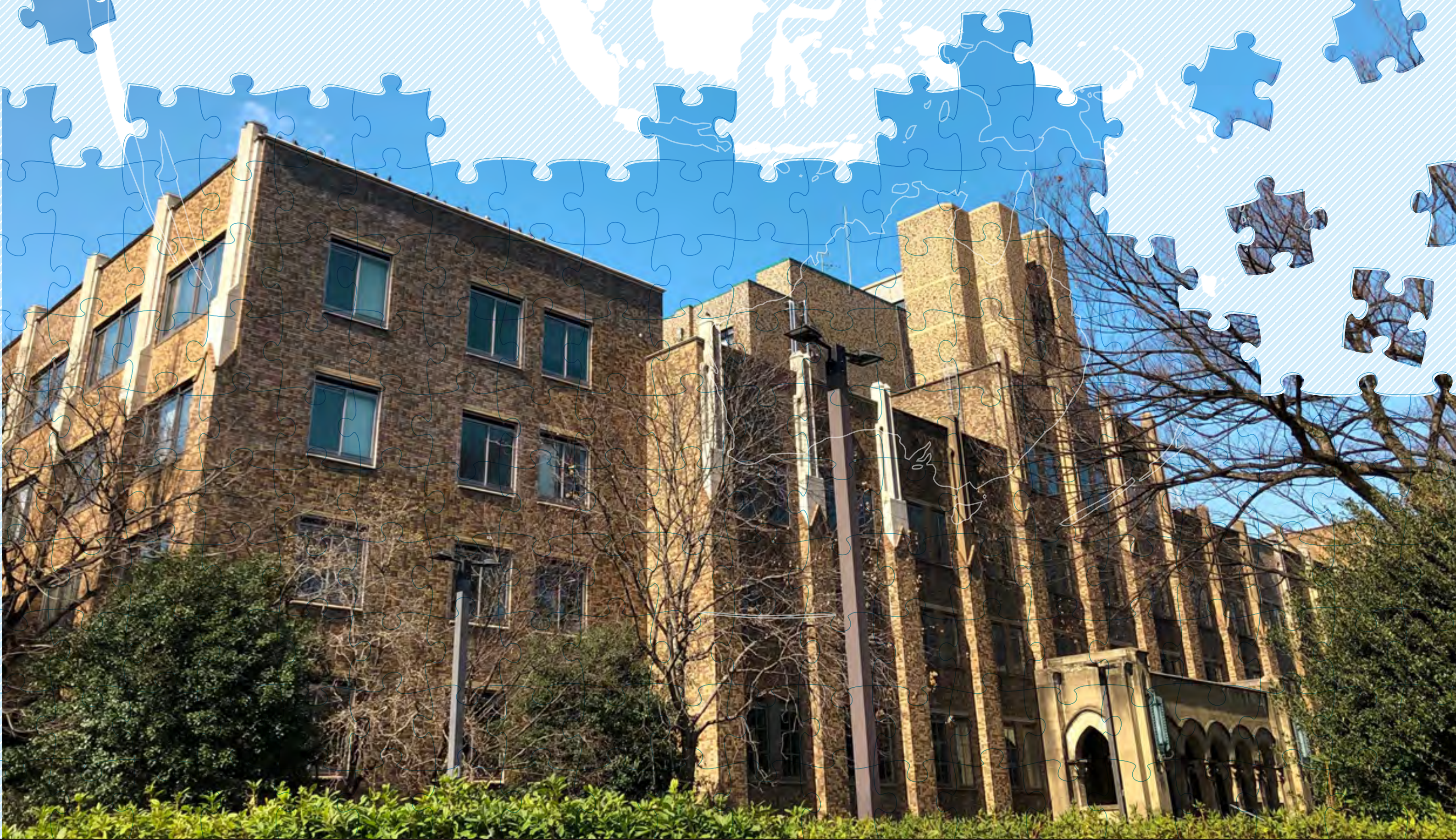
CSRDA
Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo
SUSTAINABLE DEVELOPMENT GOALS
CSRDA supports the Sustainable Development Goals

CSRDA Discussion Paper

Stepping-down or Skydiving?: A Feature-wide Effect Heterogeneity of Retirement on Subsequent Health



No. 124

Date July.2025

SDGs
3 GOOD HEALTH AND WELL-BEING
8 DECENT WORK AND ECONOMIC GROWTH
10 REDUCED INEQUALITIES

Name Shingo Nitta

Stepping-down or Skydiving?

A Feature-wide Effect Heterogeneity of Retirement on Subsequent Health

Shingo Nitta

2025-07-30

This study investigates how retirement's effect on subsequent health varies by individual attributes and their interaction. Instead of spotlighting heterogeneity for a specific, theoretically predetermined feature, I consider the feature-wide effect heterogeneity to find possible data-driven candidates that generate this heterogeneity. To accomplish this, I use causal forests and harmonized data from the Health and Retirement Studies family. The results show that temporal dimensions (such as age and birth cohort) and financial dimensions (such as income and wealth) are the most important features in heterogenizing retirement's effect on health. Based on these findings, I suggest the "double scaffold" model to comprehensively explain feature-wide effect heterogeneity.

Acknowledgments

A replication package is available at github.com/sittaningo/replication/tree/main/rethet. When the paper is accepted, the replication package in gitub will be archived on Dataverse. This analysis uses data or information from the Harmonized HRS, MHAS, SHARE, ELSA, CHARLS, and KLoSA dataset and Codebook, Version D for HRS, Version G for SHARE, Version G.3 for ELSA, Version D for CHARLS, Version E.2 for KLoSA, and Version C.2 for MHAS as of 7th November, 2024 developed by the Gateway to Global Aging Data. The development of the Harmonized HRS, MHAS, SHARE, ELSA, CHARLS, and KLoSA dataset and Codebook were funded by the National Institute on Aging (R01 AG030153, RC2 AG036619, 1R03AG043052). For more information, please refer to g2aging.org. I thank Ryota Mugiyama, Wataru Yoshida, Aguru Ishibashi, Aram Kwon, Kazuko Sano, and Jiajie Zhang, participants in Japanese Association of Mathematical Sociology 78th Conference for helpful feedbacks.

Funding

This work was supported by JSPS KAKENHI Grant Numbers JP24KJ1919.

Introduction

Divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.

—— René Descartes. *Discours de la méthode/A Discourse on Method* (1637/1957: 15)

Extending working lives is a growing concern for policymakers aiming to sustain social welfare as the life expectancies rapidly increase. Simultaneously, some suggests that extending working lives worsens the health of older workers (Taylor 2019). Examining how retirement affects subsequent health is thus crucial for policymakers. If retirement *worsens* subsequent health, extending working lives and health promotion are mutually exclusive, and policymakers need to choose between them. If retirement *improves* subsequent health, policies extending working lives can result in better health, thereby contributing to a sustainable care system.

Both theoretical expectations and empirical evidence are, however, mixed. The health capital model (Grossman 1972), disengagement theory (Hochschild 1975) and activity theory (Havighurst 1961) derive competing hypotheses on the positive and negative effect of retirement on subsequent health. Echoing this theoretical inconsistency, empirical evidences are also mixed. Positive effects (e.g., Zhao, Konishi, and Noguchi 2017; Oshio and Kan 2017), negative effects (e.g., Bloemen, Hochguertel, and Zweerink 2017; Xu 2023; Furuya and Fletcher 2024) and null effects (e.g., Grøtting and Lillebø 2020; Gorry and Slavov 2021) of retirement on health have been reported.

One reason why evidence is mixed is *heterogeneity*; the effect of retirement on health differs by subgroups, especially those defined by social determinants. Previous studies have examined whether the effect of retirement on health differs by gender (Quaade et al. 2002; McDonough et al. 2017; Zhao, Konishi, and Noguchi 2017; Xu 2023; Oshio and Kan 2017; Furuya and Fletcher 2024), timing of retirement (Hult et al. 2010; Calvo, Sarkisian, and Tamborini 2013; Barban et al. 2020; Mizuochi 2023), job before retirement (Bogaard, Henkens, and Kalmijn 2016; Bloemen, Hochguertel, and Zweerink 2017; Okamoto, Okamura, and Komamura 2018; Qvist 2022; Grøtting and Lillebø 2020; Scotti 2022) and social networks (Pilehvari, You, and Lin 2023). Although they undoubtedly advance theory and policy by delving into the effect heterogeneity, they still suffer from mixed results.

To further clarify the mixed evidence even after considering the effect heterogeneity, I apply a *feature-wide effect heterogeneity* approach. This approach identifies the comprehensive effect heterogeneity across various social determinants, including not only sociodemographic status such as gender and education but also work and family characteristics, and their interaction. Instead of spotlighting the specific feature deduced by theoretical prior, I explore the pattern of partitioning among the features and estimate the effect of retirement on health in each partition in a data-driven way. To examine feature-wide effect heterogeneity, I use causal forests (Athey and Imbens 2016) with the global harmonized data which provides a large sample size. Causal forests maximize predictive power by partitioning data and estimate the effect within each partition.

Feature-wide effect heterogeneity advances both theory and policy of retirement and subsequent health. Theoretically, I localize theory how retirement affects subsequent health in a more disaggregated fashion. Previous studies attempt to localize the effect of retirement on health by one feature expected to be theoretically important, but they simultaneously assume that the feature does not further generate heterogeneity or interact with other features. Feature-wide effect heterogeneity approach relaxes the assumption and explores the effect heterogeneity in a more flexible way. It speaks to policymakers to pivot interests to differences in effects based on various social determinants. Feature-wide effect heterogeneity approach optimizes the scope of intervention and the allocation of resources to gauge the effect heterogeneity, instead of averaging the intervention. For example, if positive effect of retirement on health is observed for high income retirees (Okamoto, Kobayashi, and Komamura 2022), policymakers can target only the low income workers to improve their health.

From the feature-wide effect heterogeneity approach, I newly propose the "double scaffold" model to comprehensively explain how the consequence of retirement on health is modified by individual attributes. The double scaffold model consists of two conditioning factors for retirement-health dynamics: the temporal scaffold and the physical scaffold. The temporal scaffold refers to the temporal dimension of the effect of retirement on health. Retirees of younger cohorts, younger ages and those retiring during recession years are exposed to further negative effects of retirement on health. The physical scaffold refers to financial circumstances and work characteristics. The health of retirees with higher income, lower wealth and longer working hours worsens more than that of their counterparts. The temporal scaffold first sets their initial condition similar to altitude, and physical scaffold further adds to the effect heterogeneity like a ladder. The distance from sea level is determined by the temporal scaffold and the physical scaffold determines the effect heterogeneity.

The contributions in this study are threefold. First and most importantly, I investigate the feature-wide effect heterogeneity of retirement on health. I further disaggregate the effect heterogeneity into various social determinants and within-determinant relationships. In doing so, clearer and specific explanations, which I call the "double scaffold", are provided, enhancing theoretical plausibility and feasibility. Second, to accomplish this, the exploration in this study is formalized by using causal forests, a machine learning method which collapses the evaluation criteria into predictive power. Causal forests aim to maximize predictive power in each partition while clarifying the partitioning process (Brand et al. 2021), thereby achieving interpretability and empirical plausibility (Ludwig and Mullainathan 2024). To obtain a sufficient sample size, I use harmonized data of HRS family and overcome one problem of causal forests; large sample size (Jawadekar et al. 2023). Third, I contribute to policy by identifying the responsive groups with whom policymakers should intervene. How to allocate the limited resources to maximize the outcome of intervention is a significant issue in the unequal society. By taking into account the partitioning and the effect heterogeneity of each partitioning into account, policymakers optimally allocate resources to those who are most likely to benefit from the intervention (Brand et al. 2021).

