Intergenerational Correlation in Mortality: The Role of Chronic Conditions

Dan Zeltzer, Liran Einav, Joseph Rashba,

Nicholas Scott-Hearn, and Ran Balicer*

November 4, 2025

Abstract

We examine the contribution of intergenerational transmission of chronic conditions to the correlation in mortality across generations. Utilizing comprehensive administrative claims data linked across generations from Israel's largest healthcare system, we analyze the morbidity and mortality of individuals and their parents. We first estimate the elevation in morbidity hazard associated with family history of different chronic conditions, and the subsequent elevation in mortality hazard following the onset of each condition. We then embed these estimates into a model in order to quantify the life-expectancy reduction associated with family history of different conditions and the contribution of chronic conditions to the overall intergenerational correlation in mortality. Results show that family history of common chronic conditions decreases life expectancy by several months. The intergenerational transmission of these conditions accounts for 9–14% of the overall intergenerational correlation in mortality.

*Zeltzer (corresponding author): Tel Aviv University and Stanford University, dzeltzer@tauex.tau.ac.il; Einav: Stanford University and NBER. Rashba: Tel Aviv University and Clalit Health Services. Scott-Hearn: Stanford University. Balicer: Clalit Health Services and Ben Gurion University. Zeltzer acknowledges financial support from the Foerder Institute for Economic Research. We thank Gueyon Kim, Gina Li, Torben Nielsen, and participants in ASHEcon 2024, AEA 2025 meetings, and CEPRA/NBER Aging and Health 2025 for helpful comments. Clalit IRB 0162-21-COM2.

JEL Codes: I11, I14, J10.

Keywords: health mobility, chronic conditions, life expectancy.

1 Introduction

The persistence of health and health inequality across generations has long been a focus of economics research (Currie and Moretti, 2007; Black et al., 2011; Currie, 2011; Lu and Vogl, 2023). The length of life (lifespan), perhaps the most fundamental measure of health, is well known to be significantly correlated across generations (Beeton and Pearson, 1901; Pearl, 1931; Black et al., 2023). However, the underlying drivers of this correlation in mortality remain largely unexplored. Understanding the role of family disease history in shaping individual health outcomes is important for understanding fundamental aspects of inequality and social mobility.

In this study, we take advantage of a unique data opportunity to quantify the role of chronic conditions in the intergenerational correlation in mortality. The analysis is facilitated by linking detailed morbidity patterns of a large, diverse population from administrative healthcare records with vital statistics data on mortality. Existing work on intergenerational correlation in longevity traditionally uses data on linked generations that lack morbidity information. Our key contribution lies in linking morbidity and mortality records across generations, which allows for novel insights into how intergenerational correlations in morbidity and mortality are related.

The backbone of our exercise is a detailed and comprehensive medical data, covering about half of the Israeli population from 2009–2022. The data come from Clalit Health Services, the largest of four non-profit health maintenance organizations (HMOs) in Israel that provide all residents universal, tax-funded health insurance. The data include detailed chronic condition information and date of death. We link these data across generations based on information from the population registry on family relations.

The longitudinal nature of the data provides an average observation period of several years per individual, allowing us to track their health trajectories, including the onset of a rich set of chronic conditions and, in some cases, mortality. Utilizing this dataset, we observe two

critical risk processes: the age-specific incidence of different chronic conditions for individuals with and without family history of the condition, and the age-specific mortality rates for individuals with and without each condition.

To quantify the reduction in life expectancy associated with family history of different conditions and to determine the contribution of chronic conditions to the intergenerational correlation in mortality, we integrate morbidity and mortality hazard estimates into a life-cycle model. We use descriptive evidence to parameterize the model, and estimate its parameters. With the estimated model, we calculate the difference in life expectancy between individuals with and without family histories of specific conditions, and simulate the overall intergenerational correlation in mortality attributable to chronic conditions and compare it to previous estimates of overall correlation in mortality.

Our results indicate that family history of neoplasms (which includes cancers), cardiovascular conditions, and psychiatric conditions induces the largest life expectancy reductions. Individuals with family history of these conditions have shorter life expectancies by 0.70, 0.67, and 0.52 years, respectively. These three conditions are also the strongest contributors to intergenerational correlation in mortality, collectively accounting for approximately half of the overall contribution of chronic conditions to intergenerational mortality correlation. Overall, chronic conditions predict the intergenerational correlation in mortality to be 1.3%, which amounts to about 9% to 14% of the total intergenerational correlation in mortality, estimated by a rich existing literature to be between 9% and 14% (Black et al., 2023).

Our study contributes to the broader understanding of intergenerational health mobility. It adds to a growing body of recent literature focusing on intergenerational mobility in different aspects of health. Johnston et al. (2013) documents intergenerational correlation in mental health, and Halliday et al. (2021) estimates intergenerational correlation in self-reported health status. Thompson (2014, 2017) use samples of adoptees to study the genetic mechanisms underlying intergenerational associations in chronic health conditions. Jia et al. (2019) analyzes the heritability of specific diseases from Electronic Health Record

(EHR) data. Andersen (2021) uses Danish registry data to quantify the correlation in latent health measures based on large administrative data across parent-child, sibling, and twin pairs. Chang et al. (2024) utilizes administrative data from Taiwan to investigate the intergenerational persistence in latent health, measured based on condition indicators and outpatient care utilization.

This study also adds to a demographic literature that explores inherited drivers of morbidity (Mahmood et al., 2014; Polubriaginof et al., 2018) and mortality (Vaupel, 1988; Ruby et al., 2018; Slagboom et al., 2018; Christensen et al., 2020). Finally, this work is related to Danesh et al. (2024), which shows that the early-life accumulation of chronic conditions drives income-related health disparities. While they focus on chronic conditions' impact on socioeconomic disparities, we highlight and quantify their contribution to intergenerational mortality persistence.

The rest of this paper continues as follows. Section 2 describes the data and presents descriptive patterns, Section 3 lays out the lifecycle model and its estimation, and Section 4 reports the main results. The last section concludes.

2 Data and descriptive evidence

2.1 Data source

Our data are sourced from Clalit Health Services (in short, Clalit), the largest among Israel's four non-profit health maintenance organizations (HMOs) that provide universal, mandatory, tax-funded healthcare coverage to all Israeli residents from birth, in accordance with the National Health Insurance Law (1995). Under Israeli health insurance, covered services are fully funded by government capitated payments.

The data cover a large and stable population. Clalit serves over half of the Israeli population, with approximately 5 million enrollees at the end of our sample period. All four HMOs offer identical coverage and share the same hospital networks but have distinct outpatient

provider networks. Individuals can switch HMOs twice a year while retaining their universal coverage. However, the annual switching rate is remarkably low (around 1%), and newborns are assigned to their mother's HMO by default. Thus, most individuals remain enrolled with Clalit throughout their lifetime, and it is common for children and parents to be enrolled with the same HMO.

The data provide a comprehensive record of the population's chronic conditions. It has three main advantages in that respect. First, Electronic Health Record (EHR) systems were adopted well before the start of our study period, so data quality is generally high. Second, by law, Clalit and all other HMOs maintain a registry of over one hundred conditions for care management and reimbursement purposes (for example, certain chronic conditions qualify for a different maximum out-of-pocket cap on drug spending). If individuals switch insurers, the law mandates transferring their chronic condition registry records, so even new enrollees have a record of any previously-diagnosed chronic conditions. Third, under-diagnosis is likely minimal, thanks to universal coverage and free primary care.

We augment these data with detailed and complete vital statistics data on mortality and family links, sourced from the Israeli population registry and available to us through Clalit. We observe birth and death dates for any individual who has been a Clalit enrollee at some point in their life. These data also include unique identifiers that allow us to link Clalit enrollees to their children, when the latter are also Clalit enrollees.

2.2 Sample construction and key variables

Sample Our sample consists of adults aged 30 and older with at least 12 months of coverage. We begin with the full population of 6.2 million individuals ever enrolled during the study period (2009–2022). We restrict the sample to adults aged 30 or older during the study period (i.e., born before 1993), as most chronic conditions typically emerge after this age, reducing the sample to approximately 3.3 million. We then exclude individuals without at least 12 months of continuous enrollment, yielding a final sample of 3.1 million enrollees,

which we refer to as the *general sample*. Within this sample, 3.0 million enrollees were observed at an age equal to or lower than 80. We use this sample to estimate overall mortality and morbidity hazard rates.

For all other analyses, we use the *linked-generations sample*, which includes all observed parent-child pairs. Starting from the general sample, we construct this sample by including every individual who has at least one parent also present in that sample. Because we observe very few child-parent pairs where the child is older than 80 years old, we exclude such pairs, and analyze morbidity and mortality rates up to the age of 80. Restricting to adults aged 30–80 means our analysis abstracts from intergenerational correlations driven by mortality at very young or very old ages. The linked-generations sample includes a total of 2.3 million unique individuals, with 1.5 million unique children and 1 million unique parents (some individuals who are part of three-generation families appear in both roles).[1] We observe both parents for 71% of the children in this sample and just one parent for the remaining 29%. Appendix Figure A1 describes the joint age distribution of parents and children in our linked generations sample.

Morbidity and mortality measures Our analysis focuses on mortality, chronic conditions, and their intergenerational relationships. In order to avoid small cells related to less common conditions, we group all chronic conditions into 15 broader categories; Appendix Table A1 provides the full list of conditions in each group (for brevity, in what follows we refer to each such group as "condition").

We measure the overall prevalence of each condition in both the general sample and the linked-generations sample, separately for (unique) parents and children. Using the linked-generations sample, we estimate the age-specific hazard rate of obtaining each condition; that is, the probability of being diagnosed with the condition at a certain age, conditional

[1]Families with multiple parents and children are represented in the sample by multiple observations—one for each parent-child pair. For example, a family with two parents and two children would have four unique parent-child pairs. In all, 31% of the unique individuals in the linked-generations sample are part of three-generation families.

on reaching that age without a prior diagnosis of the condition. We estimate this hazard separately for individuals with and without family history of the condition, defined as a binary indicator equal to one if at least one parent has ever been diagnosed with the condition. Similarly, we estimate the age-specific mortality hazard rate, both overall and separately for individuals with and without a prior diagnosis of the condition at each age.

