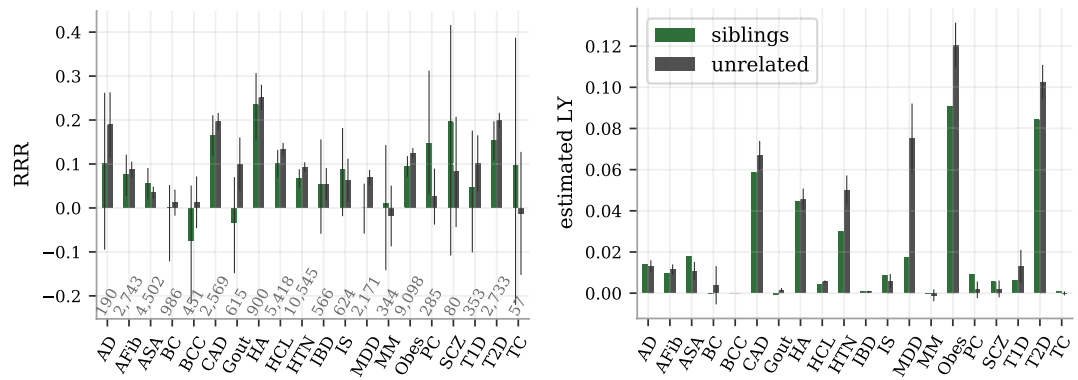
Figure 7. Index selection between 22,667 pairs of genetic siblings retain the overall benefits. In both figures, selection experiments among pairs of genetic siblings are compared to selection among pairs of unrelated individuals. The index performances are qualitatively very similar despite that siblings share half their genomes and have more similar environments. As expected, we do see a general performance attenuation among siblings, but also a few exceptions. Left: the RRR for each disease. The error bars for siblings are theoretical 95% C.I. using Wilson score interval for the prevalences among the selected siblings. The error bars for the selection among the unrelated pairs are again estimated 95% CI from 25 separate runs. The case numbers are shown above the x-axis. Right: the component-wise index gain for the selections among pairs of siblings and among pairs of unrelated individuals. The sibling results are presented without error bars since no theoretic uncertainty was calculated; statistical significance is therefore not established from this data. The error bars for the selection among unrelated individuals are 95% CI from 25 separate runs.

demonstrates the magnitude of the comorbidities, for example the very strong coincidences of (HA, CAD), (SCZ, MDD) and (T2D, T1D), and the moderate (HTN, AFib), (HTN, CAD) and (HCL, HA). The PRS correlations (upper blue triangle—triangles) are relatively small in magnitude and in general agreement with the phenotypic coincidences. As such, most PRS are relatively uncorrelated. Some notable exceptions are (HCL, CAD) and (MM, BCC). Just as the large amount of comorbidity facilitates the simultaneous positive RRRs, there are also some explanations for the lesser reductions here. The mutually exclusive tendency of (TC, CAD) complicates simultaneous risk reduction on a phenotypic level. (We are not aware of any research supporting this finding in other data sets. On the contrary, there are several examples of either inconclusive results or increased comorbidity of CAD among patients having undergone chemotherapy in TC treatment[51–53]. With our barely significant finding and small TC statistics, we view this result as peculiarity of the test set rather than a general epidemiological result.) This is in accordance with Fig. 4, where the RRR of TC is much stronger in PRS selection than index selection. The only examples of PRS level conflicts are the moderate anti-correlations between (T1D, IBD), (T1D, MDD) and (T2D, IBD), and the milder (BCC, ASA) and (IBD, ASA) anti-correlations, despite that these disease pairs are independent or have mild comorbidities. The combined index weights for ASA, T1D and T2D dwarf the impact of IBD on the index while BCC has no weight and is almost independent from everything else but MM (which is also independent from everything else). This contributes to the stronger RRR of PRS selection for ASA, BCC, IBD, and MM as compared to index selection.

Discussion

It is commonly believed that genetic factors influence overall health and longevity. With modern genomic methods we can test the scientific veracity of this hypothesis. By combining Polygenic Risk Scores (PRS) across the most impactful disease conditions, we can build a composite predictor of 20 diseases as part of an individual's overall health. The specific implementation studied in this paper used lifespan impact of each disease condition as the weighting factor in the index. We could then test whether this index predicts individual disease risks, as well as estimated longevity or disability adjusted life years.