Theoretical Background and Previous Studies

The senescence process is a key to demographic shifts (Harris and Hummer 2023). It is not independent of contextual factors, but rather intertwines with them. One modifiable contextual factor is retirement, an exit from the labor market (Robert C. Atchley 1982). Retirement is a major transition in the life course, making it a worthwhile inquiry for demography to understand how health changes after retirement (Kuh et al. 2003). Several theories predict the effect of retirement on health in varied and sometimes contradictory ways.

The health capital model (Grossman 1972) views health as a capital sensitive to daily health investment and life transitions. Retirement decreases the opportunity costs of health investment, and thus retirement is expected to improve health. Contrary to this expectation, health capital model also suggests retirement can worsen health because retirees no longer have the benefits of investigating health, which is a linchpin of maximizing lifetime earnings (Dave, Rashad, and Spasojevic 2008). Hochschild (1975) conceptualized retirement as disengagement from past roles, whereas other scholars regarded it as active involvement in new roles such as grandparenthood and volunteering (Havighurst 1961; see also Hochman and Lewin-Epstein 2013; Hämäläinen et al. 2024). Life course literature suggests that retirees are more likely to maintain their roles irrespective of retirement (Robert C. Atchley 1971; Kim and Moen 2002). There are competing and opposite expectations on the effect of retirement on health.

Aligning with multiple theories, previous studies have reported mixed results for the effect of retirement on health: positive effects (e.g., Zhao, Konishi, and Noguchi 2017), negative effects (e.g., Bloemen, Hochguertel, and Zweerink 2017), and null effects (e.g., Grøtting and Lillebø 2020). Even in novel (quasi-)experimental designs like pension reform, effects are not consistent; Bloemen, Hochguertel, and Zweerink (2017), for example, reported the negative effect, whereas Zhao, Konishi, and Noguchi (2017) reported the positive effect.

One strong possibility for a consistent explanation is effect heterogeneity by subgroups. A growing body of studies has examined how the effect differs across subgroups, especially across gender (Quaade et al. 2002; McDonough et al. 2017; Zhao, Konishi, and Noguchi 2017; Xu 2023; Oshio and Kan 2017; Furuya and Fletcher 2024), timing of retirement (Hult et al. 2010; Calvo, Sarkisian, and Tamborini 2013; Barban et al. 2020; Mizuochi 2023), job before retirement (Bogaard, Henkens, and Kalmijn 2016; Bloemen, Hochguertel, and Zweerink 2017; Okamoto, Okamura, and Komamura 2018; Qvist 2022; Grøtting and Lillebø 2020; Scotti 2022), and social networks (Pilehvari, You, and Lin 2023). These studies provide meaningful insights, yet it still reports the mixed results. For example, retirement worsens subjective health for men but not (at least statistically significantly) for women (Xu 2023) on the one hand. On the other hand, the effect of retirement on biological age is not different by gender (Furuya and Fletcher 2024). Studies with mixed results suggest there may be a plausible source of the inconsistency besides heterogeneity by a *single candidate* feature.

In this study, I argue that *feature-wide effect heterogeneity* can explain the mixed results more appropriately. I define feature-wide effect heterogeneity as heterogeneity by multiple features and their interaction. Instead of examining effect heterogeneity by one feature derived from theoretical priors, assuming that the feature does not further generate heterogeneity, I explore effect heterogeneity by various social determinants and within-determinants relationships.

Abductive Reasoning through Feature-wide Effect Heterogeneity

The feature-wide effect heterogeneity approach is a type of abductive approach, which generates hypotheses rather than tests them (Merton 1987; Swedberg 2016). Previous studies apply deductive approaches and use theory to build and test hypotheses. In an abductive sense, theory is considered to be a set of beliefs that can be modified by subsequent evidence (Engzell and Mood 2023). In abductive setting, theoretical priors are not essential components to explore effect heterogeneity [1]. Instead, theoretical mechanisms are built through the results of the feature-wide effect heterogeneity approach. In doing so, I comprehensively but optimally theorize the effect of retirement on health in accordance with social determinants, rather than spotlighting specific features deduced by theoretical prior.

To accomplish this, novel methodology and data are applied to implement feature-wide effect heterogeneity. For methodology, I use causal forests, which partition into subgroups as well as estimate the unbiased effect within each subgroup (Athey and Imbens 2016; Athey, Tibshirani, and Wager 2019; Brand et al. 2021; Jawadekar et al. 2023). Causal forests are based on a machine learning approach, which aims to maximize the predictive power of the outcome (Athey and Imbens 2016). For causal forests, partitioning (i.e., prediction) and estimation (i.e., causal inference) are carried out by different training groups, a concept called "honesty". The machine learning approach sets predictive power as the evaluation criterion, and it has an obvious advantage compared to conventional approaches like linear regression. In a conventional manner, discoveries are made by running hundreds of multivariate regressions and arbitrarily picking features that are statistically significant (Brand et al. 2021). By the nature of statistical tests, this convention is an obviously questionable research practice (Brand et al. 2021). Apart from such

[1] Though setting theoretical priors is not a must-do, it still has advantages to take them into account. As all data do not contain all features, omitted features threaten the explanation of effect heterogeneity. Theoretical priors help determine what features are selected and why.

questionable practices, causal forests discover responsive features and their interactions by maximizing predictive power.

For data, I use the harmonized data of the HRS family (Lee et al. 2021), including the Health and Retirement Study, the Mexican Health and Aging Study, the English Longitudinal Study of Ageing, the Survey of Health, Ageing and Retirement in Europe, the Chinese Health and Retirement Longitudinal Study, and the Korean Longitudinal Study of Aging. Sufficient sample size is a prominent concern for the feature-wide effect heterogeneity approach (Jawadekar et al. 2023). The harmonized data overcomes this concern and enables us to divide the sample into many responsive subgroups.

Features to Heterogenize

Although the feature-wide effect heterogeneity approach may discover new features not yet considered, I cannot take *all* features into account because features in data are finite, as always. Thus, I select features that are common to all surveys. These features include sociodemographic status, income and wealth, job before retirement, and family characteristics. The selection criteria depend entirely on the data: specifically, features measured in the harmonized data As Peirce (1974) noted, plausibility, verifiability, simplicity, and economy are criteria to bring hypotheses into inference. Features measured in all data can be regarded as plausible because it implies their importance for understanding the senescence process. Data, especially social survey data, are not collected through a purely objective and neutral process; rather, they are tightly connected to human intervention (Egbert 2025). Although theoretical priors are not necessarily prerequisites for feature-wide effect heterogeneity, I complement the aforementioned nature of data by offering theoretical expectations for how these features heterogenize the effect of retirement on health. I only argue how the features produce heterogeneity and remain agnostic about the complex interactions among the features.

Sociodemographic Status

Gender Even though results are inconsistent, gender is seen as one of the important sources of effect heterogeneity of retirement on health. Men are more likely to enjoy higher wages and higher status than women before retirement. As higher wages and higher status are still related to social class in old age (Henretta and Campbell 1976), the loss of them results in worse health. Thus, the effect of retirement on health is expected to differ by gender.

Age Age Age is a multidimensional concept (Johfre and Saperstein 2023), but scholars usually regard it as the timing of retirement, or distance from the age of pension eligibility. Hult et al. (2010) and Calvo, Sarkisian, and Tamborini (2013) suggested that retirement at an older age results in worse health. Qvist (2022) has shown that the effect of timing of retirement is pronounced by social class. Skilled manual workers are less likely to visit a doctor when they retire at older age. Evidence is not consistent, but there is heterogeneity by age.

Birth cohort Birth cohort is also related to specific sociohistorical locations that shape life course trajectories (Ryder 1965; see also Elder 1998). Previous studies often regard birth cohort as reflecting differences in pension eligibility. In the case of Qvist (2022), those born in June 1939 and those born in July 1939 have an negligible difference in early retirement, and it results in worse health.

Survey year The survey year indirectly captures economic growth, politics, and other institutional settings. In terms of the effect of retirement on subsequent health, policies such as extending working lives, pension reform, and de-familialization can differentiate the relationship between retirement and health. A global disruptive event such as a recession, pandemic, or war may also affect this relationship.

Education Education affects both retirement and health. Those with higher education are likely to retire early because they earned enough throughout their careers. Education also affects health through money, knowledge, prestige, and beneficial networks (Link and Phelan 1995). In this view, higher education moderates health by mobilizing resources, thus heterogenizing the relationship between retirement and health.

Marital Status Marital status plays a crucial role in later life (Carr and Utz 2020). Grey divorce, separation, and widowhood are now undeniable transitions in later life. They are related to economic and psychological well-being (Carr and Utz 2020); thus, those who are divorced, separated, or widowed are more likely to experience worsened health.

Income and Wealth

Income and wealth Income and wealth are two financial dimensions. Income is regarded as a 'flow,' whereas wealth is viewed as a 'stock' (Spilerman 2000). As argued by Spilerman (2000), wealth principles are consumed during economic crises. This holds true for retirement, which is depicted as a loss of earnings. If retirees have enough wealth, they tend to maintain their health by complementing income loss through its mobilization. On the contrary, having enough income before retirement is tightly linked to worsening health after retirement because it directly captures the deterioration of financial circumstances. I expect that more wealth improves health after retirement and more income worsens it.

Job before Retirement

Weekly working hour Working hours gauge both higher job demand and the importance of working lives for individuals. On the one hand, overwork makes incumbents physically and psychologically exhausted, and exiting overwork improves their health (Bogaard, Henkens, and Kalmijn 2016). On the other hand, people working longer hours devote themselves to their working role (Rodgers 1978), thus the loss of it results in worsening health.

Employment status Self-employment is usually related to a lower likelihood of retirement because of a strong attachment to work and a lack of retirement savings (Radl 2013; Visser et al. 2016). For older cohorts, self-employed individuals often cannot attain employersponsored pensions. Retirement from self-employment is thus linked to precarious lives in later life, resulting in worsening health.

Family Characteristics

Parents Parental characteristics have a long-term effect on health. They affect health through childhood adversity, socioeconomic achievement, and lifestyle factors of individuals (Morton and Ferraro 2020). Not only in childhood, but parents can also modify individuals' effect of retirement on health in old age. If parents are alive, retirees can sustain a social network that supports adaptation to retirement, ultimately maintaining their health. Alive parents are also regarded as a caregiving responsibility due to growing life expectancy. Rodrigues et al. (2023) showed that recent cohorts are more likely to provide care in their 60s. Taking into account the fact that caregiving is related to worse health (Nolan, Aaltonen, and Danielsbacka 2024), having alive parents is also related to worsening health. These two channels are related to parental age; if parents are older, they are more likely to need caregiving. Parental gender is also important for identifying these two channels.