2.3 Descriptive evidence

Sample composition Appendix Table A2 shows the prevalence of chronic conditions at age 80 for our different samples. The most common conditions at this age are endocrine (which include diabetes and thyroid disorders, 80%), cardiovascular (79%), and musculoskeletal (56%). Reassuringly, prevalence rates are similar between the general and linked-generations samples, suggesting selection into the linked sample is limited.

Figure 1 displays the age-specific hazards for the onset of various conditions among individuals with and without family history of each condition. As expected, the hazard for most conditions increases with age. The figure also shows a consistent difference, across all conditions, between individuals with and without family history, indicating that family history increases the likelihood of developing these conditions. Appendix Figure A4 shows the difference in the log hazard of diagnosis between individuals with and without family history, by condition. This difference remains relatively stable across ages for most conditions, suggesting that family history leads to a proportional increase in the hazard of developing a condition. The magnitude of this increase in hazard varies by condition: for hematological conditions, family history is associated with a 2.7 log-point increase, whereas for genitourinary conditions, the increase is only 0.11 log points.

Appendix Figure A3 shows the estimated mortality rate by age in the general population. Mortality hazard increases steadily with age. A nearly linear-in-logs relationship between age and mortality hazard is well known; it is consistent with the common use of a Gompertz distribution, which posits that the mortality hazard rate increases exponentially with age.

Figure 2 presents the mortality rates by age for individuals with and without each condition. The near-linear relationship in logs persists across subgroups. As may be expected, individuals with chronic conditions exhibit higher mortality rates. Appendix Figure A5 shows the differences in mortality rates between individuals with and without each condition. It reveals that the elevated mortality rates associated with all conditions are higher at younger ages and diminish with age. The decrease with age is approximately linear for most conditions.

Taken together, the descriptive evidence in this section shows that the incidence rates of nearly all chronic conditions are higher among individuals with family history. Moreover, the onset of these conditions is associated with elevated mortality risk. These patterns imply that parents with chronic conditions are more likely to die at younger ages and are also more likely to have children who develop chronic conditions, and who, in turn, also face higher mortality risk. In other words, these findings point to an intergenerational correlation in mortality. However, quantitatively linking these patterns across different conditions and ages to our primary object of interest, the intergenerational correlation in mortality, requires a more structured framework.[2] We turn to that next.

3 Model and estimation

To quantify the contribution of chronic conditions to the intergenerational correlation in mortality, we embed morbidity and mortality hazard estimates into a lifecycle model.[3]

Consider a parent-child pair, and denote by T_c and T_p the child's and parent's realized age at the time of death, respectively. The intergenerational correlation in mortality is governed by the joint distribution of T_c and T_p, and in particular by how the child's expected life varies as a function of the parent's age at death, $E(T_c|T_p)$. Within this framework, our key

[2]Most individuals are still alive at the end of the observation period, so complete mortality is not observed. This is a common feature of modern administrative health data.

[3]This approach follows the tradition of synthetic cohort models in demography, which use age-specific hazard rates observed in cross-sectional or short-panel data to simulate complete life trajectories of hypothetical individuals (Preston et al., 2001).

focus is on the contribution of chronic conditions to this correlation.

Model If $m_i(a)$ denotes the mortality hazard rate of individual i at age a, her probability of dying at age T_i is given by

$$Pr_i(T_i = t) = m_i(t) \prod_{a=a_0}^{t-1} \big(1 - m_i(a)\big). \tag{1}$$

and her life expectancy is given by

$$E[T_i] = \sum_{a=a_0}^{\bar{a}} a Pr_i(T_i = a), \tag{2}$$

where a_0 and $\bar{a} > a_0$ define the relevant age range for the analysis.

We now consider J chronic conditions indexed by $j = 1, 2, ..., J$. We denote by $G_i(a) \subseteq J$ the set of chronic conditions individual i has at age a, and we make the strong assumption that individual i's mortality hazard rate is only a function of their age and the chronic conditions they have, that is $m_i(a) = m\big(a, G_i(a)\big)$.

Let $h_{ij}(a)$ be the hazard rate of individual i developing condition j at age a and let O_{ij} denote the age of onset of condition j for individual i. The probability that the onset of condition j occurs at age o (conditional on survival) is given by

$$Pr_i(O_{ij} = o) = h_{ij}(o) \prod_{a=a_0}^{o-1} \big(1 - h_{ij}(a)\big), \tag{3}$$

We set, conventionally, $O_{ij} = \infty$ for individuals who never develop condition j before the maximal age $\bar{a}$; For such individuals, $Pr_i(O_{ij} = \infty) = \prod_{a=a_0}^{\bar{a}} \big(1 - h_{ij}(a)\big)$.

Collecting all these objects together, expected life is then given by applying equation (2) and integrating over the stochastic arrival rates of the J chronic conditions.

Parameterization and estimation We parameterize the model by making several tractable assumptions, which are motivated by the descriptive patterns we presented in Section 2.3.

First, we assume chronic conditions arrive independently of each other. That is, $h_{ij}(a)$ does not depend on which other conditions individual i already has at age a, and the arrival of each condition, conditional on the hazard rate, is independent of the arrival of the others.

Second, the mortality hazard rate can be written as

$$m_i(a) = m(a, \emptyset) \prod_{j \in G_i(a)} e^{\gamma_j(a)}, \tag{4}$$

where $m(a, \emptyset)$ denotes the baseline mortality hazard at age a without any diagnosed condition, and $\gamma_j(a) = \gamma_j^0 + \gamma_j^1 a$. That is, each condition contributes multiplicatively to the mortality hazard, with an effect that is log-linear in age.

Similarly, we assume the incidence hazard only depends on parental history, which has a constant, age-invariant multiplicative effect on it. Formally, let F_{ij} be an indicator for whether individual i has family history of condition j. We assume that $h_{ij}(a)/h_{i'j}(a) = e^{\beta_j(F_{ij} - F_{i'j})}$ for all individuals i, i'. Hence we can write the incidence hazard at age a as

$$h_{ij}(a) = h_j(a)\, e^{\beta_j F_{ij}}, \tag{5}$$

where $h_j(a)$ is the baseline hazard of incidence of condition j for individuals without a family history of j, and e^{β_j} is the constant hazard ratio associated with having such a history. We refer to β_j as the *intergenerational transmissibility* of condition j.

Taken together, the model parameters are given by four components: (i) the baseline, age-specific mortality hazard rate for healthy individuals, $m(a, \emptyset)$; (ii) the age-specific hazard rate of the arrival of each chronic condition j for individuals without family history of the condition, $h_j(a)$; (iii) the impact of each chronic condition on mortality, γ_j; and (iv) the intergenerational transmissibility of each chronic condition, β_j. These objects are closely related to the descriptive patterns presented in Section 2.3.

We estimate age-specific mortality and condition-onset hazards semi-parametrically. First, we recover the age-specific baseline hazards non-parametrically, directly from the data. That

is, using the general sample, for each age a we estimate $m(a, \emptyset)$ as the share of at-risk person-years (defined as individuals observed alive at age a with no chronic conditions) who die between ages a and $a+1$. Using the linked-generations sample, for each chronic condition j and age a, we estimate $h_j(a)$ as the number of first diagnoses of condition j at age a among individuals with no family history of j, divided by the corresponding person-years at risk.

We then fit Cox proportional-hazards models to recover the morbidity increase associated with family history and the mortality increase associated with the onset of different conditions. First, in the linked-generations sample, we regress time to onset of condition j on the family-history indicator F_{ij}, recovering β_j in equation (5) as the hazard ratio associated with family history (defined as having at least one parent diagnosed with the condition), while leaving the baseline hazard unspecified in age. Second, in the general sample, we regress time to death on each condition indicator and its interaction with age, yielding γ_j^0 (the log-multiplier intercept) and γ_j^1 (the log-linear age slope) in equation (4). Appendix A provides detailed estimation specifications.

Together with our non-parametric estimates of baseline hazards, these semi-parametric estimates fully parameterize the life-cycle model used to simulate life expectancies and evaluate the contribution of chronic conditions to intergenerational mortality correlations.

Model-Based Calculations and Simulations To clarify the intergenerational impact of different chronic conditions, we calculate for each condition j the *reduction in life expectancy associated with family history* of the condition, defined as:

$$\Delta LE_j = E[T_i|F_{ij} = 1] - E[T_i|F_{ij} = 0]. \tag{6}$$

Namely, the (typically negative) change in a child's life expectancy associated with a parental history of condition j. It is computed by integrating over the distribution of the child's age at condition onset and, conditional on onset, over the distribution of age at death (see Appendix B for details). It reflects a combination of elevated incidence risk linked to family

history and the increased mortality hazard associated with the condition, once it arrives. Thus, ΔLE_j can serve as an aggregate measure of the intergenerational burden associated with each condition j.

We describe the relationship between the life-expectancy reduction ΔLE_j and three condition-specific fundamentals. First, β_j, which captures the intergenerational transmissibility of condition j defined above. Second, π_j, the baseline *lifetime prevalence* of condition j, defined as the probability of developing the condition by age $\bar{a}$ in a cohort with no competing mortality risks.[4] This provides a standardized, comparable measure of how common a condition is in the absence of competing-risk bias. Third, we define the condition's *mortality burden* as the difference in life expectancy between individuals who ever develop the condition and those who never do, conditional on no family history: $\lambda_j = E[T_i \mid O_{ij} \leq \bar{a}, F_{ij} = 0] - E[T_i \mid O_{ij} = \infty, F_{ij} = 0]$. This quantity captures the average life-years lost due to the condition, absent family-related risk.[5] Each of these three quantities captures a distinct channel contributing to the reduction in life expectancy associated with family history.[6]

To more comprehensively quantify the role of chronic conditions in the intergenerational correlation of mortality, we simulate synthetic cohorts using the model and estimates derived from our sample. Each cohort consists of N families, each with a parent and a child. Utilizing the nonparametric estimates for the baseline morbidity hazard for each condition, $h_j(a)$, we determine for every parent the risk profile over their lifetime for developing different conditions and simulate the age at onset of any conditions they acquire (up to age 80) by drawing realizations from this risk process. Given these morbidity realizations, we compute the age-specific mortality hazard ($m_i(a)$ from equation (4)) for each parent, and draw their age at death from this mortality hazard. We then assign each child the vector of family history of all conditions $(F_{ij})_{j \in J}$, which records the conditions their parent would have encountered

[4]Formally, $\pi_j = \Pr(O_{ij} \leq \bar{a} \mid F_{ij} = 0)$, where $\bar{a} = 80$.