Specifically, we validated this index in selection experiments using unrelated individuals and sibling pairs and trios from the UK Biobank. Individuals with higher index scores have decreased risk of individual diseases across almost all 20 diseases, with no significant risk increases, and longer calculated life expectancy. When Disability Adjusted Life Years (DALYs) due to the 20 diseases were used as the performance metric, the gain from genetic selection (highest index score vs average) among 10 individuals was found to be roughly 4 DALYs, and among 5 individuals was found to be 3 DALYs.

We found no statistical evidence for strong antagonistic trade-offs in risk reduction across these 20 diseases. Correlations between the disease risks are found to be mostly positive, and generally mild. This supports the folk notion of a general factor which characterizes overall health, sometimes described as synergistic pleiotropy. These results have important implications for public health and also for fundamental biological questions such as genetic architecture of human disease conditions.

The concept of pleiotropy was formulated before the notion of high dimensional spaces of genetic variation became familiar. The conventional logic is that, because a single gene can affect many different complex traits, it must be the case that different complex traits, such as disease risks, are themselves correlated, perhaps antagonistically (e.g., due to balancing selection, or for some deeper biochemical reason). This would entail
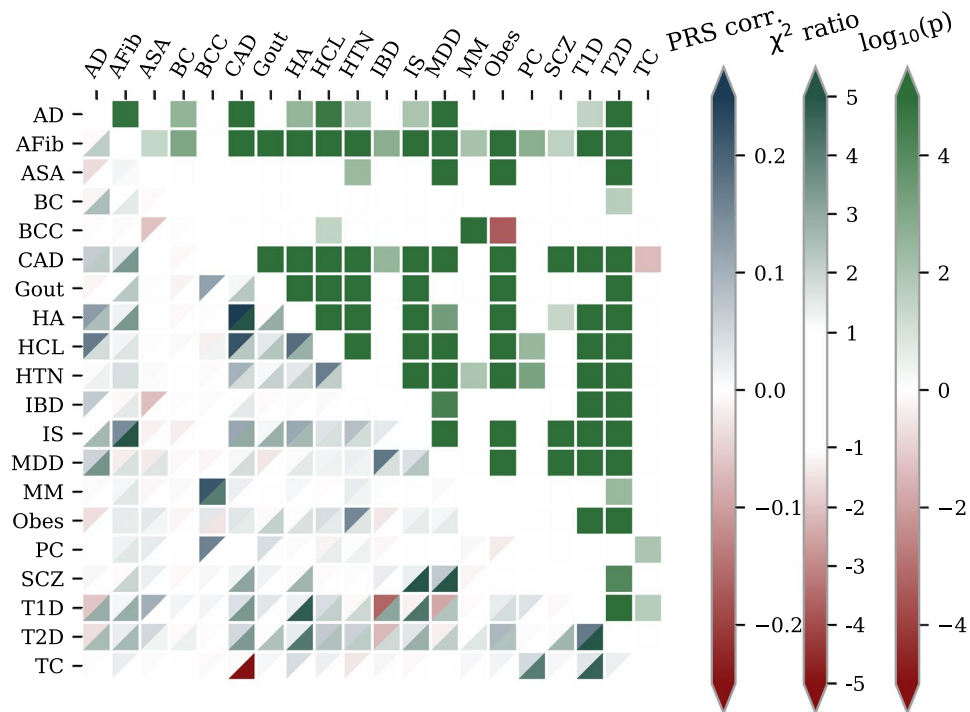
Figure 8. Phenotype dependencies and PRS correlation comparisons. This figure visualizes three different quantities for each pair of diseases: the PRS correlation, a comorbidity metric, and a χ^2 independence test p-value. Each tile below the diagonal is split into two halves where upper blue triangle = PRS corr. is the correlation between the two diseases' PRS, i.e., the genetic correlations as inferred by the predictors. The other half, lower green triangle = χ^2 ratio, is a metric of the actual disease comorbidity: how many more times is disease coincidence observed compared to what would be expected if the diseases were completely independent, where a positive (negative) sign indicates higher (lower) comorbid frequency (this is based on the ratio between the observed and expected case-case cell in a χ^2-test contingency table, hence referred to as the χ^2 ratio). The green/red squares = log(p), above the diagonal indicate the statistical significance of the dependence: the (signed) logarithm of the p-value in a χ^2-test. The sign is positive (negative) for more (less) frequent coincidence. Both the p-value and the χ^2 ratios are masked for disease pairs without statistically significant ($p = .05$) dependence. For example, the deep green square above the diagonal at (CAD,HCL) indicates that the CAD-hypercholesterolemia comorbidity is highly significant (we can reject phenotype independence at p-value $< 10^{-4}$). Below the diagonal, we see for the same disease pair that the lower triangle is gently blue-green, i.e., case coincidence for CAD-hypercholesterolemia is about 2.3 times more common than random chance. Lastly, the upper triangle is dark blue meaning that the PRS correlation between CAD and hypercholesterolemia is among the very strongest, at about 0.22. Overall, we see that most disease pairs have statistically significant comorbidity with 1–2 times more coincidence than chance, and that their PRS are not, or slightly positively, correlated. This phenotypic and genetic background not only allows but facilitates the construction of a useful health index. The most prominent outliers are discussed in the main text.