Kinship Not only the parental relationship, but children and siblings are also important sources to generate effect heterogeneity of retirement on health. If there are many children or siblings who are alive, they function as a social network. Pilehvari, You, and Lin (2023) showed that a shrinking size of social network due to retirement has a substantial effect on health. Children and siblings are usually not expected to change their relationship due to retirement, so it can suppress health deterioration by retirement by maintaining relationships between retirees and kin. Additionally, the number of members in a household is also important. If there are many household members, intergenerational contact occurs more frequently (Heylen et al. 2012). This adds new roles after retirement and thus facilitates health investment.

Intergenerational transfer Intergenerational transfer of money and/or time to (grand)children can be regarded as a new role after retirement. For example, retirees acquire the role of grandparents, thereby avoiding role loss after retirement. However, the other scenario is also possible. Drentea et al. (2024) has shown that grandparenthood lowers physical activity, thereby deteriorating health. Intergenerational transfer is usually expected to improve health, but some evidence implies worsening health.

Data and Methods

Data and Sample

I used harmonized datasets produced by The Gateway to Global Aging Data (Lee et al. 2021). The harmonized data (hereafter, the Gateway) feature an integrated survey questionnaire and consistent variables under a representative survey design based on the Health and Retirement Study (HRS). In 2021, the Gateway included eleven datasets across

38 countries. The total number of individuals was approximately 357400. To obtain as many observations and features as possible, I chose six surveys: HRS, the Mexican Health and Aging Study (MHAS), the Survey of Health, Ageing and Retirement in Europe (SHARE), the English Longitudinal Study of Ageing (ELSA), the China Health and Retirement Longitudinal Study (CHARLS), and the Korean Longitudinal Study of Aging (KLoSA). The version of each study is Version D for HRS, Version G for SHARE, Version G.3 for ELSA, Version D for CHARLS, Version E.2 for KLoSA, and Version C.2 for MHAS. A total of 272,352 individuals were obtained from the harmonized data in this study. The Gateway is suitable to investigate the feature-wide effect heterogeneity of retirement on subsequent health in this study because it has large sample size. A large sample size allows us to divide the sample into fine-grained but responsive subgroups. Consequently, I can maximize the predictive power of the heterogeneous treatment effect as well as partitioning.

"I restricted the sample to individuals aged 50 to less than 80, and those who were observed for at least three waves. I chose three waves because explicit temporary pathways are prerequisites for causal forests (Brand et al. 2021). Restricting samples to people who have complete information for selected features across three waves, the analytical sample consists of 101509 observations. The processes of sample restriction are shown in Figure S1. I use the Gateway as a pooled panel sample: if a given respondent retires twice, he or she is counted as a different respondent. The composition of the sample by survey and the percentages of missingness by feature and survey are shown in Figures S2 and S3, respectively.

Features

Outcome

I consider subjective health status as a measure of health. Codes for subjective health range from 1 for excellent to 5 for poor. Subjective health is a good indicator of health in that it predicts a wider range of other health statuses such as mortality (Idler and Benyamini 1997), genetics, and emotional health (Finkel et al. 2024). Some criticisms regarding the reliability of subjective health have been raised (Zajacova and Dowd 2011; Huisman and Deeg 2010; Bollen et al. 2021), but I still use it because specific measurements of health gauge different dimensions of the morbidity process (Crimmins, Kim, and Vasunilashorn 2010), thus it can provide a local explanation. Rather than spotlighting a specific dimension of the morbidity process, I examine the process-agnostic effect on health after retirement. Admittedly, future research should extend the framework of feature-wide effect heterogeneity to process-specific explanations of health.

Treatment

I simply measure retirement as a transition from working at $t-1$ to not working at t. I code "retire" when respondents reported working for pay at $t-1$ and not working for pay at t. I code "not retire" when respondents reported working for pay at $t-1$ and also working for pay at t. I compare the two groups and estimate the causal effect of retirement on health.

Modifier

As modifiers, sociodemographic status, income and wealth, and work and family characteristics are included. Sociodemographic status includes gender, age, birth cohort, survey years, education, and marital status. Gender includes men and women. Age, birth cohort, and survey years are continuous. Education is harmonized into three levels: less than upper secondary, upper secondary and vocational training, and tertiary. Marital status is defined as married, partnered, separated/divorced, widowed, and never married. Some waves in HRS do not distinguish between divorce and separation, so I merged them into one category. In HRS, within the married category, whether a spouse is absent or not is asked. I merged these two sub-categories into one category as "married" because no other studies apart from HRS ask this question.

Income and wealth are measured at the household level. Household income includes all income sources such as labor income, pension, social security and disability income, retirement income, unemployment income, and other government transfers for respondents and their spouses. In addition, total income from other household members is also included [2]. Household wealth is net worth, which is the sum of all assets (housing wealth and financial wealth) minus all debts (e.g., debt). The measurement of financial wealth differs among datasets. SHARE and MHAS measure financial wealth as the value of stocks, shares, and bonds, whereas the other datasets measure it as the value of stocks, mutual funds, investment trusts, bonds, and bond funds. The value of checking, savings, or money market accounts is measured in all datasets.

Work characteristics include weekly working hours and employment status before retirement. Weekly working hours range from 0 to 168. Working 0 hours (not working before retirement) and working 168 hours (working 24 hours a day) are infeasible, so I excluded the bottom 1% and top 1% of the distribution as outliers. Employment status is measured as whether the respondents are self-employed or not.

Family characteristics include parental characteristics and kinship. Parental characteristics include the number of living parents and parental age. The number of living parents ranges from 0 to 2. If all parents are alive, the number is 2. Parental age measures both the age of living parents and already deceased parents. If parents die, parental age indicates age at death. The number of household members is the total number of household members, ranging from 1 to 20. Kinship includes the number of living children and the number of living siblings. Transfer to (grand)children is a binary variable indicating whether the respondents transfer money or time to (grand)children or not.

[2] SHARE collects household income as a single question about total household income. Thus, the components of household income are not available in SHARE.

Data Analytic Plan

Estimand

The theoretical estimand in this study (Lundberg, Johnson, and Stewart 2021) begins with the conditional average treatment effect (CATE). CATE is the average treatment effect when a feature is fixed to a given value (Athey and Imbens 2016). Going beyond *one* feature, feature-wide CATE considers a wider range of features and their interactions. As argued, feature-wide CATE reveals implicit patterns of effect heterogeneity, generates hypotheses, and ultimately enhances our knowledge on retirement and subsequent health (Ludwig and Mullainathan 2024). The feature-wide CATE is defined as shown in Equation Equation 1.

$$\tau = \frac{1}{n}\Sigma_i\left[Y_i(1) - Y_i(0) \mid \vec{X}_i \in \ell(\vec{x};\Pi)\right] \qquad (1)$$

Here, τ is the feature-wide CATE, n is the number of individuals, $Y_i(1)$ and $Y_i(0)$ are the outcomes for the individual i under the treatment (1) and control (0), respectively. $\vec{X}_i$ is the covariate space of individual i, $\ell(\vec{x};\Pi)$ is the partitioning within the covariate space, and Π is the set of partitioning. I take the difference in subjective health between those who retire and those who continue to work, conditioned on a given partition $\vec{X}_i \in \ell(\vec{x};\Pi)$. The target population in this study is those at risk of retirement. In other words, I do not include those who are not alive, those who did not work at $t-1$, or those aged less than 50 and 80 or more.

Identification

To link the target estimand to observables, I first set assumptions for causal inference. These include the assumptions of positivity, consistency, no interference, and conditional exchangeability (Morgan and Winship 2015). Under these assumptions, I estimate the feature-wide CATE as the conditional average treatment effect of retirement on subsequent health within a partition. The empirical estimand is shown in Equation Equation 2.

$$\theta = \frac{1}{n}\Sigma_{i\in\mathcal{S}_{\text{retirees}}}\left[Y_i \mid \vec{X}_i \in \ell(\vec{x};\Pi)\right] - \frac{1}{n}\Sigma_{i\in\mathcal{S}_{\text{workers}}}\left[Y_i \mid \vec{X}_i \in \ell(\vec{x};\Pi)\right] \qquad (2)$$

To link feature-wide CATE to observables, I use the directed acyclic graph (DAG) in Figure 1, which clarifies the causal assumptions in this study. I address two potential sources of bias: reverse causation and classical confounding. Reverse causation refers to cases where health affects retirement. Classical confounding means the relationship between retirement and subsequent health is spurious and explained by other features. To deal with these two problems, I use a treatment weight W. The causal flow from Y_{t-1} and A_{t-1} is blocked with the weight of selection into treatment, thereby partly overcoming reverse causation and classical confounding.

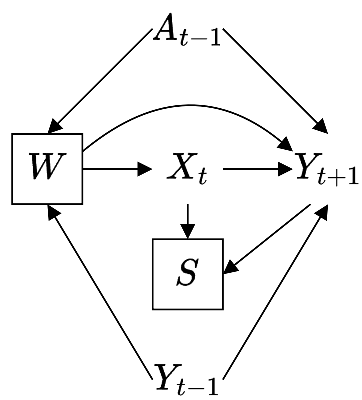


Figure 1: Directed Acyclic Graph in This Study
Source: author's own.
Note: Directed acyclic graph (DAG) is shown to visualize the causal assumptions in this study (Lundberg, Johnson, and Stewart 2021). X_t is the treatment: whether workers retire or not at a given time t. Y_{t+1} is the outcome; subjective health at $t+1$. Y_{t-1} is subjective health at $t-1$. I take Y_{t-1} into account to deal with the reverse causation that health affects retirement (Barban et al. 2020). A_{t-1} represents modifiers that can heterogenize the effect. W is a treatment weight. S is a collider.

Of course, the treatment weight is not a panacea. If there are unobserved confounders, classical confounding is not completely overcome. Furthermore, the Gateway has, like many panel surveys, attrition bias as shown by S in Figure 1. If worsening health affects the probability of participating in the survey, the estimate is upwardly biased. "Due to the nature of observational data, the precision of the estimate itself is not as satisfactory as that from (quasi-)experimental data that aim to estimate the local average treatment effect (Bloemen, Hochguertel, and Zweerink 2017; Qvist 2022; Okamoto, Kobayashi, and Komamura 2022; Mizuochi 2023; Pilehvari, You, and Lin 2023; Calvo, Sarkisian, and Tamborini 2013). However, this does not mean that the estimate here is not informative. As Lundberg, Johnson, and Stewart (2021, 562) argues,"a paper that develops a compelling theoretical estimand but relies on less-than-perfect identification assumptions should be recognized for making an important contribution: it sets the stage for future work to explore that theoretical estimand under different identification assumptions." The primary contribution of this study is to develop the estimand of feature-wide effect heterogeneity. Subsequent studies can enhance this estimand by relaxing the assumptions made in this study. Taking the triangulation of experimental design, observational data, and assumptions into account (Bailey et al. 2024), I believe the estimate from the observational data is still informative and useful.