[5]The mortality burden of a condition is closely related to the average γ_j, its age-specific impact on mortality hazard, weighted over the age-of-onset distribution.

[6]In the simple case of a congenital condition (arriving only at a_0), $\Delta LE_j \approx \pi_j \beta_j \lambda_j$. See Appendix B.

had they lived to 80. Similar to lifetime prevalence, this definition is free of competing-risk bias. Using this family history, we calculate each child's lifetime risk process for developing each condition and draw their realized ages of onset from these profiles. Finally, based on the child's realized morbidity trajectory, we compute their age-specific mortality risk process and draw their age at death. All draws are independent across families. In the resulting simulated population, we calculate the correlation between realized child and parent lifespan and its standard error using 200 bootstrap replications.

With this simulation framework, we conduct several analyses. First, we estimate the intergenerational correlation in lifespan when all chronic conditions are included. To quantify the contribution of specific conditions to this correlation, we then calculate the correlation in alternative scenarios where we neutralize the intergenerational transmission ($\beta_j = 0$) of one condition at a time, while retaining each condition's direct impact on mortality. Additionally, we examine how this correlation accumulates over the entire life duration by calculating the intergenerational correlation with the maximum lifespan winsorized at different ages.

4 Results

Parameter estimates and model-based calculations Table 1 (columns (1)–(3)) presents our estimates of the key model parameters. Column (1) displays estimates of the intergenerational transmissibility parameter β, the increase in incidence hazard of each condition with family history of the condition, derived from fitting the difference between the nonparametric hazard rates of onset for these groups. For example, family history of cardiovascular conditions increases the hazard by a factor of $e^{0.558} = 1.747$, meaning individuals with family history are 74.7% more likely to develop the condition at any age compared to those without. Loosely, this corresponds to the average vertical difference between the nonparametric log hazard curves in Figure 1 (a difference that is also directly depicted in Appendix Figure A4). Mirroring this evidence, column (1) shows that conditions exhibit considerable variability in their intergenerational transmissibility, with a proportional increase (e^{β}) ranging from 1.11

for genitourinary conditions to 15 for hematological conditions.

Columns (2) and (3) of Table 1 show the elevated mortality risk associated with the onset of each condition (γ_0) and its change with age (γ_1). Loosely, these parameters correspond to fitting a linear model to the vertical differences between the age-specific log mortality hazards between individuals with and without each condition, as depicted by the difference between the fitted curves in Figure 2, explicitly plotted in Appendix Figure A5. Both the baseline (age 30) increase in mortality risk and its age gradient vary substantially across conditions.

Columns (4)-(6) of Table 1 show additional characteristics important for intergenerational correlation in mortality that we calculated from the estimated model. Column (4) summarizes the mortality burden (λ)—the difference in life expectancy between those with and without each condition (which depends on both the baseline mortality hazard m and its increase with morbidity, γ). Neoplasms (the condition class that includes solid cancers) have the highest direct mortality burden, 10 life years. Dermatological conditions have the lowest, at 0.15 life years.

Column (5) shows baseline lifetime prevalence (π), capturing the prevalence of a condition (at $\bar{a} = 80$) in a population not subject to mortality risk.[7] Conditions vary greatly in their lifetime prevalence, from approximately 1% for hematological conditions to most of the population for endocrine conditions. Prevalence is important for intergenerational correlation in mortality because rare conditions, by definition, can, at most, affect only a small share of the population.

Life expectancy reduction associated with family history of different conditions

Column (6) of Table 1 reports a key summary measure of intergenerational mortality impacts of chronic condition—ΔLE, the reduction in life expectancy associated with family history of each condition. Family history of neoplasms and cardiovascular diseases leads to the largest

[7]This counterfactual measure correlates strongly with the actual prevalence of conditions in our sample (column (2) of Appendix Table A2; Corr = 0.972).

life expectancy reductions: 0.7 and 0.67 years, respectively. Individuals with family history of psychiatric conditions have a life expectancy half a year shorter than those without such a history. In contrast, other conditions, such as musculoskeletal and dermatological conditions, have negligible intergenerational impacts.

Comparing different conditions highlights the combined importance of different factors in determining the intergenerational life-expectancy gaps. For example, cardiovascular diseases are both highly prevalent and correlated across generations, and thus have intergenerational impact on mortality. In contrast, neoplasms are not as prevalent, and less correlated across generations, but are more deadly. As these examples show, the transmissibility of mortality across generations is not driven by a single factor but by an interplay of factors: prevalence, intergenerational transmissibility, and mortality burden. Moreover, these facturs are not particularly correlated within conditions (see Appendix Figure A6).

Further, the relationship between different factors and ΔLE is non-linear. Figure 3 presents scatter plots of the association between ΔLE and each of intergenerational transmissibility, mortality burden, and lifetime prevalence across different conditions. No single factor is strongly linearly associated with ΔLE. However, the simple product of a condition prevalence, transmissibility, and mortality burden captures much better the variation across conditions in life expectancy reduction associated with family history (Appendix Figure A7), showing the effects are approximately multiplicative.[8]

Intergenerational correlation in lifespan Table 2 presents the results of simulation-based estimates for this correlation in the lifespan—the realized length of life—and its association with chronic conditions from a simulated population of child-parent pairs. Panel A shows that the overall correlation when all chronic conditions are included is 1.325% (s.e. 0.033). Given that prior research estimates intergenerational lifespan correlations to be between 9% and 14% (Black et al., 2023), this finding implies that approximately 9–14% of

[8]Appendix B shows that in a simple example of condition arrival concentrated at a_0, they are exactly multiplicative.

the intergenerational correlation in mortality can be attributed to the transmission across generations of chronic conditions.

Panel B of Table 2 shows the (reduced) intergenerational correlation in the duration of life in scenarios when we neutralize the transmissibility across generations of one condition at a time (by setting $\beta_j = 0$). Consistent with Table 1, neoplasm, psychiatric, and cardiovascular conditions again emerge as the most substantial contributors to the intergenerational correlation in lifespan, as reflected in the fact that when they are neutralized, the correlation decreases sharply (by 24%, 21.9%, and 12.2%, respectively). These conditions are prevalent, transmissible across generations, and have a significant impact on mortality. Other conditions lead to more modest reduction in overall correlation when their transmissibility is zeored (9.4% for neurological conditions, 7.9% for gastrointestinal, and 4.8% for genitourinary). Other conditions result in even smaller reductions, suggesting they account for only a small part of the overall correlation in the lifespan.

Panel C of Table 2 shows the intergenerational correlation when we winsorize the maximal age of parents and children and recalculate the correlation. As we decrease the maximal age to 60, the correlation decreases by 56.5%, indicating that the contribution of chronic conditions to the intergenerational correlation in mortality accumulates late in the lifecycle. In contrast, the correlation captured before age 50 is minimal: only 15.4% of the total is realized by this age, and just 3.5% before age 40. These findings suggest that most of the intergenerational persistence in lifespan attributable to chronic conditions is driven by variation in mortality timing at older ages, rather than early-life health shocks.

5 Conclusion

We leverage comprehensive administrative claims data from Israel's largest healthcare system to explore the role of chronic conditions in intergenerational mortality correlation. By directly linking morbidity and mortality patterns across generations, this work contributes to the literature on intergenerational health mobility that typically relies on data with lim-

ited clinical information. Our use of comprehensive administrative healthcare records, which contain detailed information on chronic conditions, enables us to quantify the contribution of specific conditions to mortality correlation across generations. Our findings reveal that chronic conditions, specifically neoplasms (which include cancers), cardiovascular diseases, and psychiatric conditions, are significant contributors to the intergenerational transmission of mortality risk, together explaining more than half of the contribution of chronic conditions to the intergenerational correlation in mortality. Our results suggest that chronic conditions not only reduce life expectancy but also perpetuate health disparities across generations.

The findings have important implications for understanding health inequality and its persistence. The significant role of chronic conditions in explaining intergenerational mortality correlation suggests that public health interventions targeting prevalent and intergenerationally transmissible conditions could potentially reduce health disparities across generations. Interventions addressing the early detection and management of these conditions may yield benefits that extend beyond the current generation.

Our methodology, which combines hazard models for the onset of chronic conditions with models of subsequent mortality risk, provides a framework that could be extended to other settings where similar administrative data are available. We also provide detailed anonymized data on the morbidity and mortality hazards associated with different conditions (as supplementary materials), which future research can reuse and build on. Future research could build on this approach to explore additional determinants of intergenerational health mobility, such as the role of healthcare access, preventive care utilization, and specific health behaviors that might mediate the relationship between family health history and mortality outcomes. Future research can explore the interactions of intergenerational mobility in health and income, within and across countries. A limitation of our analysis is that we cannot fully disentangle genetic factors from shared environmental and behavioral determinants of chronic conditions. This represents an important area for future research.

References

Andersen, Carsten, "Intergenerational health mobility: Evidence from Danish registers," *Health Economics*, 2021, *30* (12), 3186–3202.

Beeton, Mary and Karl Pearson, "On the Inheritance of the Duration of Life, and on the Intensity of Natural Selection in Man.," *Biometrika*, 1901, *1* (1), 50–89.

Black, Sandra E, Neil Duzett, Adriana Lleras-Muney, Nolan G Pope, and Joseph Price, "Intergenerational correlations in longevity," Technical Report, National Bureau of Economic Research 2023.

_ **, Paul J Devereux et al.**, "Recent Developments in Intergenerational Mobility," *Handbook of Labor Economics*, 2011, *4*, 1487–1541.