specific trade-offs, hypothetically: an individual with low diabetes risk might necessarily have higher cancer risk, etc. However, results from the modern era of GWAS and machine learning on large data sets show that the number of genetic loci which control a specific complex trait is typically in the thousands. It was shown in[54] that the SNP sets used in *sparse* predictors are largely disjoint for different traits or disease risks. The fact that most of the variance *can* be disjoint across different complex traits is a manifestation of high dimensionality. In this work we focus on *sparse* algorithms applied to array data which leaves open the possibility that there could be underlying causal loci that could still be correlated. However, the relatively small genetic correlations observed here leave this as an unlikely scenario.

In an earlier paper[54], we looked at the extent to which SNPs used in polygenic predictors of risk are correlated across pairs of disease conditions. Here we went further and investigated pairwise correlations between each of 20 major disease PRS. The results, as summarized in Fig. 8, can be expressed in words as: most correlations are modest, and tend to be positive rather than negative (antagonistic). (Modest correlation is consistent with mostly but not entirely disjoint variance in the two PRS.) We also concluded, on a phenotypic level, that the 20 diseases tend to have positive significant pairwise comorbidity.

It may be counter-intuitive that variants with exclusively deleterious effects have survived natural selection, especially widespread variants and in large numbers, as these results suggest[55]. It is possible that variants which solely increases disease risk, without positive contribution to fitness, would be selected against and disappear.

However, most of the 20 studied diseases have late onset and may reduce the lifespan from say 75 years to 65 within a modern day well-developed society. The lost fitness, even for the surviving (grand)children, is small and potentially negligible for all but a very short time in evolutionary history. This weak selection pressure is competing against the natural tendency for a population to accumulate random mutations. A full evolutionary genetics analysis of this would be an interesting continuation of our findings. Meanwhile we claim the results to be plausible even from an evolutionary perspective.

The proposed proxy-phenotype I^c for general health in UKB has some clear limitations. First, and as mentioned, it only takes the impact of the listed 20 diseases into account; the still not quantified impacts from other diseases and traits may have large contributions of either sign. Second, the UKB-cohort was 40+ years old at intake and although the medical records extend back to include early onset diagnoses for the participants, there is an inherent sampling bias against diseases with high mortality early in life. Quantifying those effects would increase the applicability of the results, in particular to embryo selection. Third, the index approximates the lost life years/disease burden under *current* environmental factors. The performance of the health index applied to embryo selection would most accurately be measured in the environments 40–70 years in the future (where some disease might be easily cured and others more common due to environmental changes). However, that is a limitation that applies to any type of health intervention (including recommendations of eating less cholesterol or already well-established pre-implantation testing); we will never be able to see into the future and the best we can do is to be aware of the potential discrepancy between today's environment and the future, while making as credible assumptions as possible. While the caveat still applies, we believe the selected 20 disease will remain relevant for the notion of general health also 70 years from now.

Let us also point out another limitation of the health index. The predictors used for the studied index are built using only *common* SNPs (very rare variants were filtered out from all training). Hence they do not, and in extension neither does the index, capture disease risks arising from rare but potentially very impactful variants. In the context of embryo selection, this is a minor concern since genetic pre-implantation testing usually includes additional monogenetic screening, targeting precisely such known variants. However, a full genetic health index intended for clinical use on adults should also include such risk contributions.

