



RIETI Discussion Paper Series 25-E-128

# Healthcare Utilization in the Face of Illness and Costs: Evidence from Japan

**SAXENA, Akshar**

Nanyang Technological University

**YIN, Ting**

RIETI

**FAN, Mingxuan**

National University of Singapore

**WANG, Wenjie**

Nanyang Technological University



Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry  
<https://www.rieti.go.jp/en/>

Healthcare Utilization in the Face of Illness and Costs: Evidence from Japan<sup>\*</sup>Akshar Saxena<sup>†</sup>, Ting Yin<sup>‡§</sup>, Mingxuan Fan<sup>\*\*</sup>, Wenjie Wang<sup>††</sup>

## Abstract

This study examines how health shocks to individual family members affect healthcare utilization patterns across entire households. We investigate spillover mechanisms such as preventive behavior changes and disease impacts. Using comprehensive Japanese healthcare claims covering the period from 2005-2023, we exploit the quasi-random timing of acute health events to identify causal effects on family members' healthcare utilization. Our identification strategy leverages detailed longitudinal data linking household members over 18 years. We extract dynamic responses from immediate reactions, examining healthcare utilization and preventive care.

We find substantial clustering of cardiovascular disease (CVD) diagnoses within families: 22% of family members' CVD diagnoses occur within the first year following an index case, with the strongest and most immediate responses among spouses and rapid intergenerational effects between parents and children. We also identify preventive screening as a key protective mechanism—family members who undergo health check-ups after an index diagnosis experience a 13-percentage-point reduction in two-year CVD risk. These findings highlight that health shocks generate significant intra-household behavioral responses, with implications for healthcare cost projections, preventive program design, and family-centered policy interventions.

Keywords: Health spillovers, Cardiovascular disease, Family health behaviours, Preventive screening  
Household health production

JEL classification: I12, I18, J13

The RIETI Discussion Paper Series aims at widely disseminating research results in the form of professional papers, with the goal of stimulating lively discussion. The views expressed in the papers are solely those of the author(s), and neither represent those of the organization(s) to which the author(s) belong(s) nor the Research Institute of Economy, Trade and Industry.

<sup>\*</sup>This study is conducted as a part of the Project “Economic Analysis on the problem of an aging population and a declining birthrate in China and Japan in the COVID-19 pandemic” undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the data provided by the JMDC data center, with the funding support from the Research Institute of Economy, Trade and Industry (RIETI).

The draft of this paper was presented at the RIETI DP seminar for the paper. I would like to thank participants of the RIETI DP Seminar for their helpful comments.

<sup>†</sup> Division of Economics, Nanyang Technological University, Singapore. [aksharsaxena@ntu.edu.sg](mailto:aksharsaxena@ntu.edu.sg)

<sup>‡</sup> Department of International Economics Faculty of Economics, Toyo University, Japan. [yin@toyo.jp](mailto:yin@toyo.jp)

<sup>§</sup> Research Institute of Economy, Trade and Industry (RIETI), Japan. [yin-ting@rieti.go.jp](mailto:yin-ting@rieti.go.jp)

<sup>\*\*</sup> Department of Real Estate, NUS Business School, National University of Singapore, Singapore. [mfan@nus.edu.sg](mailto:mfan@nus.edu.sg)

<sup>††</sup> Division of Economics, Nanyang Technological University, Singapore. [wang.wj@ntu.edu.sg](mailto:wang.wj@ntu.edu.sg)

# 1 Introduction

Cardiovascular disease (CVD) is the leading global cause of death, accounting for nearly 18 million deaths annually ([World Health Organization 2021](#)). While CVD has been extensively studied in clinical and epidemiological settings, its spillover effects within families remain understudied, particularly regarding how health shocks affect the diagnosis and preventive behavior of other household members. Healthcare decisions are fundamentally embedded within household contexts, yet empirical research has predominantly focused on individual-level responses to health shocks and policy changes.

This study examines how CVD diagnoses within families induces spillover effects among other household members, addressing two primary questions. First, does the diagnosis of a cardiovascular event in one family member increase the likelihood of diagnosis in others? Second, does this health shock induce preventive behavior, specifically uptake of health check-ups, and does such screening mitigate subsequent disease risk? Using comprehensive administrative claims data from Japan spanning 2005-2023, we leverage the quasi-random timing of acute cardiovascular events to identify effects on family members' subsequent CVD diagnoses and preventive screening behaviors. Our analysis reveals substantial spillover effects operating through multiple channels: shared risk recognition leading to increased disease detection, behavioral responses in preventive care uptake, and protective effects from family-motivated screening.