Chang, Harrison, Timothy J Halliday, Ming-Jen Lin, and Bhashkar Mazumder, "Estimating intergenerational health transmission in Taiwan with administrative health records," *Journal of Public Economics*, 2024, *238*, 105194.

Christensen, Kaare, Mary K Wojczynski, Jacob K Pedersen, Lisbeth A Larsen, Susanne Kløjgaard, Axel Skytthe, Matt McGue, James W Vaupel, and Michael A Province, "Mechanisms underlying familial aggregation of exceptional health and survival: a three-generation cohort study," *Aging Cell*, 2020, *19* (10), e13228.

Currie, Janet, "Inequality at birth: Some causes and consequences," *American Economic Review*, 2011, *101* (3), 1–22.

_ **and Enrico Moretti**, "Biology as destiny? Short-and long-run determinants of intergenerational transmission of birth weight," *Journal of Labor economics*, 2007, *25* (2), 231–264.

Danesh, Kaveh, Jonathan T Kolstad, Johannes Spinnewijn, and William D Parker, "The chronic disease index: Analyzing health inequalities over the lifecycle," *NBER Working Paper No. 32577*, 2024.

Halliday, Timothy, Bhashkar Mazumder, and Ashley Wong, "Intergenerational mobility in self-reported health status in the US," *Journal of Public Economics*, 2021, *193*, 104307.

Jia, Gengjie, Yu Li, Hanxin Zhang, Ishanu Chattopadhyay, Anders Boeck Jensen, David R Blair, Lea Davis, Peter N Robinson, Torsten Dahlén, Søren Brunak et al., "Estimating heritability and genetic correlations from large health datasets in the absence of genetic data," *Nature Communications*, 2019, *10* (1), 5508.

Johnston, David W, Stefanie Schurer, and Michael A Shields, "Exploring the intergenerational persistence of mental health: Evidence from three generations," *Journal of Health Economics*, 2013, *32* (6), 1077–1089.

Lu, Frances and Tom Vogl, "Intergenerational persistence in child mortality," *American Economic Review: Insights*, 2023, *5* (1), 93–109.

Mahmood, Syed S, Daniel Levy, Ramachandran S Vasan, and Thomas J Wang, "The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective," *The Lancet*, 2014, *383* (9921), 999–1008.

Pearl, Raymond, "Studies on human longevity. IV. The inheritance of longevity. Preliminary report," *Human Biology*, 1931, *3* (2), 245.

Polubriaginof, Fernanda CG, Rami Vanguri, Kayla Quinnies, Gillian M Belbin, Alexandre Yahi, Hojjat Salmasian, Tal Lorberbaum, Victor Nwankwo, Li Li, Mark M Shervey et al., "Disease heritability inferred from familial relationships reported in medical records," *Cell*, 2018, *173* (7), 1692–1704.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot, *Demography: Measuring and Modeling Population Processes*, Blackwell Publishers, 2001.

Ruby, J Graham, Kevin M Wright, Kristin A Rand, Amir Kermany, Keith Noto, Don Curtis, Neal Varner, Daniel Garrigan, Dmitri Slinkov, Ilya Dorfman et al., "Estimates of the heritability of human longevity are substantially inflated due to assortative mating," *Genetics*, 2018, *210* (3), 1109–1124.

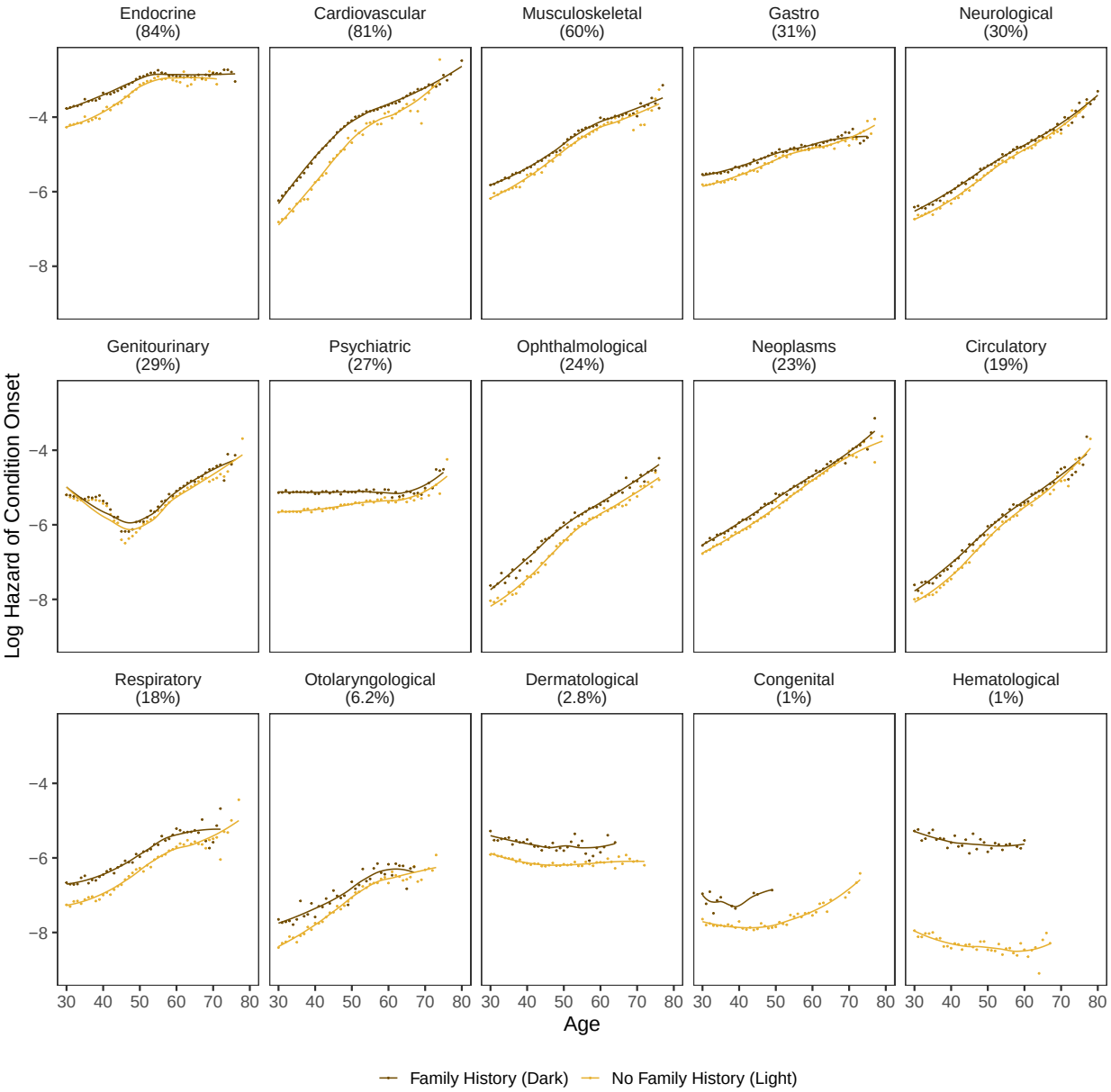
Slagboom, P Eline, Niels van den Berg, and Joris Deelen, "Phenome and genome based studies into human ageing and longevity: An overview," *Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease*, 2018, *1864* (9), 2742–2751.

Thompson, Owen, "Genetic mechanisms in the intergenerational transmission of health," *Journal of Health Economics*, 2014, *35*, 132–146.

_ , "Gene–Environment Interaction in the Intergenerational Transmission of Asthma," *Health Economics*, 2017, *26* (11), 1337–1352.

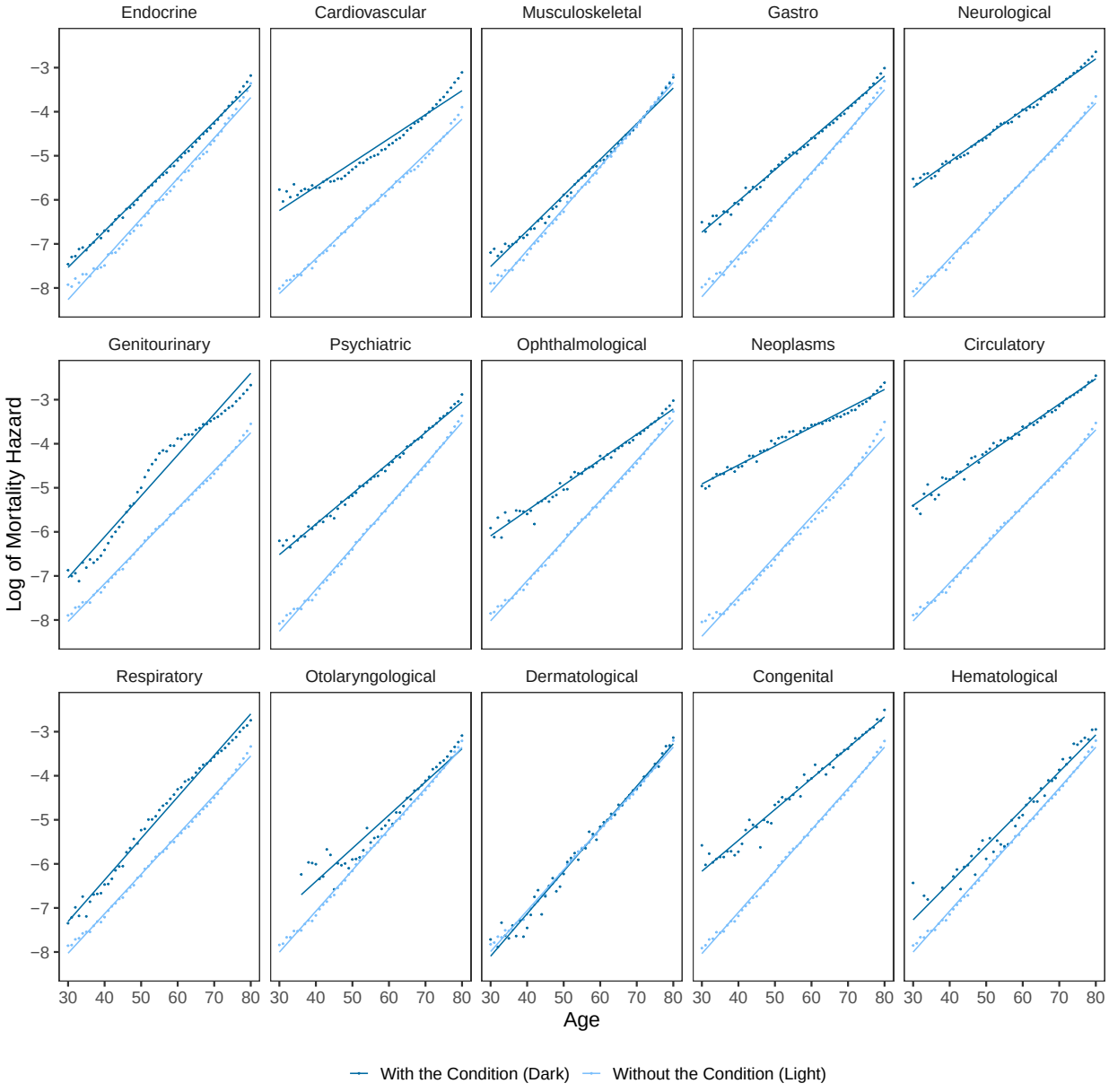
Vaupel, James W, "Inherited frailty and longevity," *Demography*, 1988, *25* (2), 277–287.