Estimation

To estimate the feature-wide CATE, I apply honest causal forests (Athey and Imbens 2016; Brand et al. 2021; Jawadekar et al. 2023). Honest causal forests are a variant of random forests, which are ensembles of regression trees (Jawadekar et al. 2023). The difference between random forests and (honest) causal forests is that the former aims to maximize the predictive power for the outcome, while the latter aims to rigorously estimate the CATE across partitions. To obtain an unbiased estimate as well as the highest predictive power, an "honest" approach is encouraged. The honest approach involves splitting the sample into two parts: a training sample for constructing the tree and a test sample for estimating the CATE in each leaf (i.e., partition). To achieve an unbiased estimate with the highest predictive power, an expected mean squared error (EMSE) as shown in Equation Equation 3 is utilized. τ indicates the CATE in each partition $l(\vec{x};\Pi)$, $\vec{X}_i$ is the covariate space, S^{tr} is the training sample, N^{tr} is the size of the training sample, N^{est} is the size of the test sample, and $V_{S^{tr}}$ is the within-leaf variance of the CATE, respectively.

$$-\widehat{EMSE}_\tau(S^{tr}, N^{est}, \Pi) \\ = \frac{1}{N^{tr}} \Sigma_{i \in S^{tr}} \hat{\tau}^2\left(\vec{X}_i; S^{tr}, \Pi\right) - \left(\frac{1}{N^{tr}} + \frac{1}{N^{est}}\right) \cdot \Sigma_{i \in \Pi}\left(V_{S^{tr}}\left(l(\vec{x};\Pi)\right)\right) \quad (3)$$

The first term represents the magnitude of heterogeneity in the treatment effect across leaves, and the second term evaluates the variance within each leaf. Honest causal forests ultimately gauge the heterogeneous treatment effect and obtain an unbiased estimate in each leaf.

As well as the EMSE, I use feature importance to evaluate the importance of each feature. Feature importance is simply a weighted mean of the frequency of the feature's appearance in partitioning processes. Here, the weight is the depth of the feature in the partitioning (the shallower, the more important).

For honest inference for treatment effect, I use a causal tree estimator for splitting and cross-validation, which is recommended by Athey and Imbens (2016). I set the minimum sample size for splitting to 3000, and the minimum sample size within each leaf to 1000. For the importance of each feature, I set the maximum depth of the tree to 30. Although interpretation becomes more difficult as the depth of the tree deepens, it can consider more features to do so, thereby facilitating the precision of feature importance.

To complement the results, I carry out three supplementary analyses. First, I change the definition of retirement. Even though one stops working in later life, it is not necessarily retirement; it may be unemployment. I label "retired" when respondents answer "not working for pay" at t *and* subjectively report being retired at t. Second, I include country both in the treatment model and in causal forests. Country can be a promising driver of effect heterogeneity, but country itself is hard to interpret: it measures both institutional and cultural dimensions, as well as specific policies. Although I do not interpret the country, I include it in the model to quantify the importance of country. Finally, I allow for missingness of modifiers. Causal forests can estimate the treatment effect even if modifiers are missing. However, when allowing missingness, I cannot estimate the treatment weight because the treatment weight requires complete information of modifiers. Thus, I do not use the treatment weight in this scenario.

Results

Feature-wide Effect Heterogeneity

Figure 2 shows one representative causal tree for subjective health. The overall estimate is 0.142, confirming that retirement generally worsens subjective health status. However, the estimate varies from -0.131 to 0.387, indicating high heterogeneity. The first partitioning refers to age. If workers retire at 57 years old or older, health deterioration is suppressed, but retirement before 57 years old (basically viewed as early retirement) results in worse health. Among the non-early retirement group (left bucket), the second partitioning refers to the total household number. If the total household number is less than 3, their health worsens less. Among those, the third partitioning mentions household income. If household income is less than 7,058 USD (in 2010 USD), their health finally *improves*. Although the sample size is small, I detect that those who do not retire early, who have a small household number, and who have low household income report better health than those who retire. Among the group with a large household number, the third partitioning refers to weekly working hours: overwork or non-overwork. Consistent with previous studies (e.g., Bogaard, Henkens, and Kalmijn 2016), retirees who previously overworked report worse health than those who did not overwork.

Next, we move to the right bucket, the early retirement group. The second partitioning considers birth cohort, i.e., whether the birth cohort is older than 1953 or not. Basically, younger cohorts suffer worsening health. Within the older birth cohort group (left bucket), the second partitioning refers to net worth. Contrary to income, less net worth results in worse health. Finally, the third partitioning refers to the age of the father. If the father is younger than 62 years old, health worsens less.

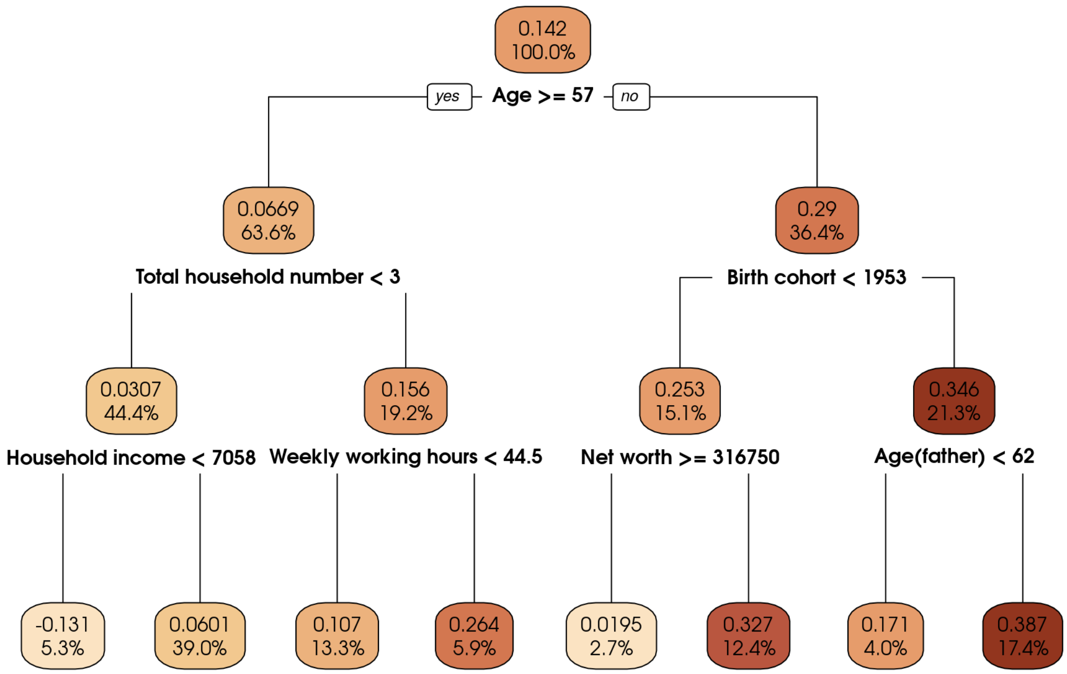


Figure 2: A representative causal tree of subjective health
Source: the Gateway
Note: This figure shows a representative causal tree estimated by a causal forests. The sample is divided by a feature and its threshold from top to bottom. If an individual is included in the threshold, he or she is allocated to the left group. If not, he or she is allocated to the right group. Among the divided group, the difference in subjective health between retirees and workers is shown, with an inverse weight of treatment probability. The selected features may be tree-specific, so I consider the feature importance among the trees below. A positive and higher value of the effect indicates worse subjective health

Figure 3 summarizes the importance of features that generate heterogeneity of retirement on subsequent health. The most important feature is age, followed by birth cohort and year. These features relate to the temporal dimension of retirement. As shown in Figure 2, people who retired at a younger age and were born in a younger birth cohort suffer greater health decline due to retirement. In terms of survey year, Figure S4 clarifies that retiring before 2009 suppresses the health decline. Next to these features, total household number, number of living parents, household income, and net worth are also informative. In detail, *more* total household number, *greater* number of living parents, *more* household income, and *less* net worth before retirement all result in worse health compared to their counterparts.

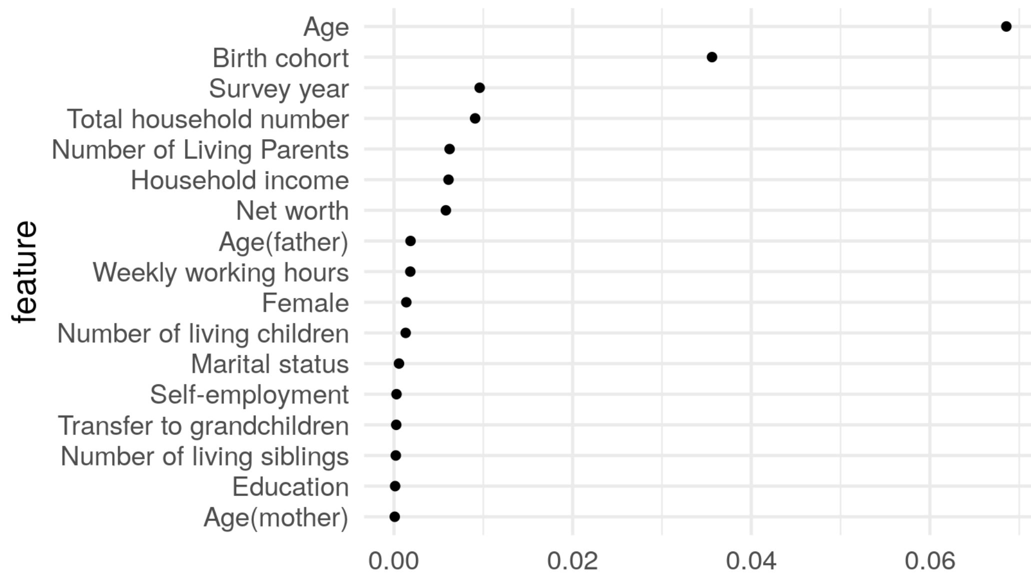


Figure 3: Feature importance of subjective health
Source: the Gateway
Note: This figure shows the importance of features in terms of effect heterogeneity. A higher value indicates that the feature matters more in generating effect heterogeneity across all causal trees

Supplementary Analyses

In Figure 4, we employ three alternative scenarios: changing the definition of retirement, including country as a feature, and allowing for missingness of modifiers. These scenarios robustly confirm the importance of age, birth cohort, household income, and net worth. On the contrary, the importance of weekly working hours and age of father, which are important in Figure 2, and the number of living parents shown in Figure 3 are not consistently detected in the alternative scenarios. Total household number, referred to in both Figure 2 and Figure 3, is more complex: it is important when we allow missingness, but not in the scenario of changing the definition of retirement and including country as a feature. "To summarize, I conclude that the importance of age, birth cohort, household income, and net worth is robust in a conservative way, and the importance of total household number may also be important but requires further research.

Country is also important, but its interpretation is quite complex, as shown in Figure S5. While this study is agnostic about how a social context across countries shapes heterogeneity, I preliminarily show that country is important and calls for future research to disentangle the country effect.