The research addresses several important gaps in the health economics literature. First, while extensive work has documented individual responses to health shocks, relatively little attention has been paid to how these shocks propagate through family networks. Second, most studies of preventive care focus on individual decision-making, overlooking how family health events can serve as powerful motivators for screening behavior. Third, the temporal dynamics of intra-household health spillovers, particularly the timing and

persistence of responses, remain poorly understood.

Our findings contribute to these strands of literature in several ways. We document that 42% of family member CVD diagnoses occur within one year of an index case, with particularly strong clustering among spouses and rapid responses between parents and children. The probability of CVD diagnosis varies significantly by family relationship, with children and siblings facing the highest risk but longest delays to diagnosis. Importantly, we find that preventive screening undertaken after a family member's CVD diagnosis is associated with a 13 percentage point reduction in two-year CVD risk, highlighting the protective potential of family-motivated health behaviors.

These results have important policy implications. Healthcare cost projections that ignore household-level spillovers may systematically underestimate total expenditure following major health events. Prevention programs that target families rather than individuals may achieve greater effectiveness by leveraging natural information channels and behavioral responses. Insurance design that accounts for family-level risk clustering and screening responses could improve both coverage adequacy and cost efficiency.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and positions our contribution. Section 3 describes the data and empirical methodology. Section 4 presents the main results on CVD spillovers within families and the timing of these effects. Section 5 examines the role of preventive screening as a protective mechanism. Section 6 concludes with policy implications and directions for future research.

## 2 Literature Review

### 2.1 Individual Responses to Health Shocks

The theoretical foundation for understanding healthcare demand following health shocks was established by [Grossman \(1972\)](#), who modeled health as both a consumption and investment good. This framework predicts that individuals will adjust their healthcare utilization following changes in health status or information about health risks. Subsequent empirical work has extensively documented such individual-level responses across various contexts.

[Finkelstein et al. \(2012\)](#) demonstrate that individuals substantially increase healthcare utilization following health insurance expansion in Oregon, with particularly large effects on preventive care and chronic disease management. [Card et al. \(2008\)](#) find significant discontinuities in healthcare utilization at Medicare eligibility, suggesting that insurance coverage changes trigger immediate behavioral responses. More recently, [Einav and Finkelstein \(2013\)](#) show how individuals learn about their health status over time and adjust utilization accordingly, while [Polyakova \(2016\)](#) documents how health experiences influence subsequent insurance choices.

Studies of acute health shocks have found similar patterns of behavioral adaptation. [Dobkin et al. \(2018\)](#) show that individuals reduce labor supply and increase healthcare utilization following hospitalizations, with effects persisting for several years. [Fadlon and Nielsen \(2019\)](#) examine responses to spousal hospitalization, finding that healthy spouses increase their own healthcare utilization and reduce risky behaviors. These studies establish that health shocks generate significant and persistent behavioral responses, but focus primarily on direct effects on the affected individual or their immediate caregiver.

## 2.2 Household-Level Health Decisions and Spillovers

Despite extensive evidence on individual responses, the literature on household-level health spillovers remains limited. Theoretical extensions of the Grossman framework to consider families as health producers predict that health shocks to one member can affect others through budget constraints, caregiving responsibilities, health production complementarities, and information spillovers about shared risk factors (Jacobsen 2000).

Our findings relate to a growing body of work on peer effects and spillovers in health behavior. Social networks influence health through information diffusion, behavioral imitation, and emotional contagion (Manski 2000; Kremer and Levy 2001). Family-specific spillovers have been documented in vaccination uptake (Banerjee et al. 2010), fertility decisions (Munshi 2003), and cancer screening (Saxena et al. 2025). However, evidence on cardiovascular disease spillovers within families remains limited.

Empirical evidence on such spillovers has been constrained by data limitations and identification challenges. Early work by Christakis and Fowler (2007) documented clustering of health behaviors and outcomes in social networks, but could not establish causal relationships. More recent studies have made progress using quasi-experimental variation. Fadlon and Nielsen (2019) examine spillover effects of spousal job loss on healthcare utilization, finding significant reductions in care following income shocks.