Figure 1: Morbidity Risk and Family History



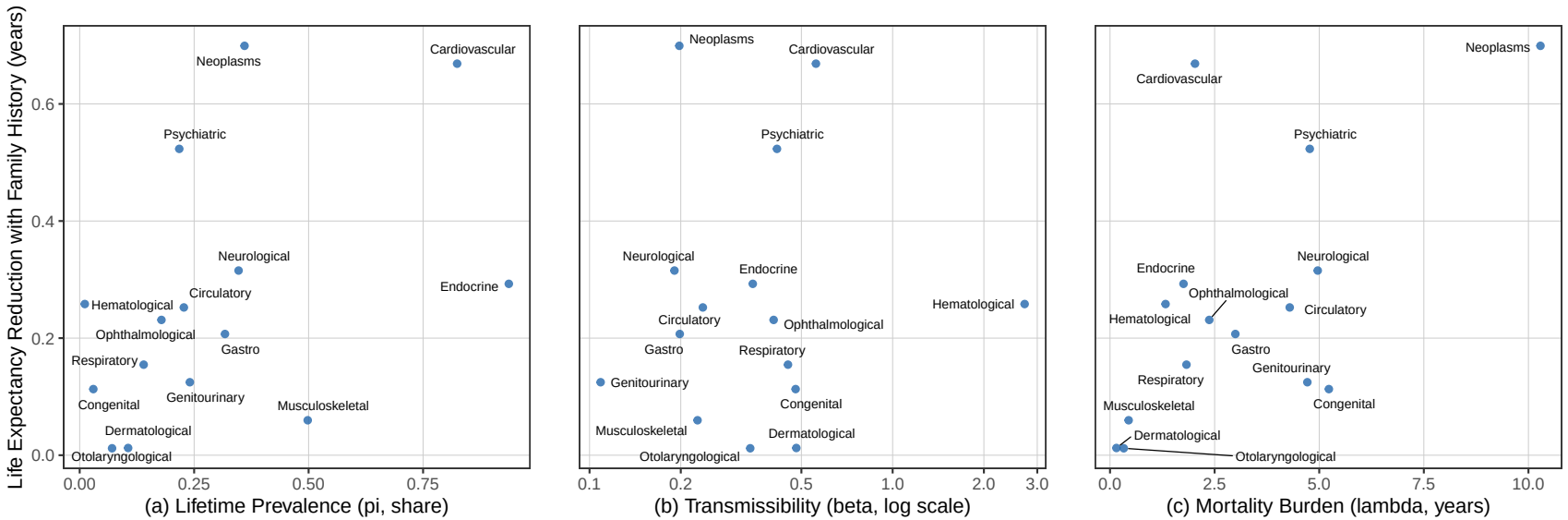
The figure shows nonparametric estimates of the hazard of chronic condition onset for individuals with and without family history of each condition. Each panel corresponds to a different condition group, with its sample prevalence shown in parentheses. We use the linked-generations sample to construct these estimates and report the sample prevalence of parents in the linked-generation sample at age 80 in parentheses. The x-axes show age in years; the y-axes show the log of the age-specific hazard of first diagnosis—defined as the probability of being diagnosed at a given age, conditional on surviving to that age without a prior diagnosis. Family history is defined as having at least one parent with the condition. Y-axis scales vary across panels. Estimates are overlaid with a loess smoother. Cells with fewer than 10 individuals are suppressed to preserve privacy.

Figure 2: Mortality Hazard by Chronic Condition



The figure shows nonparametric estimates of age-specific mortality rates for individuals with and without a diagnosis of each chronic condition. Each panel corresponds to a different condition group, sorted by decreasing order of prevalence. The x-axes show age in years; the y-axes show the log mortality hazard rate at each age, separately for individuals with and without a prior diagnosis of the condition at that age. Estimates are overlaid with a linear trend. Cells with fewer than 10 individuals are suppressed to preserve privacy.

Figure 3: Life Expectancy Reduction with Family History of Different Conditions against Different Condition Characteristics



The figure shows the association between life expectancy and family history for each condition plotted against relevant statistics. Each dot in the scatter plots corresponds to one chronic condition group. The vertical axis in all three panels is ΔLE_j, the reduction in life expectancy associated with a parental history of the condition. Panel A relates ΔLE_j to the condition's lifetime prevalence, π_j, defined as the share of individuals without parental history who would be diagnosed by age 80 in a cohort free of competing mortality risks. Panel B relates ΔLE_j to the intergenerational transmissibility, β_j, the log hazard ratio by which parental history multiplies a child's incidence hazard; this panel uses log scale for better legibility. Panel C plots ΔLE_j against the condition's mortality burden, λ_j, the difference in life expectancy between individuals who ever develop the condition by age 80 and those who never do. See Section 3 and Appendix B for detailed definitions. All quantities are computed from the lifecycle model in Section 3 using the non-parametric baseline hazards and the parameter estimates reported in Table 1, which also tabulates the data underlying this figure.

Table 1: Prevalence, Intergenerational Transmissibility, and Mortality Effects of Different Conditions

	Sample Estimates (log points)				Model-Based Calculations		
		Mortality Risk Change with the Condition					
	Morbidity Risk Change with Family History (β)	Change in Initial Risk (γ_0)	Change in Risk with Age (γ_1)		Mortality Burden (λ, years)	Lifetime Prevalence (π, population share)	Life Expectancy Gap with Family History (ΔLE, years)
	(1)	(2)	(3)		(4)	(5)	(6)
Neoplasms	0.198	5.813	−0.061	(<0.001)	10.29	0.360	0.70
	(0.009)	(0.032)					
Cardiovascular	0.558	2.016	−0.016	(0.001)	2.03	0.824	0.67
	(0.013)	(0.039)					
Psychiatric	0.415	2.536	−0.025	(<0.001)	4.77	0.218	0.52
	(0.008)	(0.034)					
Neurological	0.191	3.214	−0.027	(<0.001)	4.96	0.347	0.32
	(0.010)	(0.034)					
Endocrine	0.346	1.147	−0.010	(0.001)	1.76	0.937	0.29
	(0.009)	(0.038)					
Hematological	2.726	0.770	−0.004	(0.002)	1.33	0.011	0.26
	(0.034)	(0.127)					
Circulatory	0.236	3.513	−0.030	(0.001)	4.29	0.228	0.25
	(0.014)	(0.042)					
Ophthalmological	0.405	2.826	−0.032	(0.001)	2.37	0.179	0.23
	(0.015)	(0.051)					
Gastro	0.198	2.227	−0.024	(<0.001)	3.00	0.317	0.21
	(0.008)	(0.034)					
Respiratory	0.452	1.500	−0.010	(0.001)	1.83	0.140	0.15
	(0.014)	(0.039)					
Genitourinary	0.109	2.325	−0.017	(0.001)	4.72	0.241	0.12
	(0.009)	(0.039)					
Congenital	0.479	2.729	−0.024	(0.001)	5.23	0.030	0.11
	(0.064)	(0.085)					
Musculoskeletal	0.227	0.647	−0.007	(0.001)	0.45	0.498	0.06
	(0.009)	(0.035)					
Dermatological	0.481	0.072	0.001	(0.001)	0.15	0.106	0.01
	(0.019)	(0.091)					
Otolaryngological	0.339	0.520	−0.005	(0.001)	0.33	0.071	0.01
	(0.028)	(0.097)					

This table reports parameter estimates and model-based calculations for different chronic-condition groups. Columns 1–3 show Cox proportional hazard estimates associated with each group and their standard errors in parentheses. Column 1 shows β, the estimated change in each condition incidence hazard associated with family history. The parameter β_j is estimated using the linked-generations sample. Columns 2 and 3 report the change in mortality hazard associated with each condition at age 30 and with years of age, estimated using the general sample. Columns 4–6 report magnitudes calculated using the estimated model. Column 4 shows λ_j, the difference in life expectancy between individuals who ever develop the condition by age 80 and those who never do, conditional on no parental history. Column 5 shows π_j, the lifetime prevalence of the condition—the probability that an individual without parental history would be diagnosed by age 80 absent competing mortality risks. Column 6 presents ΔLE_j, the reduction in life expectancy (years) associated with havingfamily history of the condition. Rows are sorted by Column 6. See Section 3 for detailed definitions.