We focused this paper on index performance in a single cohort, and carried out cross-cohort analyses in other populations. We found substantial index performance in all populations, despite the expected and observed decreases in distant non-training populations. With expanded data availability, these cross-cohort analyses will be expanded in scope. There are already many research efforts dedicated to making the benefits of PRS available to more population groups, with efforts directed toward data collection, analysis, and clinical tools as end-products. It is an urgent task to make polygenic precision medicine not only as effective but also as equitable as it can be. To this end, follow-up health index studies in more cohorts are planned.

Data availibility

Access to the UK Biobank resource is available via application (http://www.ukbiobank.ac.uk).

References

1. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med.* **12**, 1–11 (2020).
2. Lewis, A. C. & Green, R. C. Polygenic risk scores in the clinic: New perspectives needed on familiar ethical issues. *Genome Med.* **13**, 1–10 (2021).
3. Richardson, T. G., Harrison, S., Hemani, G. & Smith, G. D. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *eLife* **8**, e43657 (2019).
4. Wray, N. R. *et al.* From basic science to clinical application of polygenic risk scores: A primer. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2020.3049 (2020)
5. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581 (2018).
6. Lello, L., Raben, T. G., Yong, S. Y., Tellier, L. C. & Hsu, S. D. H. Genomic prediction of 16 complex disease risks including heart attack, diabetes, breast and prostate cancer. *Sci. Rep.* **9**, 1–16 (2019).
7. Widen, E., Raben, T. G., Lello, L. & Hsu, S. D. H. Machine learning prediction of biomarkers from SNPs and of disease risk from biomarkers in the UK biobank. *Genes* **12**. ISSN: 2073-4425. https://www.mdpi.com/2073-4425/12/7/991 (2021).
8. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6**, 1000864 (2010).
9. Veenstra, D. L., Roth, J. A., Garrison Jr, L. P., Ramsey, S. D. & Burke, W. A formal risk 493 benefit framework for genomic tests: Facilitating the appropriate translation of genomics into clinical practice. *Genet. Med.* **12**, 686 (2010).
10. Amir, E., Freedman, O. C., Seruga, B. & Evans, D. G. Assessing women at high risk of breast cancer: A review of risk assessment models. *JNCI J. Natl. Cancer Inst.* **102**, 680–691 (2010).
11. Euesden, J., Lewis, C. M. & Oreilly, P. F. PRSice: Polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2014).
12. Abraham, G. *et al.* Accurate and robust genomic prediction of celiac disease using statistical learning. *PLOS Genet.* **10**, 1–15. https://doi.org/10.1371/journal.pgen.1004137 (2014).
13. Priest, J. R. & Ashley, E. A. Genomics in clinical practice (2014).
14. Jacob, H. J. *et al.* Genomics in clinical practice: Lessons from the front lines. *Sci. Translat. Med.* **21**, 5194cm5 (2013).
15. Shieh, Y. *et al.* Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res. Treat.* **159**, 513–525 (2016).
16. Bowdin, S. *et al.* Recommendations for the integration of genomics into clinical practice. *Genet. Med.* **18**, 1075 (2016).
17. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392 (2016).
18. Liu, L. & Kiryluk, K. Genome-wide polygenic risk predictors for kidney disease. *Nat. Rev. Nephrol.* **14**, 723–724 (2018).

19. Nelson, H. D., Pappas, M., Cantor, A., Haney, E. & Holmes, R. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: Updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* **322**, 666–685 (2019).

20. Kulm, S., Marderstein, A., Mezey, J. & Elemento, O. A systematic framework for assessing the clinical impact of polygenic risk scores. *medRxiv* 2020-04 (2021).

21. Wray, N. R. *et al.* From basic science to clinical application of polygenic risk scores: A primer. *JAMA Psychiatry* **78**, 101–109 (2021).

22. Meisner, A. *et al.* Combined utility of 25 disease and risk factor polygenic risk scores for stratifying risk of all-cause mortality. *Am. J. Hum. Genet.* **107**, 418–431 (2020).

23. Jukarainen, S., Kiiskinen, T., Havulinna, A. S. & Karjalainen, J. Genetic risk factors have a substantial impact on healthy life years. *medRxiv* 1–55. https://www.medrxiv.orgcontent/10.1101/2022.01.25.22269831v1 (2022).