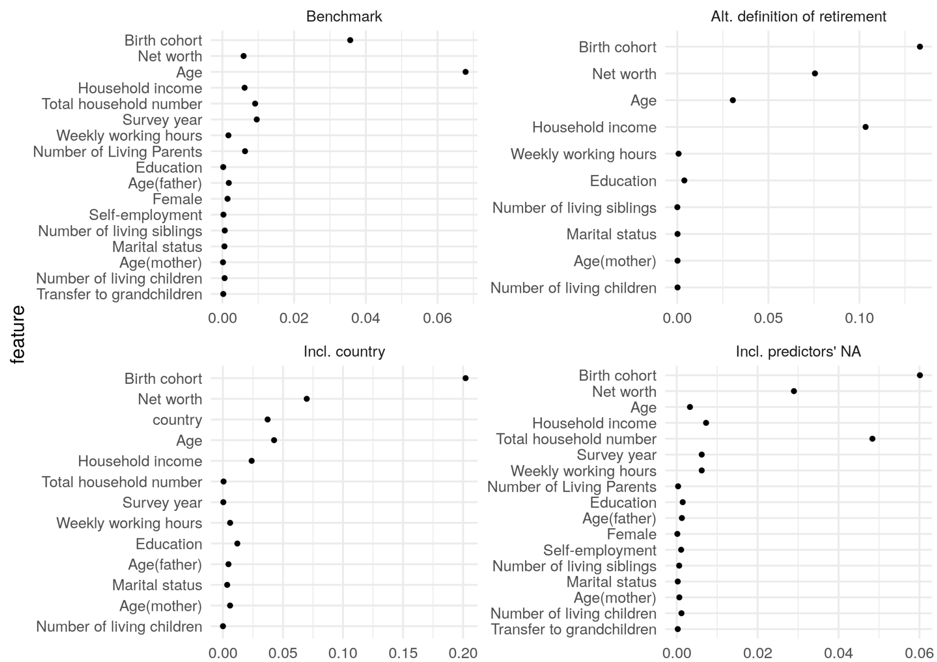


Figure 4: Feature importance of subjective health by each scenario

Discussion

Summary of Findings

In this study, I investigate the feature-wide effect heterogeneity of retirement on subsequent health through causal forests. While there's a consensus that retirement's effect on health is heterogeneous, the sources of this heterogeneity are not yet fully understood. The feature-wide effect heterogeneity approach reveals that retirement's impact on health varies significantly by age, birth cohort, survey year, household income, and household net worth. The results also suggest that total household number and survey year may further heterogenize the relationship between retirement and health. From the results, two discussions are drawn.

First and most importantly, the relationship between retirement and health has a temporal dimension. In terms of age, early retirement is related to worse health. The consequences of early retirement are debated (e.g., Qvist (2022) reports a positive effect, whereas Barban et al. (2020) reports a negative effect). I show a negative effect of early retirement on health, as well as the possibility of further disaggregation. Younger cohorts who retire at a younger age suffer worse health after retirement. Birth cohort is frequently used as an instrumental variable for pension eligibility (e.g., Furuya and Fletcher 2024; Mizuochi 2023) or policies extending working lives (e.g., Qvist 2022). The role of survey year, if it has an effect, may be highly local, specifically whether retirement occurred before 2009 or after it. In 2009, the Great Recession cast a shadow, and retiring after 2009 can be regarded as unemployment or firing rather than voluntary retirement. Taken together, the data-driven results show the temporal dimension is a prominent feature for understanding retirement and subsequent health, urging future research to disentangle its components and determine what aspects of the temporal dimension generate heterogeneity.

Second, I jointly consider total household number, total household income, and net household wealth as indicators of *financial (in)security*. A larger household number and higher household income worsen health after retirement, whereas greater net household wealth protects against health deterioration. Retirement is, by its nature, a loss of earnings, so those with higher pre-retirement incomes and those supporting more household members are more likely to suffer from this earnings loss. Wealth can be consumed during economic crises (Spilerman 2000), and it is used similarly after retirement

Double Scaffold Model

Based on the results from the feature-wide effect heterogeneity approach, I propose the "double scaffold model" of retirement and subsequent health to interpret the latent dimensions of the findings. The double scaffold model posits that the relationship between retirement and health is determined by a temporal scaffold and a financial scaffold. The conceptual framework is illustrated in Figure 5. The temporal scaffold establishes an individual's fundamental "position value," akin to an intercept. Primarily determined by age, the temporal scaffold is shaped by a myriad of factors such as pension eligibility, health capital, and human capital. While disentangling these factors is challenging, early retirees consistently appear to experience worse health outcomes following retirement.

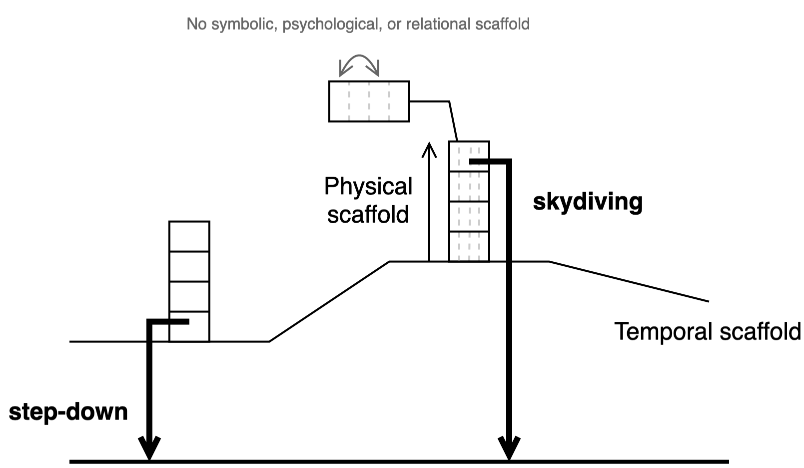


Figure 5: Double Scaffold of Retirement and Subsequent Health

Above the temporal scaffold, the financial scaffold is constructed. Retirement brings about multiple changes, including to finances, time structure, social contact, collective purpose, activities, and identity (Beehr and Bennett 2015). I reveal that the relationship between retirement and health is significantly influenced by financial factors. Household income (a flow) tends to worsen health, whereas household net worth (a stock) tends to suppress health deterioration. Though suggestive, total household number is also related to financial necessity.

The effect of retirement on subsequent health is determined by the "height" of both the temporal and financial scaffolds. For individuals at a lower temporal scaffold and lower financial scaffold, retirement might merely be a *step-down* toward a life with less work and lower income, similar to their pre-retirement life. Conversely, for those at a higher temporal and financial scaffold, retirement can be perceived as *skydiving* to a fundamentally different life. This latter group may, therefore, suffer more from the loss of daily routine and changes in financial circumstances, leading to more pronounced health deterioration.

Surprisingly, I do not confirm the symbolic, relational, or psychological scaffold. While some fundamental sociodemographic features, such as gender and education, are occasionally linked to opportunity (Beehr and Bennett 2015), they were scarcely mentioned in the results. Given the tangible experience of financial changes, the symbolic dimensions of gender and education were not confirmed as primary drivers of heterogeneity in this study.

The double scaffold model complements the health capital model (Grossman 1972). Unlike the health capital model, which explains *why* retirement affects health through individuals' rational behavior, the double scaffold model considers *whose* health retirement affects. It argues that the expectations of the health capital model are not universal but vary across individuals. Going beyond the simple notion that "everyone is different," the double scaffold model further specifies temporal and financial scaffolds as the core sources of heterogeneity in the effect of retirement on subsequent health. Going beyond the simple notion that "everyone is different," the double scaffold model further specifies temporal and financial scaffolds as the core sources of heterogeneity in the effect of retirement on subsequent health.

The feature-wide effect heterogeneity approach is instrumental in developing the double scaffold model. By simultaneously considering many features and their interaction, this approach can uncover both temporal and financial scaffolds. While a temporal scaffold might be identified by only considering heterogeneity by birth cohort, it is not well established how significant the temporal scaffold is compared to other potential scaffolds (physical, symbolic, relational, and psychological). In terms of comparing features, I also demonstrate that feature-wide effect heterogeneity is a powerful tool for providing a comprehensive landscape of heterogeneity.

Theoretical and Policy Implications

I contribute significantly to existing theory and literature. I move beyond examining heterogeneity based on a single candidate feature and instead explore feature-wide heterogeneity in a data-driven way. While a growing body of research on effect heterogeneity often identifies prominent features through theoretical considerations, I complement this approach by incorporating numerous features and their interactions. This constitutes an abductive approach, inferring the best explanation (Merton 1987; Swedberg 2016). Through the abductive approach, I contribute to theory by pinpointing where heterogeneity is generated. To achieve this, I utilized large harmonized longitudinal data and a novel methodology incorporating a machine learning approach. I suggest the double scaffold model, derived from the feature-wide effect heterogeneity approach, which highlights temporal dynamics, financial insecurity, and work characteristics as key drivers of effect heterogeneity.

Drawing from the results, I set two agendas for future research to further elaborate on the double scaffold model of retirement on subsequent health. First, as mentioned, the components of the temporal dimension should be disaggregated and directly measured. For example, birth cohort might capture pension eligibility on one hand, but it could also reflect other policy changes such as retirement policies or long-term care policies on the other. Second, other financial conditions should be considered. For instance, it may be possible to disaggregate household income into labor income, pension income, and other income such as social security. As well, household net worth can be disaggregated into housing wealth, financial wealth, and debts. These disaggregated features can be used to further investigate the financial scaffold of retirement and subsequent health.

From a policy perspective, I prioritize identifying responsive groups for intervention. First and foremost, the relationship between retirement and health is dynamic, so policymakers should be cautious about age at retirement. Considering it, I suggests potentially responsive groups: those who are younger cohorts, those with more household members, and those with less wealth and more income. In terms of inequality, the intervention strategy is not straightforward. There are two broad intervention strategies: (1) extending working lives and (2) mitigating health deterioration given retirement. Some responsive groups might have pre-retirement advantages (e.g., higher income), so intervening in their retirement behavior might inadvertently perpetuate inequality. To maximize public health while minimizing inequality, I argue that interventions should primarily focus on those whose health worsens after retirement due to a lack of advantages (i.e., lower wealth).

Limitations and Future Research

Two limitations are noteworthy mentioned. First, the results are derived from an observational setting. While observational data still plays an important role in causal inference (Bailey et al. 2024) and should be complemented by (quasi-)experimental designs, the unconfoundedness assumption inherent in causal forests is strong, as always. Further investigation into the internal validity and precise estimation of this study through (quasi-)experiments would be valuable. Second, due to the nature of machine learning, the results should be regarded as "relational, not ontological" (So and Roland 2020, 63). The Gateway, as well as all other data, only captures only a finite number of features. The results thus suffer from the so-called omitted variable bias. Despite this limitation, I attempt to abstract an ontological dimension of heterogeneity in the retirement-health relationship as temporal and financial conditioning through our discussion, partially overcoming this bias.

References

Atchley, Robert C. 1982. "Retirement as a Social Institution." *Annual Review of Sociology* 8 (1): 263–87. https://doi.org/10.1146/annurev.so.08.080182.001403.

Atchley, Robert C. 1971. "Retirement and Leisure Participation: Continuity or Crisis?" *The Gerontologist* 11 (1_Part_1): 13–17. https://doi.org/10.1093/geront/11.1_part_1.13.

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences* 113 (27): 7353–60. https://doi.org/10.1073/pnas.1510489113.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized random forests." *The Annals of Statistics* 47 (2): 1148–78. https://doi.org/10.1214/18-aos1709.

Bailey, Drew H., Alexander J. Jung, Adriene M. Beltz, Markus I. Eronen, Christian Gische, Ellen L. Hamaker, Konrad P. Kording, et al. 2024. "Causal inference on human behaviour." *Nature Human Behaviour* 8 (8): 1448–59. https://doi.org/10.1038/s41562-024-01939-z.