Table 2: Simulated Intergenerational Correlation in Lifespan

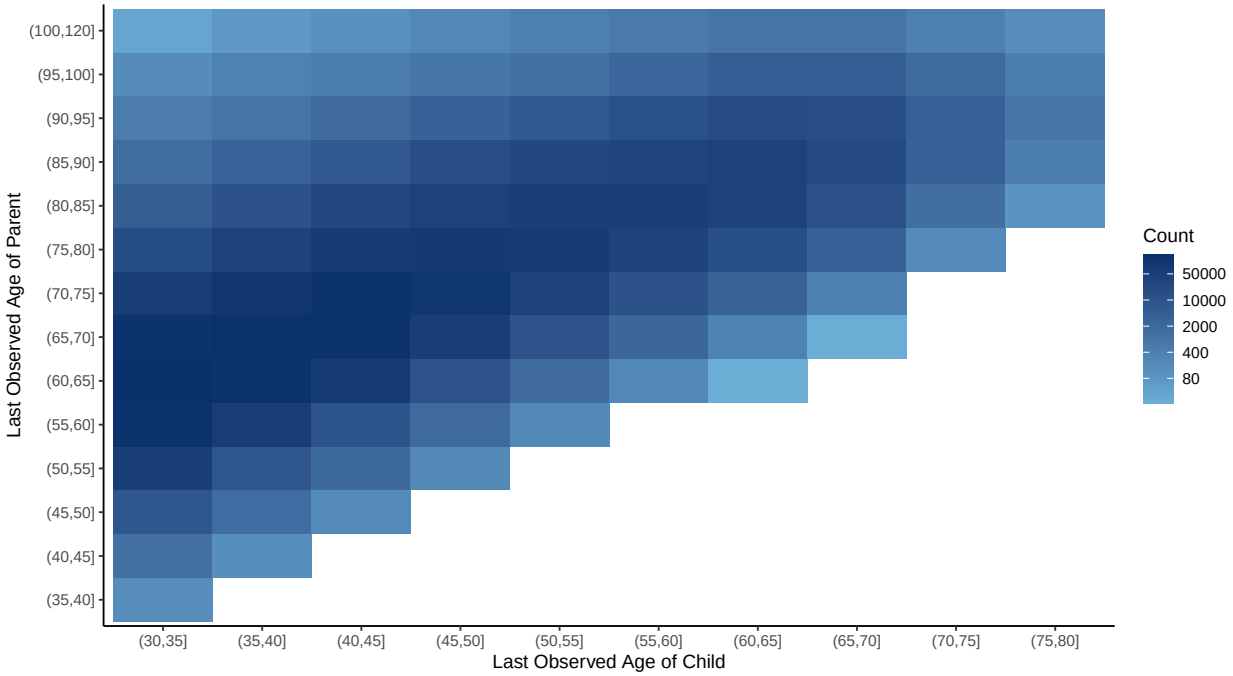
	Corr. (%)	SE	Reduction in Corr. (% of Baseline)
	(1)	(2)	(3)
A. All Conditions	1.325	(0.033)	baseline
B. Muting the Intergenerational Transmissibility of:			
Neoplasms	1.007	(0.030)	24.0%
Psychiatric	1.035	(0.031)	21.9%
Cardiovascular	1.164	(0.031)	12.2%
Neurological	1.201	(0.033)	9.4%
Gastro	1.220	(0.032)	7.9%
Genitourinary	1.262	(0.033)	4.8%
Endocrine	1.270	(0.031)	4.2%
Circulatory	1.287	(0.031)	2.9%
Ophthalmological	1.296	(0.030)	2.2%
Respiratory	1.302	(0.031)	1.8%
Musculoskeletal	1.315	(0.036)	0.8%
Congenital	1.319	(0.033)	0.5%
Hematological	1.320	(0.031)	0.4%
Otolaryngological	1.325	(0.030)	0.0%
Dermatological	1.325	(0.033)	0.0%
C: Winsorizing Lifespan at:			
Max Age = 70	1.042	(0.028)	21.4%
Max Age = 60	0.577	(0.031)	56.4%
Max Age = 50	0.204	(0.031)	84.6%
Max Age = 40	0.047	(0.030)	96.4%

This table shows correlations in mortality from simulated parent-child pairs using the estimated life cycle model described in Section 3 (N=10^7 pairs). For each parent-child pair, we simulate the parent's chronic condition onsets over their life course and draw the parent's age at death from the resulting mortality hazard, which increases with each condition onset. We then assign each parent's realized condition history to their child, simulate the child's condition onsets from morbidity hazards that are augmented by their family history, and draw the child's age at death from its mortality hazard. Panel A shows results when all conditions are included (the baseline scenario). Column (1) shows the intergenerational correlation in lifespan as a percentage (i.e., multiplied by 100). Column (2) shows standard errors in parentheses, calculated from 200 bootstrap replications. Panel B shows results of simulations where the intergenerational impact of one condition at a time is excluded (by setting $\beta_j = 0$ while keeping its direct impact on mortality). Panel C shows results where the simulated lifespans from Panel A are winsorized at different ages for both children and parents. For Panels B and C, Column (3) shows the percentage reduction in correlation in each scenario relative to the baseline scenario. In Panel B, reductions do not sum to 100% because the model involves non-linear interactions across conditions through survival; hence, the reduction is not a decomposition of the total.

On-Line Appendix for Intergenerational Correlation in Mortality: The Role of Chronic Conditions

Appendix Figures and Tables

Appendix Figure A1: Joint Age Distribution of Linked Generations

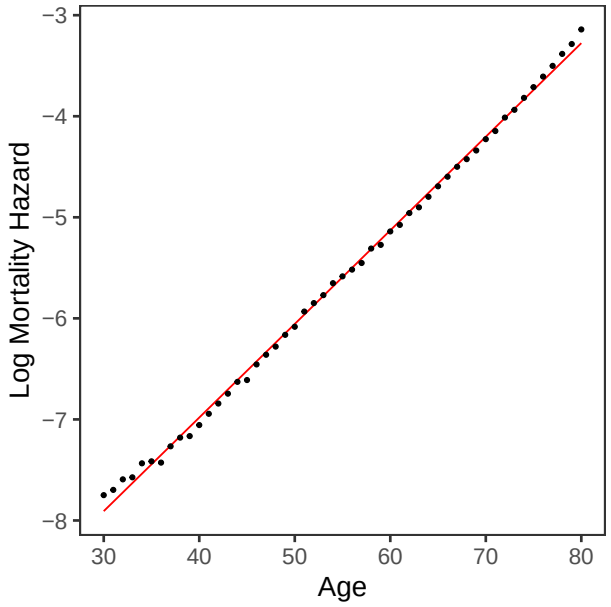


The figure presents a heatmap representing the joint distribution of the last observed ages of children and their parents in our sample. Each cell on the heatmap corresponds to a pair of 5-year age bins. To address privacy concerns, cells with fewer than 15 individuals are not plotted, and ages above 100 are categorized into broader bins. Note that the last observed ages might differ from actual age differences due to differing observation years. For instance, a child born in 1945 and last observed in 2018 at age 73, with a parent alive from 1920 to 2010 (last observed in 2010), would have a recorded age difference of 17 years in our data, which is smaller than their actual age difference of 25 years.

Appendix Figure A2: Intergenerational Relationship in Morbidity and Mortality Risk

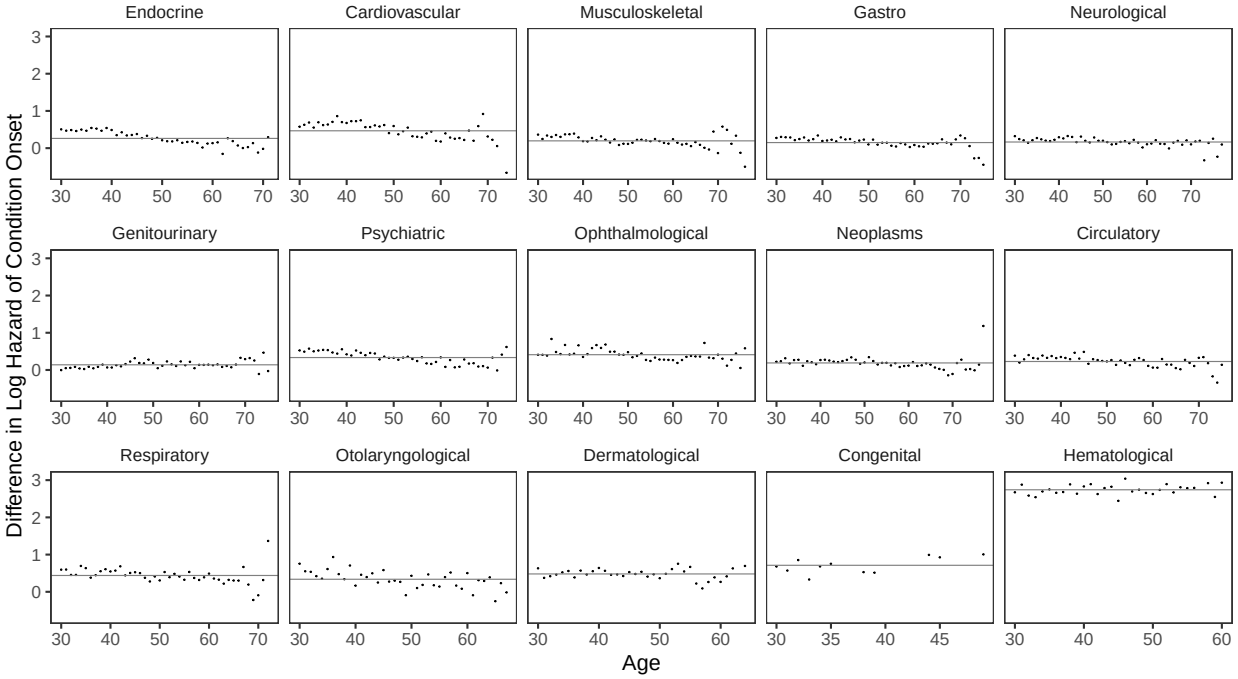
The figure shows an illustration of the key elements of our model of intergenerational correlation in mortality risk driven by intergenerational correlation in morbidity risk. Panel A shows the rate of onset of a chronic condition j, which increases proportionally at a constant rate β with family history of the condition. Panel B shows the mortality hazard rate, which increases upon the condition's onset by a rate γ_j, which is linear in age. The baseline age-specific hazards for condition onset and mortality, depicted as linear in age, are unconstrained by the model and estimated non-parametrically for population data.