24. Crawford, D. C., Cooke Bailey, J. N. & Briggs, F. Mind the gap: Resources required to receive, process and interpret research-returned whole genome data. *Hum. Genet.* **138**, 691–701 (2019).

25. Haga, S. B. *et al.* Public knowledge of and attitudes toward genetics and genetic testing. *Genet. Test. Mol. Biomark.* **17**, 327–335 (2013).

26. Hurle, B. *et al.* What does it mean to be genomically literate?: National Human Genome Research Institute meeting report. *Genet. Med.* **15**, 658–663 (2013).

27. Lea, D. H., Kaphingst, K. A., Bowen, D., Lipkus, I. & Hadley, D. W. Communicating genetic and genomic information: Health literacy and numeracy considerations. *Public Health Genomics* **14**, 279–289 (2011).

28. Dwyer, A. A. *et al.* Evaluating co-created patient-facing materials to increase under standing of genetic test results. *J. Genet. Counsel.* **30**, 598–605 (2021).

29. Moscarello, T., Murray, B., Reuter, C. M. & Demo, E. Direct-to-consumer raw genetic data and third-party interpretation services: More burden than bargain?. *Genet. Med.* **21**, 539–541 (2019).

30. Davis, K. W., Hamby Erby, L., Fiallos, K., Martin, M. & Wassman, E. R. A comparison of genomic laboratory reports and observations that may enhance their clinical utility for providers and patients. *Mol. Genet. Genomic Med.* **7**, e00551 (2019).

31. Kaye, C. & Korf, B. Genetic literacy and competency. *Pediatrics* **132**, S224–S230 (2013).

32. Henneman, L., Marteau, T. M. & Timmermans, D. R. Clinical geneticists' and genetic counselors' views on the communication of genetic risks: A qualitative study. *Patient Educ. Counsel.* **73**, 42–49 (2008).

33. The alarming rise of complex genetic testing in human embryo selection. *Nature* **603**, 549–550 (2022).

34. Forzano, F. *et al.* The use of polygenic risk scores in pre-implantation genetic testing: an unproven, unethical practice. *Eur J Hum Genet* **30**, 493–495. https://doi.org/10.1038/s41431-021-01000-x (2022).

35. Buddeke, J. *et al.* Comorbidity in patients with cardiovascular disease in primary care: A cohort study with routine healthcare data. *Eng. Br. J. Gen. Pract.* **69**, e398–e406 (2019) (**(ISSN:1478-5242 (Electronic); 0960-1643 (Print); 0960-1643 (Linking))**).

36. Institute of Medicine. *Cardiovascular Disability: Updating the Social Security Listings*. ISBN: 978-0-309-15698-1. https://nap.nationalacademies.org/catalog/12940/cardiovasculardisability-updating-the-social-security-listings. (The National Academies Press, 2010).

37. Long, A. N. & Dagogo-Jack, S. Comorbidities of diabetes and hypertension: Mechanisms and approach to target organ protection. *Eng. J. Clin. Hypertens. (Greenwich)* **13**, 244–251 (2011) (**(ISSN: 1751-7176 (Electronic); 1524-6175 (Print); 1524-6175 (Linking))**).

38. Bähler, C., Schoepfer, A. M., Vavricka, S. R., Brüngger, B. & Reich, O. Chronic comorbidities associated with inflammatory bowel disease: Prevalence and impact on healthcare costs in Switzerland. *Eur. J. Gastroenterol. Hepatol.* **29**. https://journals.lww.com/eurojgh/Fulltext/2017/08000/Chronic_comorbidities_associated_with_inflammatory.8.aspx (2017).

39. Wang, J.-H., Wu, Y.-J., Tee, B. L. & Lo, R. Y. Medical comorbidity in Alzheimer's disease: A nested case-control study. *Eng. J. Alzheimers Dis.* **63**, 773–781 (2018) (**(ISSN: 1875- 8908 (Electronic); 1387-2877 (Linking))**).

40. Santiago, J. A. & Potashkin, J. A. The impact of disease comorbidities in Alzheimer's disease. *Eng. Front. Aging Neurosci.* **13**, 631770 (2021) (**(ISSN: 1663-4365 (Print); 1663-4365 (Electronic); 1663-4365 (Linking))**).