Barban, N., X. de Luna, E. Lundholm, I. Svensson, and F. C. Billari. 2020. "Causal Effects of the Timing of Life-course Events: Age at Retirement and Subsequent Health." *Sociological Methods & Research* 49 (1): 216–49. https://doi.org/10.1177/0049124117729697.

Beehr, Terry A., and Misty M. Bennett. 2015. "Working After Retirement: Features of Bridge Employment and Research Directions." *Work, Aging and Retirement* 1 (1): 112–28. https://doi.org/10.1093/workar/wau007.

Bloemen, Hans, Stefan Hochguertel, and Jochem Zweerink. 2017. "The causal effect of retirement on mortality: Evidence from targeted incentives to retire early." *Health Economics* 26 (12): e204–18. https://doi.org/10.1002/hec.3493.

Bogaard, Levi van den, Kène Henkens, and Matthijs Kalmijn. 2016. "Retirement as a Relief? The Role of Physical Job Demands and Psychological Job Stress for Effects of Retirement on Self-Rated Health." *European Sociological Review* 32 (2): 295–306. https://doi.org/10.1093/esr/jcv135.

Bollen, Kenneth A., Iliya Gutin, Carolyn T. Halpern, and Kathleen M. Harris. 2021. "Subjective health in adolescence: Comparing the reliability of contemporaneous, retrospective, and proxy reports of overall health." *Social Science Research* 96: 102538. https://doi.org/10.1016/j.ssresearch.2021.102538.

Brand, Jennie E., Jiahui Xu, Bernard Koch, and Pablo Geraldo. 2021. "Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning." *Sociological Methodology* 51 (2): 189–223. https://doi.org/10.1177/0081175021993503.

Calvo, Esteban, Natalia Sarkisian, and Christopher R. Tamborini. 2013. "Causal Effects of Retirement Timing on Subjective Physical and Emotional Health." *The Journals of Gerontology: Series B* 68 (1): 73–84. https://doi.org/10.1093/geronb/gbs097.

Carr, Deborah, and Rebecca L. Utz. 2020. "Families in Later Life: A Decade in Review." *Journal of Marriage and Family* 82 (1): 346–63. https://doi.org/10.1111/jomf.12609.

Crimmins, Eileen, Jung Ki Kim, and Sarinnapha Vasunilashorn. 2010. "Biodemography: New approaches to understanding trendsand differences in population health and mortality." *Demography* 47 (Suppl 1): S41–64. https://doi.org/10.1353/dem.2010.0005.

Dave, Dhaval, Inas Rashad, and Jasmina Spasojevic. 2008. "The Effects of Retirement on Physical and Mental Health Outcomes." *Southern Economic Journal* 75 (2): 497–523. https://doi.org/10.1002/j.2325-8012.2008.tb00916.x.

Drentea, Patricia, Greg Pavela, Lu Tian, and Julie L Locher. 2024. "Grandparenting and Physical Activity." *Journal of Aging and Health*, 8982643241273197. https://doi.org/10.1177/08982643241273197.

Egbert, Simon. 2025. "Predictive Analytics: A Sociological Perspective." In *The Oxford Handbook of the Sociology of Machine Learning*, edited by Christian Borch and Juan Pablo Pardo-Guerra, 715–36. https://doi.org/10.1093/oxfordhb/9780197653609.001.0001.

Elder, Glen H. 1998. "The Life Course as Developmental Theory." *Child Development* 69 (1): 1–12. https://doi.org/10.1111/j.1467-8624.1998.tb06128.x.

Engzell, Per, and Carina Mood. 2023. "Understanding Patterns and Trends in Income Mobility through Multiverse Analysis." *American Sociological Review* 88 (4): 600–626. https://doi.org/10.1177/00031224231180607.

Finkel, Deborah, Margaret Gatz, Carol E Franz, Vibeke S Catts, Kaare Christensen, William Kremen, Marianne Nygaard, et al. 2024. "Age and Sex Differences in the Genetic Architecture of Measures of Subjective Health: Relationships With Physical Health, Depressive Symptoms, and Episodic Memory." *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 79 (6): gbae062. https://doi.org/10.1093/geronb/gbae062.

Furuya, Shiro, and Jason M Fletcher. 2024. "Retirement Makes You Old? Causal Effect of Retirement on Biological Age." *Demography*. https://doi.org/10.1215/00703370-11380637.

Gorry, Devon, and Sita Nataraj Slavov. 2021. "The effect of retirement on health biomarkers." *Economics & Human Biology* 40: 100949. https://doi.org/10.1016/j.ehb.2020.100949.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55. https://doi.org/10.1086/259880.

Grøtting, Maja Weemes, and Otto Sevaldson Lillebø. 2020. "Health effects of retirement: evidence from survey and register data." *Journal of Population Economics* 33 (2): 671–704. https://doi.org/10.1007/s00148-019-00742-9.

Hämäläinen, Hans, Antti O Tanskanen, Bruno Arpino, Aïda Solé-Auró, and Mirkka Danielsbacka. 2024. "Is Transition to Retirement Associated With Volunteering? Longitudinal Evidence from Europe." *Research on Aging*, 1640275241251786. https://doi.org/10.1177/01640275241251786.

Harris, Kathleen Mullan, and Robert A Hummer. 2023. "The Increasing Salience of Health in Social Forces Research." *Social Forces* 101 (4): 1644–57. https://doi.org/10.1093/sf/soad036.

Havighurst, Robert J. 1961. "Successful Aging." *The Gerontologist* 1 (1): 8–13. https://doi.org/10.1093/geront/1.1.8.

Henretta, John C, and Richard T Campbell. 1976. "Status Attainment and Status Maintenance: A Study of Stratification in Old Age." *American Sociological Review* 41 (6): 981. https://doi.org/10.2307/2094798.

Heylen, Leen, Dimitri Mortelmans, Maarten Hermans, and Kim Boudiny. 2012. "The intermediate effect of geographic proximity on intergenerational support: A comparison of France and Bulgaria." *Demographic Research* 27: 455–86. https://doi.org/10.4054/demres.2012.27.17.

Hochman, Oshrat, and Noah Lewin-Epstein. 2013. "Determinants of early retirement preferences in Europe: The role of grandparenthood." *International Journal of Comparative Sociology* 54 (1): 29–47. https://doi.org/10.1177/0020715213480977.

Hochschild, Arlie Russell. 1975. "Disengagement Theory: A Critique and Proposal." *American Sociological Review* 40 (5): 553–69.

Huisman, Martijn, and Dorly J. H. Deeg. 2010. "A commentary on Marja Jylhä's 'What is self-rated health and why does it predict mortality? Towards a unified conceptual model'(69:3, 2009, 307–316)." *Social Science & Medicine* 70 (5): 652–54. https://doi.org/10.1016/j.socscimed.2009.11.003.

Hult, Carl, Mikael Stattin, Urban Janlert, and Bengt Järvholm. 2010. "Timing of retirement and mortality - A cohort study of Swedish construction workers." *Social Science & Medicine* 70 (10): 1480–86. https://doi.org/10.1016/j.socscimed.2009.10.043.

Idler, Ellen L, and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38 (1): 21. https://doi.org/10.2307/2955359.

Jawadekar, Neal, Katrina Kezios, Michelle C Odden, Jeanette A Stingone, Sebastian Calonico, Kara Rudolph, and Adina Zeki Al Hazzouri. 2023. "Practical Guide to Honest Causal Forests for Identifying Heterogeneous Treatment Effects." *American Journal of Epidemiology* 192 (7): 1155–65. https://doi.org/10.1093/aje/kwad043.

Johfre, Sasha, and Aliya Saperstein. 2023. "The Social Construction of Age: Concepts and Measurement." *Annual Review of Sociology* 49 (1): 339–58. https://doi.org/10.1146/annurev-soc-031021-121020.

Kim, Jungmeen E., and Phyllis Moen. 2002. "Retirement Transitions, Gender, and Psychological Well-BeingA Life-Course, Ecological Model." *The Journals of Gerontology: Series B* 57 (3): P212–22. https://doi.org/10.1093/geronb/57.3.p212.

Kuh, D, Y Ben-Shlomo, J Lynch, J Hallqvist, and C Power. 2003. "Life course epidemiology." *Journal of Epidemiology and Community Health* 57 (10): 778. https://doi.org/10.1136/jech.57.10.778.

Lee, Jinkook, Drystan Phillips, Jenny Wilkens, and Gateway to Global Aging Data Team. 2021. "Gateway to Global Aging Data: Resources for Cross-National Comparisons of Family, Social Environment, and Healthy Aging." *The Journals of Gerontology: Series B* 76 (Supplement_1): S5–16. https://doi.org/10.1093/geronb/gbab050.

Link, Bruce G, and Jo Phelan. 1995. "Social Conditions As Fundamental Causes of Disease." *Journal of Health and Social Behavior* 35: 80. https://doi.org/10.2307/2626958.

Ludwig, Jens, and Sendhil Mullainathan. 2024. "Machine Learning as a Tool for Hypothesis Generation*." *The Quarterly Journal of Economics* 139 (2): 751–827. https://doi.org/10.1093/qje/qjad055.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86 (3): 532–65. https://doi.org/10.1177/00031224211004187.

McDonough, Peggy, Diana Worts, Laurie M. Corna, Anne McMunn, and Amanda Sacker. 2017. "Later-life employment trajectories and health." *Advances in Life Course Research* 34 (Sociological Methods Research 29 2000): 22–33. https://doi.org/10.1016/j.alcr.2017.09.002.

Merton, Robert K. 1987. "Three Fragments From a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials." *Annual*

Review of Sociology 13 (1): 1–29. https://doi.org/10.1146/annurev.so.13.080187.000245.

Mizuochi, Masaaki. 2023. “Retirement Pathways’ Effect on Physical and Mental Health: Evidence from Japan.” *Journal of Aging and Health*, 8982643231200931. https://doi.org/10.1177/08982643231200931.

Morgan, Stephen L., and Christopher Winship. 2015. *CounterFactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

Morton, Patricia M., and Kenneth F. Ferraro. 2020. “Early Social Origins of Biological Risks for Men and Women in Later Life.” *Journal of Health and Social Behavior* 61 (4): 503–22. https://doi.org/10.1177/0022146520966364.

Nolan, Andreas, Katri Aaltonen, and Mirkka Danielsbacka. 2024. “The Effect of Informal Caregiving on Depression: An Asymmetric Panel Fixed-Effects Analysis of In-Home and Out-Of-Home Caregivers Across Europe.” *Journal of Aging & Social Policy* ahead-of-print (ahead-of-print): 1–19. https://doi.org/10.1080/08959420.2024.2348968.

Okamoto, Shohei, Erika Kobayashi, and Kohei Komamura. 2022. “The Retirement–Health Puzzle: A Sigh of Relief at Retirement?” *The Journals of Gerontology: Series B* 78 (1): 167–78. https://doi.org/10.1093/geronb/gbac127.

Okamoto, Shohei, Tomonori Okamura, and Kohei Komamura. 2018. “Employment and health after retirement in Japanese men.” *Bulletin of the World Health Organization* 96 (12): 826–33. https://doi.org/10.2471/blt.18.215764.