Appendix Figure A3: Mortality Hazard



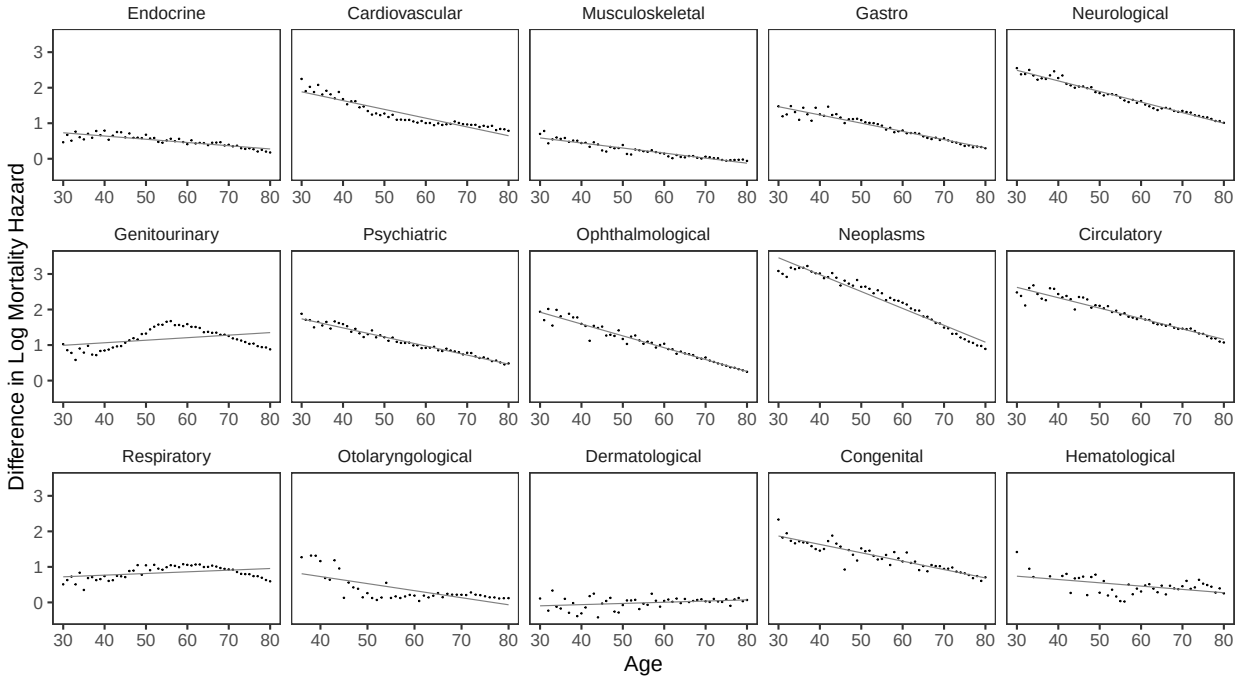
The figure shows nonparametric estimates of age-specific mortality rates for the full population in the general sample. The x-axes show age in years; the y-axes show the log mortality hazard rate at each age. Estimates are overlaid with a linear fit line.

Appendix Figure A4: Increase in Morbidity Hazard with Family History, by Condition



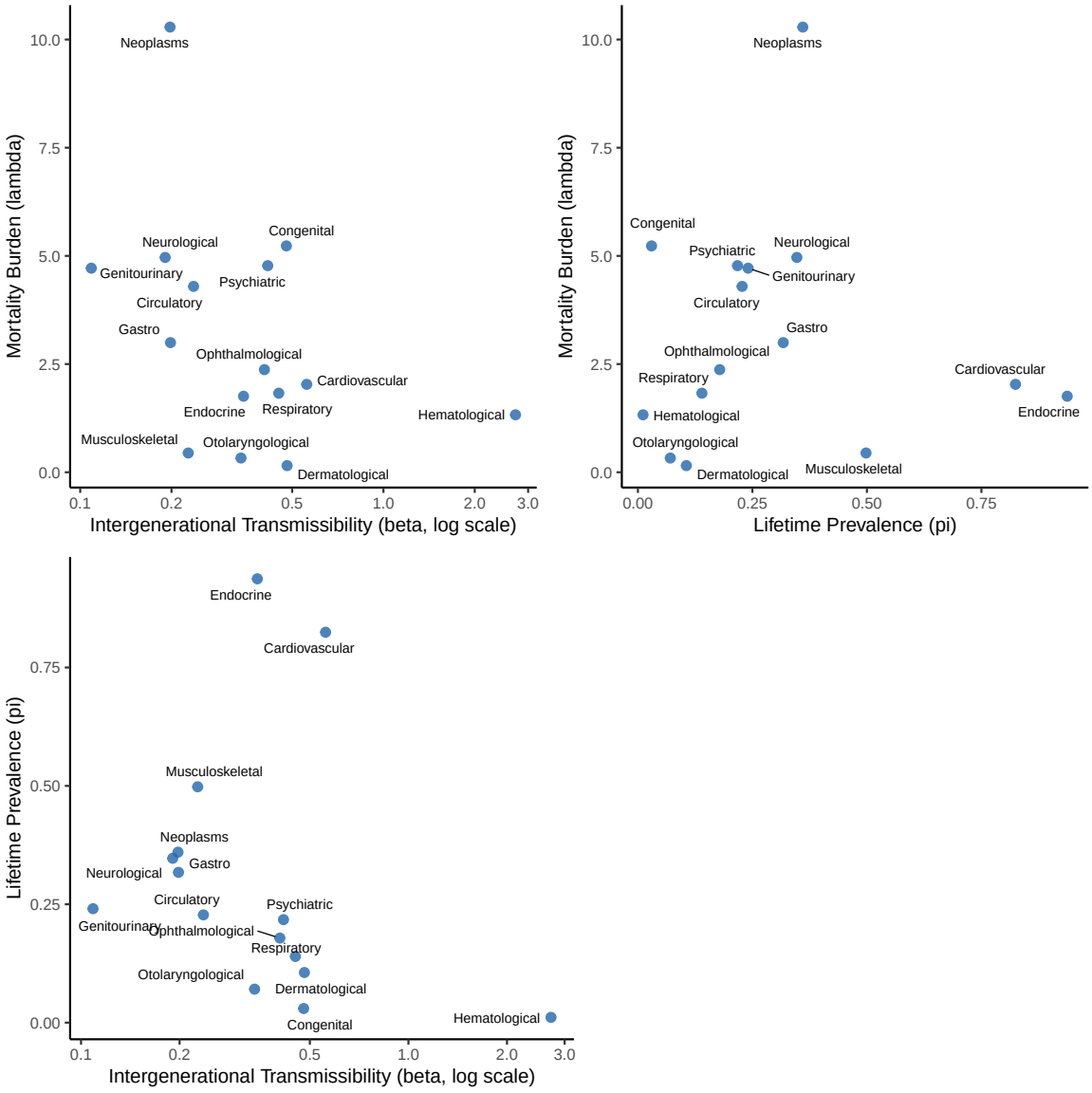
The figure shows the difference in age-specific chronic condition onset hazard rates between individuals with and without family history of each condition. Each panel corresponds to a different condition group. The x-axis shows age in years; the y-axis shows the difference in the log hazard of first diagnosis at each age. Positive values indicate higher morbidity risk among individuals with family history of the condition. The points in each panel are overlaid with their mean, roughly corresponding to the parameterization of these data in our model. Cells with fewer than 10 individuals are suppressed to preserve privacy.

Appendix Figure A5: Increase in Mortality Hazard with Condition Onset, by Condition



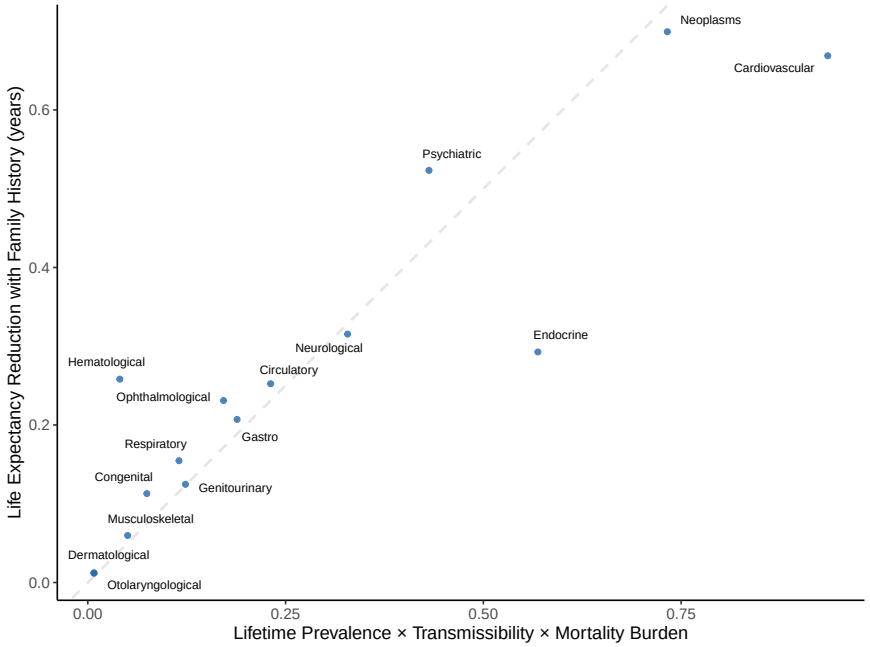
The figure shows the difference in the log of age-specific mortality rates between individuals with and without each chronic condition. Each panel corresponds to a different condition group. The x-axis shows age in years; the y-axis shows the difference in log mortality hazard. Positive values indicate higher mortality risk for individuals diagnosed with the condition. The points in each panel are overlaid with a linear best fit, roughly corresponding to the parameterization of these data in our model. Cells with fewer than 10 individuals are suppressed to preserve privacy.