41. Al-Asadi, A. M., Klein, B. & Meyer, D. Multiple comorbidities of 21 psychological disorders and relationships with psychosocial variables: a study of the online assessment and diagnostic system within a web-based population. *J. Med. Internet Res.* **17**, e55–e55. https://pubmed.ncbi.nlm.nih.gov/25803420 (2015).

42. Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Eng. Arch. Gen. Psychiatry* **62**, 617–627 (2005) (**(ISSN: 0003-990X (Print); 1538-3636 (Electronic); 0003-990X (Linking))**).

43. Farabaugh, A. *et al.* Relationships between major depressive disorder and comorbid anxiety and personality disorders. *Eng. Compr. Psychiatry* **46**, 266–271 (2005) (**(ISSN: 0010-440X (Print); 0010-440X (Linking))**).

44. Slade, T. & Watson, D. The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. *Eng. Psychol. Med.* **36**, 1593–1600 (2006) (**(ISSN: 0033-2917 (Print); 0033-2917 (Linking) (2006))**).

45. Vollebergh, W. A. *et al.* The structure and stability of common mental disorders: The NEMESIS study. *Eng. Arch Gen Psychiatry* **58**, 597–603 (2001) (**(ISSN: 0003-990X (Print); 0003- 990X (Linking))**).

46. Buckley, P. F., Miller, B. J., Lehrer, D. S. & Castle, D. J. Psychiatric comorbidities and schizophrenia. *Eng. Schizophr. Bull.* **35**, 383–402 (2009) (**(ISSN: 0586-7614 (Print); 1745-1701 (Electronic); 0586-7614 (Linking))**).

47. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584 (2019).

48. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 12–23 (2022) (**(ISSN: 15376605)**).

49. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**. ISSN: 1061-4036 (2022).

50. Shi, H. *et al.* Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat. Commun.* ISSN: 20411723. http://dx.doi.org/10.1038/s41467-021-21286-1 (2021).

51. Gugic, J., Zaletel, L. Z. & Oblak, I. Treatment-related cardiovascular toxicity in long term survivors of testicular cancer. *Eng. Radiol. Oncol.* **51**, 221–227 (2017) (**(ISSN: 1318-2099 (Print); 1581-3207 (Electronic); 1318-2099 (Linking))**).

52. Feldman, D. R. *et al.* Predicting cardiovascular disease among testicular cancer survivors after modern cisplatin-based chemotherapy: Application of the Framingham risk score. *Clin. Genitour. Cancer* **16**, e761–e769. https://pubmed.ncbi.nlm.nih.gov/29534941 (2018).

53. Zaid, M. A. *et al.* Clinical and genetic risk factors for adverse metabolic outcomes in North American testicular cancer survivors. *J. Natl. Compr. Cancer Netw.* **16**, 257–265 (2018).

54. Yong, S. Y., Raben, T. G., Lello, L. & Hsu, S. D. Genetic architecture of complex traits and disease risk predictors. *Sci. Rep.* **10**, 1–14 (2020).

55. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).

Acknowledgements

Author contributions

Authors, listed alphabetically according to last name, contributed in the following ways: conceptualization, S.D.H.H, L.L., L.T.; methodology, S.D.H.H., L.L., T.G.R., L.T., E.W.; software, L.L., E.W.; validation, S.D.H.H, L.L., T.G.R., L.T., E.W.; formal analysis, S.D.H.H, L.L., T.G.R., L.T., E.W.; investigation, L.L., E.W.; resources, S.D.H.H.; data curation, L.L., L.T., E.W.; writing-original draft preparation, E.W.; writing-review and editing, S.D.H.H, L.L., T.G.R., L.T., E.W.; visualization, L.L., L.T., E.W.; supervision, S.D.H.H., L.T.; project administration, S.D.H.H., L.T.; funding acquisition, S.D.H.H., L.T. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare the following competing interests: SH is a founder, shareholder, and serves on the Board of Directors of Genomic Prediction, Inc. (GP). LT is a founder, shareholder, serves on the Board of Directors, and is the CEO of GP. EW and LL are employees and shareholders of GP. TR declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22637-8.

Correspondence and requests for materials should be addressed to E.W. or L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.