Oshio, Takashi, and Mari Kan. 2017. “The dynamic impact of retirement on health: Evidence from a nationwide ten-year panel survey in Japan.” *Preventive Medicine* 100: 287–93. https://doi.org/10.1016/j.ypmed.2017.04.007.

Peirce, Charles Sanders. 1974. *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Pilehvari, Asal, Wen You, and Xu Lin. 2023. “Retirement’s impact on health: what role does social network play?” *European Journal of Ageing* 20 (1): 14. https://doi.org/10.1007/s10433-023-00759-w.

Quaade, Thomas, Gerda Engholm, Anne Mette T Johansen, and Henrik Møller. 2002. “Mortality in relation to early retirement in Denmark: a population-based study.” *Scandinavian Journal of Public Health* 30 (3): 216–22. https://doi.org/10.1080/14034940210134022.

Qvist, Jeevitha Yogachandiran. 2022. “Early Retirement and Social Class: A Health-Giving Choice for All?” *European Sociological Review* 39 (1): 132–44. https://doi.org/10.1093/esr/jcac029.

Radl, Jonas. 2013. "Labour Market Exit and Social Stratification in Western Europe: The Effects of Social Class and Gender on the Timing of Retirement." *European Sociological Review* 29 (3): 654–68. https://doi.org/10.1093/esr/jcs045.

Rodgers, Daniel T. 1978. *The Work Ethic in Industrial America 1850-1920*. Chicago: University of Chicago Press.

Rodrigues, Ricardo, Johan Rehnberg, Cassandra Simmons, Stefania Ilinca, Eszter Zólyomi, Afshin Vafaei, Selma Kadi, Janet Jull, Susan P Phillips, and Stefan Fors. 2023. "Cohort Trajectories by Age and Gender for Informal Caregiving in Europe Adjusted for Sociodemographic Changes, 2004 and 2015." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 78 (8): 1412–22. https://doi.org/10.1093/geronb/gbad011.

Ryder, N B. 1965. "The cohort as a concept in the study of social change." *American Sociological Review* 30 (6): 843–61.

Scotti, Benedetta. 2022. "Late-Career Employment Trajectories and Postretirement Mortality: Evidence From Italy." *Demography* 59 (6): 2187–2213. https://doi.org/10.1215/00703370-10291269.

So, Richard Jean, and Edwin Roland. 2020. "Race and Distant Reading." *PMLA/Publications of the Modern Language Association of America* 135 (1): 59–73. https://doi.org/10.1632/pmla.2020.135.1.59.

Spilerman, Seymour. 2000. "Wealth and Stratification Processes." *Annual Review of Sociology* 26 (1): 497–524. https://doi.org/10.1146/annurev.soc.26.1.497.

Swedberg, Richard. 2016. "Theorizing in Sociological Research: A New Perspective, a New Departure?" *Annual Review of Sociology* 43 (1): 1–18. https://doi.org/10.1146/annurev-soc-060116-053604.

Taylor, Philip. 2019. "Working Longer May Be Good Public Policy, But It Is Not Necessarily Good for Older People." *Journal of Aging & Social Policy* 31 (2): 99–105. https://doi.org/10.1080/08959420.2019.1576487.

Visser, Mark, Maurice Gesthuizen, Gerbert Kraaykamp, and Maarten H. J. Wolbers. 2016. "Inequality among Older Workers in the Netherlands: A Life Course and Social Stratification Perspective on Early Retirement." *European Sociological Review* 32 (3): 370–82. https://doi.org/10.1093/esr/jcw013.

Xu, Yuanrong. 2023. "The effect of retirement on health and mortality in the United States." *Journal of Population Research* 40 (2): 12. https://doi.org/10.1007/s12546-023-09307-w.

Zajacova, Anna, and Jennifer Beam Dowd. 2011. "Reliability of Self-rated Health in US Adults." *American Journal of Epidemiology* 174 (8): 977–83. https://doi.org/10.1093/aje/kwr204.

Zhao, Meng, Yoshifumi Konishi, and Haruko Noguchi. 2017. "Retiring for better health? Evidence from health investment behaviors in Japan." *Japan and the World Economy* 42: 56–63. https://doi.org/10.1016/j.japwor.2017.06.003.

Supplementary Materials

Descriptive statistics

Table S1: Descriptive statistics

	All	%	Mean	SD	min	max
Subjective Health	101509	100	2.78	1.03	1	5
Retirement	101509	100	0.18	0.38	0	1
Birth cohort	101509	100	1946.97	8.95	1918	1966
Age	101509	100	59.26	5.87	50	75
Sex						
Male	51286	50.52				
Female	50223	49.48				
Marital status						
Married	78443	77.28				
Partnered	4133	4.07				
Separated/Divorced	11501	11.33				
Widowed	5161	5.08				
Never married	2271	2.24				
Number of living siblings	101509	100	3.02	2.41	0	21
Survey year	101509	100	2006.23	7.19	1992	2016
Self-employment	101509	100	0.22	0.41	0	1
Weekly working hours	101509	100	37.90	15.68	0	168
Education						
Less than upper secondary	23253	22.91				
Upper secondary and vocational training	52351	51.57				
Tertiary	25905	25.52				

	All	%	Mean	SD	min	max
Number of Living Parents	101509	100	0.53	0.68	0	2
Age(father)	101509	100	71.92	13.80	10	120
Age(mother)	101509	100	75.96	12.80	10	117
Net worth	101509	100	235787.22	324832.18	-49900	5384157.17
Household income	101509	100	58772.31	66666.17	0	1510327.12
Transfer to grandchildren	101509	100	0.42	0.49	0	1
Total household number	101509	100	2.60	1.29	1	20
Number of living children	101509	100	2.93	1.74	0	15