Appendix Figure A6: Associations Between Condition Characteristics



The figure shows scatter plots of pairs of the three condition characteristics: intergenerational transmissibility (β), lifetime prevalence (π), and mortality burden (λ). Underlying data are tabulated in Table 1. Transmissibility is directly estimated from the linked-generation sample. Lifetime prevalence and mortality burden are calculated using the estimated lifecylce model. See Section 3 for model specifications and estimation details and AppendixB for detailed definitions.

Appendix Figure A7: Life Expectancy Reduction with Family History of Different Conditions and the Product of its Driving Factors



The figure shows a scatter of ΔLE_j, the difference in life expectancy associated with family history of each condition group against the product of three quantities contributing to it: lifetime prevalence (π_j), intergenerational transmissibility β_j, and mortality burden (λ_j), and a dashed 45-degree line from the origin. The units of both axes are years. All quantities are calculated using the estimated model. See Section 3 and Appendix B for detailed definitions.

Appendix Table A1: Chronic Condition Groupings

Condition Group	Specific Conditions
Endocrine	Hyperlipidemia, Obesity, Diabetes, Hypothyroidism, Hyperthyroidism, Hypo/Hyperparathyroidism, Hyperprolactinemia, Diabetes Insipidus, Addison's Disease, Cushing's Disease, Acromegaly
Cardiovascular	Hypertension, IHD, Arrhythmia, Valvular Cardiac Dis (excl. MVP), CHF, PVD, Carotid Artery Disease, Pulmonary Hypertension, Cardiomyopathy, Aortic Aneurysm, IHSS
Musculoskeletal	Arthropathy, Osteoporosis, Joint Replacement, Gout, Rheumatoid Arthritis, Polymyalgia Rheumatica, SLE, Scleroderma
Neoplasms	Malignancy
Neurological	Other Neurological Disease, Dementia / Alzheimer's / OMS, Epilepsy, Parkinson's Disease, Cerebral Palsy, Multiple Sclerosis, Motor Neuron Disease, Myasthenia Gravis, Benign Brain Tumor, Hereditary Neurological Disease
Gastro	Reflux Esophagitis / Gastritis / Deudenitis, Peptic Ulcer, Other Liver Disease, Irritable Bowel Syndrome, Celiac Disease, Crohn's Disease, Hepatitis B Carrier, Hepatitis C Carrier, Chronic Act/Per Hepatitis, Cirrhosis, Ulcerative Colitis, Gastrostomy
Genitourinary	Other kidney Disease, Chronic Renal Failure, Infertility Male/Female, Prostatic Hypertrophy, Dialysis, Kidney Transplant, Endometriosis
Psychiatric	Depression, Anxiety, Psychoses, Neuroses, Schizophrenia, Bipolar Disease
Circulatory	s/p CVA, s/p Head of Femur Fracture, s/p Pneumothorax, s/p splenectomy
Ophthalmological	Retinopathy, Glaucoma, Blindness, Retinitis Pigmentosum, Retinal vein occlusion
Respiratory	Asthma, COPD, Chronic Bronchitis, Bronchiectasis, Other pulmonary disease, Tuberculosis s/p, Tuberculosis
Dermatological	Atopic dermatitis, Psoriasis, Non-melanoma skin cancers, Hidradenitis Suppurativa
Otolaryngological	Deafness
Other	Amputation of Limb, Chronic Ventilation, Tracheostomy, Hypophysary Adenoma, Amyloidosis, Pemphigus, G-6-P-D Deficiency, Behcets Disease, Cystic Fibrosis, Other hereditary diseases, Wilson's Disease, Gaucher Disease, Mitochondrial diseases, Organ Transplantation, Sarcoidosis
Congenital	Congenital Anomalies, Mental Retardation (including Down Syndrome), Autism, Familial Mediterranean Fever
Hematological	Other Hematologic Dis (excluding Iron Def Anemia), Thalassemia, Sickle Cell Anemia, Hemophilia, ITP, Pernicious Anemia

This table reports the specific chronic conditions associated with each chronic condition group (hereafter referred to as Condition).

Appendix Table A2: Chronic Condition Prevalence

	General Sample	Linked-Generations Sample	
		Parent	Child
	(1)	(2)	(3)
Endocrine	0.799	0.843	0.900
Cardiovascular	0.789	0.811	0.743
Musculoskeletal	0.563	0.604	0.613
Gastro	0.298	0.312	0.324
Neurological	0.272	0.302	0.317
Psychiatric	0.265	0.269	0.265
Genitourinary	0.257	0.288	0.276
Ophthalmological	0.230	0.242	0.235
Neoplasms	0.221	0.229	0.262
Circulatory	0.175	0.194	0.146
Respiratory	0.164	0.182	0.150
Otolaryngological	0.053	0.062	0.048
Dermatological	0.024	0.028	0.049
Congenital	0.009	0.010	0.020
Hematological	0.008	0.010	–
Observations (80 year old individuals)	403,637	223,580	608

This table reports the empirical prevalence, at age 80, of each chronic condition group. To calculate this, we take the number of individuals in each sample who have been diagnosed with chronic condition j by age 80 divided by the total number of individuals in each sample. Columns show this prevalence in different samples: all members in the general sample (1), and the parents (2) and children (3) in the linked-generations sample. Cells with fewer than 10 individuals are suppressed to preserve privacy.

A Estimation Specifications

We estimate the two key model parameters using Cox proportional-hazards models. First, to estimate the impact of each condition on mortality (equation 4), we run separate regressions for each condition j using the specification:

$$m_i(a) = m_0(a) \exp\left(\gamma_j^0 D_{ij}(a) + \gamma_j^1 D_{ij}(a) \times (a - a_0)\right), \tag{7}$$

where $m_0(a)$ is the baseline mortality hazard (left unspecified and stratified by gender), and $D_{ij}(a) = \mathbb{1}(O_{ij} \leq a)$ is an indicator for whether individual i has been diagnosed with condition j by age a (i.e., the condition's onset occurred at or before age a). For each condition j, the coefficients γ_j^0 and γ_j^1 respectively capture the condition's effect on mortality risk at a_0 and the change of this effect with age. We estimate these individual condition models using the general sample.

Second, to estimate the impact of family history on condition onset (β_j from equation 5), we specify:

$$h_{ij}(a) = h_{j0}(a) \exp(\beta_j F_{ij}), \tag{8}$$

where $h_{j0}(a)$ is the baseline hazard function (left unspecified and stratified by gender), and F_{ij} is the indicator for family history of condition j. The parameter of interest, β_j, captures the log hazard ratio associated with family history. We estimate these regressions using the linked-generations sample.

B Model-Based Calculations

Life Expectancy Reduction Associated with Family History of Each Condition. For each condition j and family history status $z \in \{0, 1\}$,

$$\begin{aligned} E\big[T_i \mid F_{ij} = z\big] = &\sum_{o=a_0}^{\bar{a}} \Pr\big(O_{ij} = o \mid F_{ij} = z\big)\, E[T_i \mid O_{ij} = o, F_{ij} = z] \\ &+ \Pr\big(O_{ij} = \infty \mid F_{ij} = z\big)\, E[T_i \mid O_{ij} = \infty, F_{ij} = z], \end{aligned}$$

where $E[T_i \mid O_{ij} = o]$ is the expected lifespan of an individual who is first diagnosed with j exactly at age o:

$$E\big[T_i \mid O_{ij} = o\big] \;=\; \sum_{a=a_0}^{\bar{a}} a \cdot m(a|o) \prod_{t=a_0}^{a-1} \big(1 - m(t|o)\big),$$

and where $m(a|o)$ denotes the mortality hazard at age a given that condition j had onset at age o. This is a specialized version of our general mortality hazard notation $m(a, G_i(a))$, where we're focusing specifically on the impact of condition j with onset at age o while averaging over all other possible conditions in $G_i(a)$. Formally, $m(a|o) = E[m(a, G_i(a))|O_{ij} = o]$, where the expectation is taken over all possible realizations of other conditions (besides j) that could be in $G_i(a)$.

In the simple case of a single conditions for which $h_j(a) = \pi_j$ for $a = a_0$ and zero for $a > a_0$, and letting $L_0 = E[T_i|O_{ij} = \infty]$, the difference ΔLE_j simplifies to $(L_0 - \lambda_j\pi_j) - (L_0 - \lambda_j\pi_j e^{\beta_j}) = \pi_j(e^{\beta_j} - 1)\lambda_j \approx \pi_j\beta_j\lambda_j$ (when β_j is small). That is, the life expectancy reduction associated with the condition is the product of its lifetime prevalence, intergenerational transmissibility, and mortality burden.

This calculation is performed independently for each condition, using the empirical distribution of baseline morbidity and mortality hazard rates observed in the data: $\hat{h}_j(a)$, the morbidity hazard without family history, and $\hat{m}(a|o > a)$, the age-specific mortality rate for individuals without condition j. We then use the estimated parameters (β_j and γ_j) and the functional form specified by the model for the elevated morbidity with family history and mortality after the onset of each condition.

Mortality Burden of Different Conditions. For each condition j, we define the mortality burden life expectancy reduction associated with the condition

$$\lambda_j = E[T_i \mid O_{ij} \leq \bar{a}] - E[T_i \mid O_{ij} = \infty]. \tag{9}$$

where the first term integrates $\sum_{o=a_0}^{\bar{a}} \big(\Pr(O_{ij} = o)E[T_i|O_{ij} = o]\big)$ using the same $E[T_i \mid O_{ij} = o]$ as above.

Hypothetical prevalence. We compute the hypothetical prevalence of condition j at the maximal age $\bar{a}$ in a population of individuals with no family history of j who are not subject to mortality. Define $\pi_j = \Pr(O_{ij} \leq \bar{a} \mid F_{ij} = 0)$. With baseline incidence hazards $h_j(a)$,

$$\pi_j = 1 - \prod_{a=a_0}^{\bar{a}} \big(1 - h_j(a)\big),$$

We use the non-parametric estimates for $\hat{h}_j(a)$. This hypothetical measure of prevalence compares abstracts from survival bias and competing risks, to more directly compare the cumulative hazards of different conditions.