Participant flow

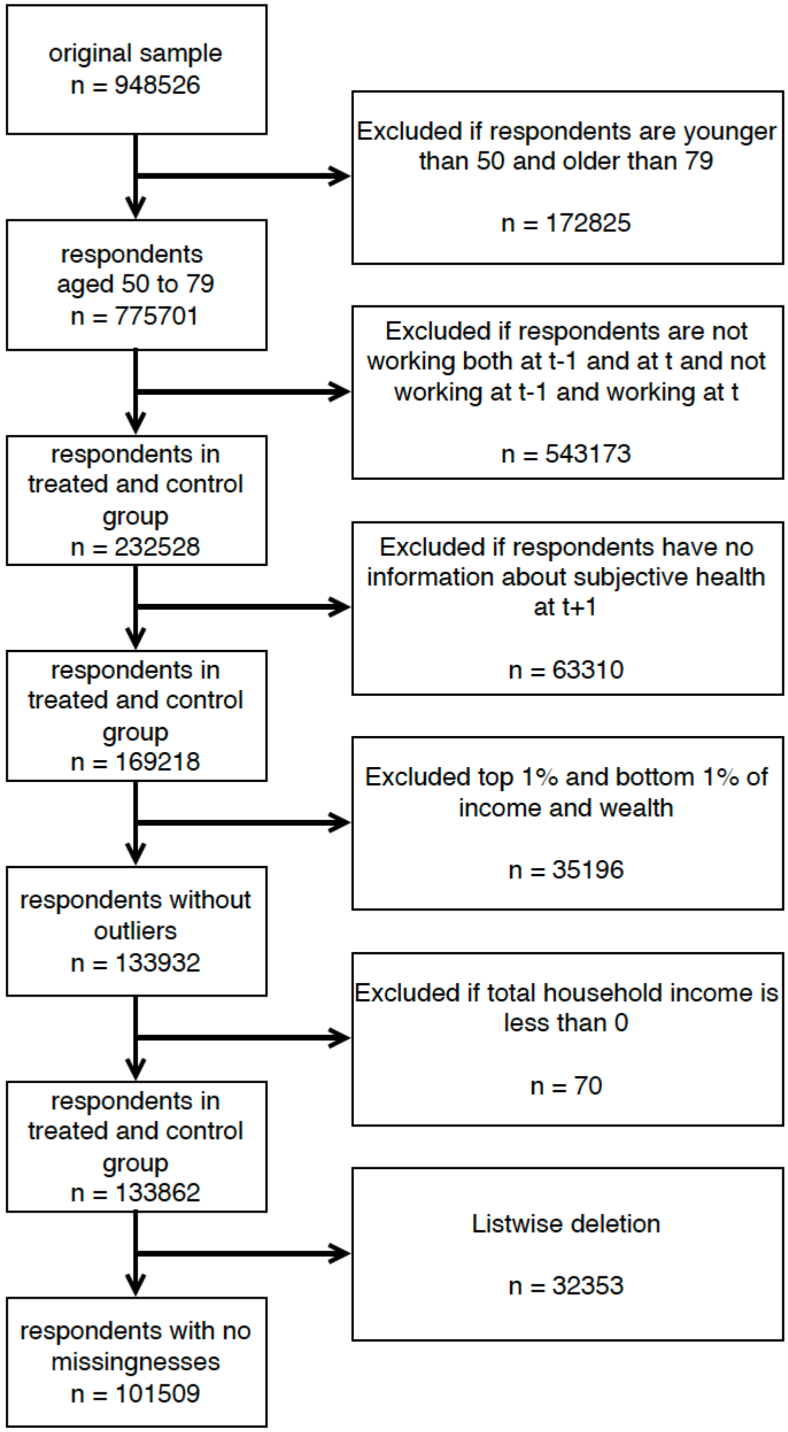


Figure S 1: Flowchart of sample selection in the Gateway

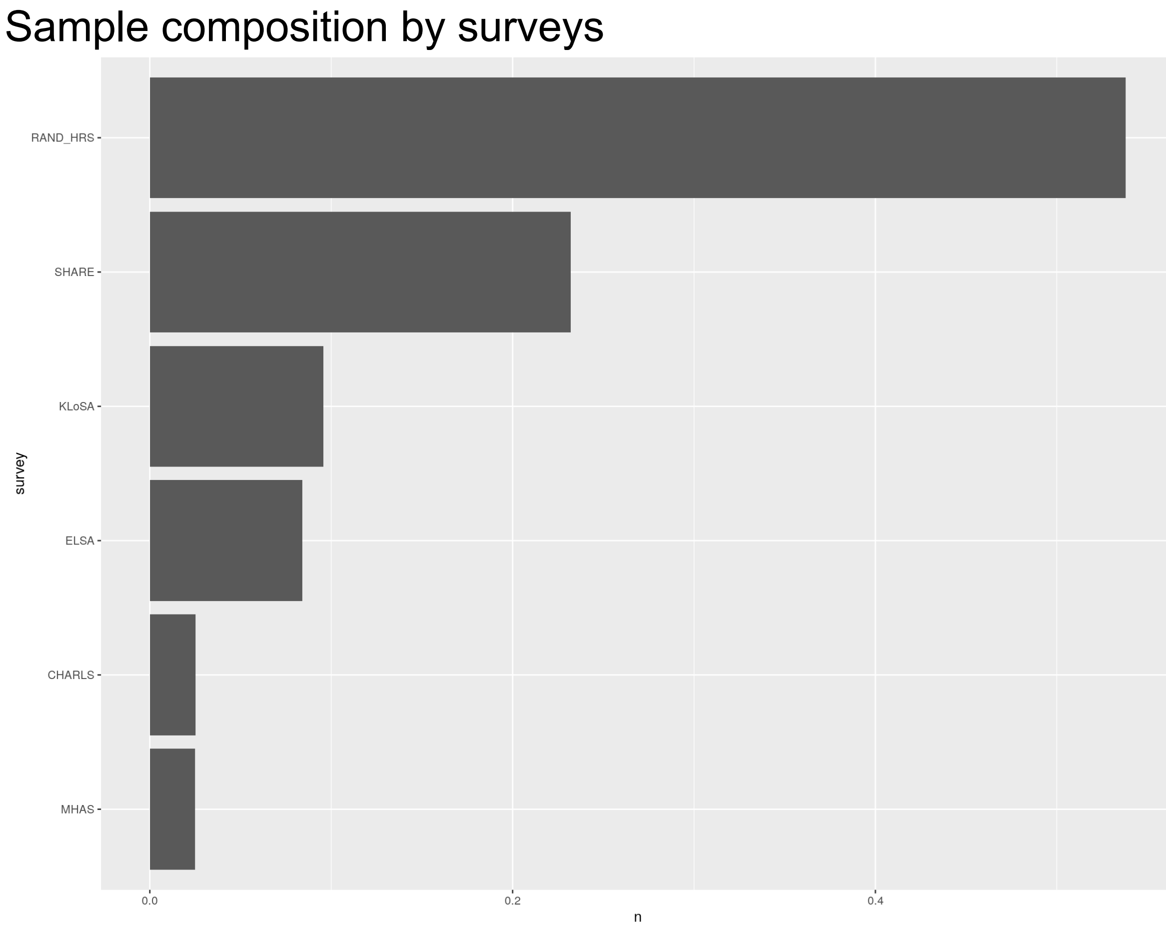


Figure S 2: Composition of sample by surveys

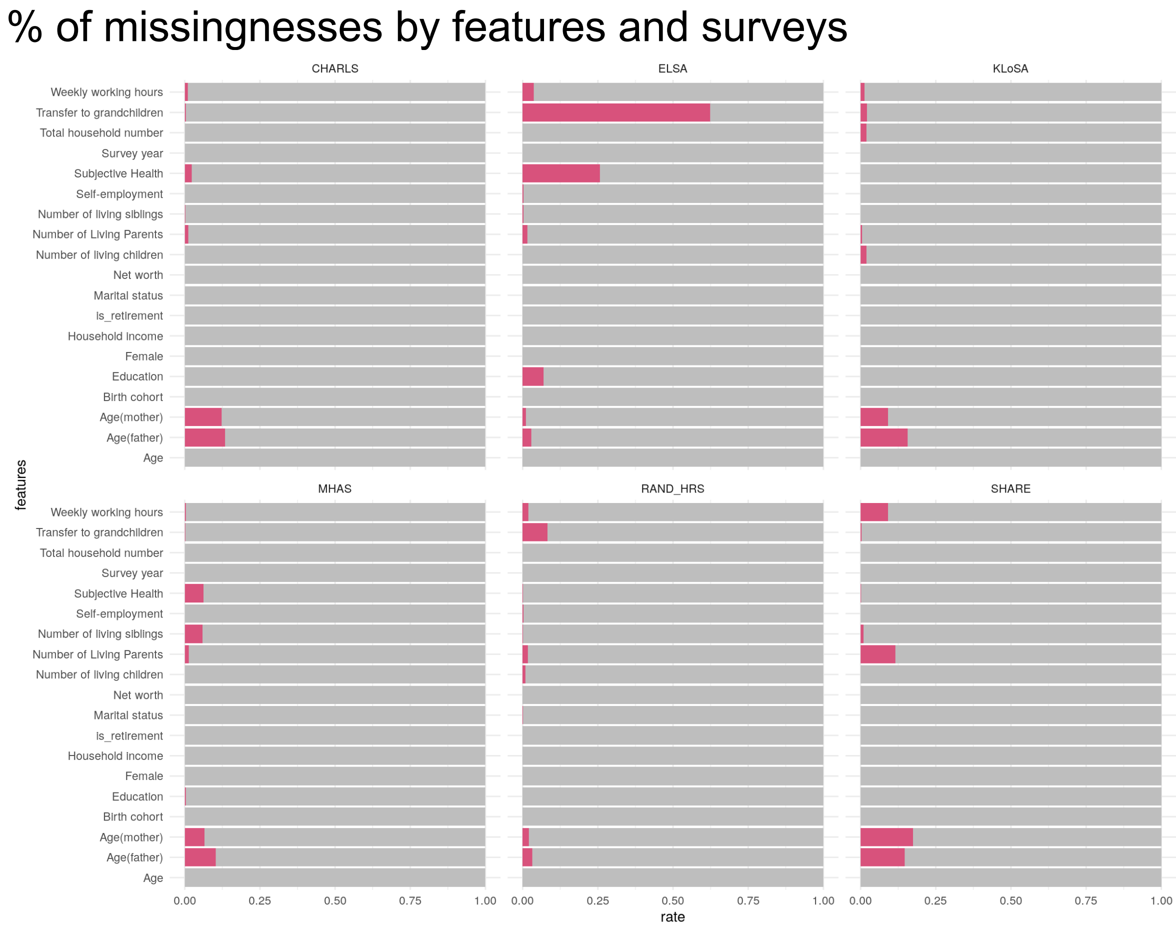


Figure S 3: % of missingnesses by features and surveys

A causa tree of subjective health with a deeper depth

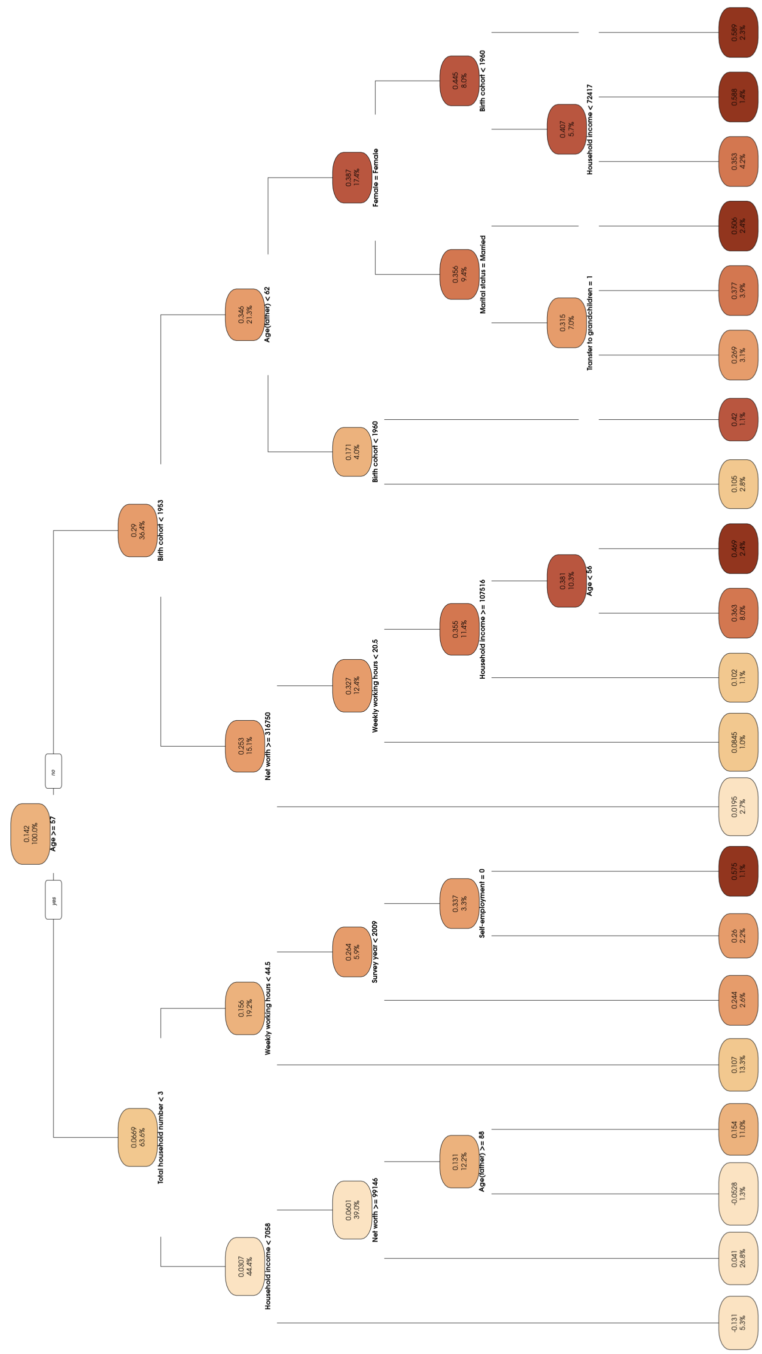


Figure S 4: A representative causal tree of subjective health, depth 6

A causa tree of subjective health including country

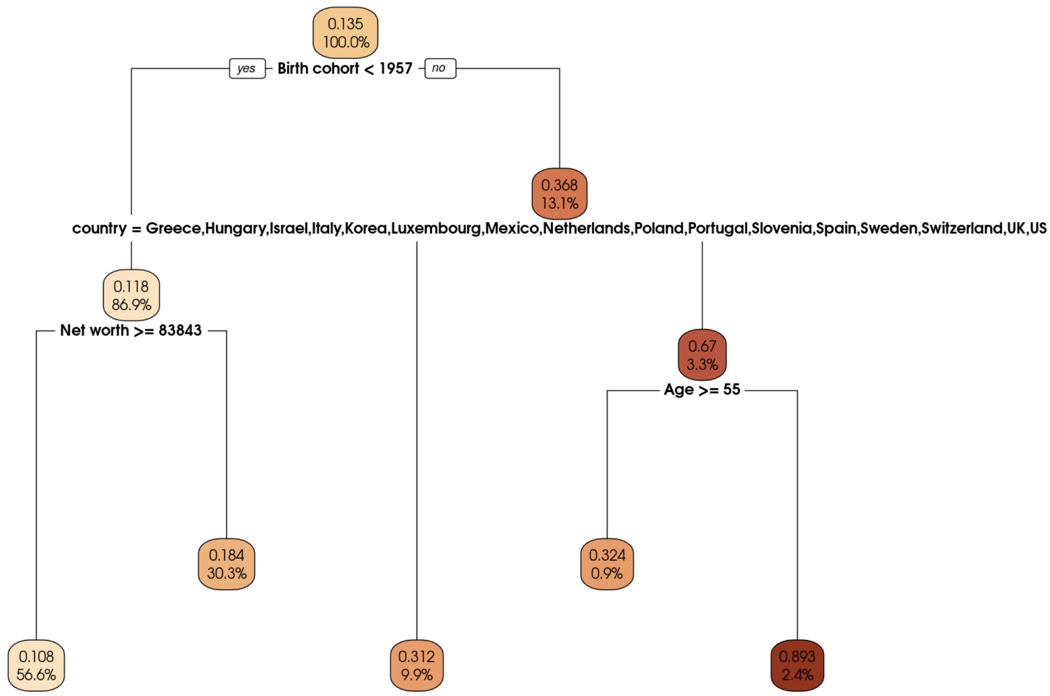


Figure S 5: A representative causal tree of subjective health including countries