The literature on disease clustering within families has focused primarily on genetic and environmental factors rather than behavioral responses to health information. Studies in genetics and epidemiology document substantial familial clustering of cardiovascular disease, but typically treat this clustering as evidence of shared risk factors rather than examining how diagnoses themselves trigger behavioral changes among family members (Saxena and Mendenhall 2022; Ma and Saxena 2025).

## 2.3 Preventive Care and Health Information

A separate literature examines how health information affects preventive care utilization. [Dranove et al. \(2003\)](#) show that public reporting of hospital quality information increases patient sorting and hospital competition. [Saxena et al. \(2025\)](#) show how detection of cancer within family promotes preventive genetic testing among other members. More broadly, studies have documented that health information campaigns, media coverage of health risks, and peer health experiences can influence preventive care decisions, though empirical evidence often struggles with selection bias and reverse causality.

However, this literature has largely overlooked family health events as sources of health information. The few studies that examine family health histories focus on genetic counseling contexts or inherited disease risks, rather than examining how acute health events within families trigger preventive behaviors among other members. This represents an important gap, as family health events may be particularly salient sources of health information given their emotional impact and relevance for shared risk factors.

## 2.4 Institutional Context and Contribution

Japan's healthcare system provides several advantages for studying household spillover effects. Universal coverage minimizes financial barriers to care, allowing focus on behavioral rather than access-driven responses. The fee-for-service payment system generates detailed claims records enabling precise measurement of utilization patterns. Comprehensive population coverage and household linkages in administrative data overcome the sample size and measurement limitations that have constrained previous research.

Our study makes several contributions to these strands of literature. We provide the first comprehensive analysis of intra-household CVD spillovers using population-scale administrative data with family linkages. Second, we establish causal identification through

the quasi-random timing of acute health events, addressing endogeneity concerns that have limited previous research. Third, we examine both the extensive margin (whether family members develop CVD) and the timing of responses, providing novel evidence on the temporal dynamics of health spillovers. Fourth, we identify preventive screening as a key mechanism through which families can mitigate negative spillover effects, contributing to studies on both health information and prevention.

The combination of rich administrative data, clean identification, and focus on both spillover mechanisms and protective responses allows us to provide comprehensive evidence on household-level health dynamics that has important implications for both academic understanding and health policy design.

## **3 Data and Methods**

### **3.1 Data and Sample Construction**

We use administrative claims data from the JMDC Database in Japan, covering the period 2005–2023. The JMDC Claims Database is an epidemiological receipt database that has accumulated receipts (inpatient, outpatient, dispensing) and medical examination data received from multiple health insurance associations since 2005. The respondents in this JMDC dataset are those covered by employer-based insurance and companies in Japan (including small ones) are required to provide employees with mandatory health check-ups under the Industrial Safety and Health Act. While the health-checks are mandatory for the employees, the dependents are encouraged to co-operate with health-checks. However, if dependents work part-time for another company, they may also be subject to mandatory health check-ups. For all others the participation is voluntary and thus we expect that around 70-80% may have received check-ups in our sample.

The dataset links individual-level records over time and across family members through

unique personal identifiers, enabling the construction of a rich longitudinal family panel. Our analysis merges three data sources, described below.

The first source is the diagnosis file, which records individual healthcare encounters by date (year and month) and includes ICD-10 diagnosis codes. We use this to identify instances of cardiovascular disease (CVD) diagnosis and treatment. The second source is the health check-up file, which records the date and results of periodic medical screenings. Linking the diagnosis and check-up files allows us to construct an individual-level history of both preventive and diagnostic healthcare use. The third source is the personal information file, which includes demographic details such as date of birth, gender, relationship to the primary insured individual, and household identifiers. This file enables us to link individuals within the same family and map intra-household health trajectories.

### **3.1.1 Sample Construction**

We construct our sample by first identifying index cases—individuals who experience a qualifying CVD diagnosis during the study period. Using household identifiers, we link these index individuals to other family members and trace healthcare utilization over time. We retain individuals whose relationship to the primary insured is clearly defined as self, spouse, or child, and exclude cases with ambiguous or missing relationship codes (e.g., “unknown,” “other”). To ensure clean identification of the index case, we exclude families where two members were diagnosed with CVD in the same month. Additionally, we restrict the sample to families with at least one recorded CVD diagnosis, as our focus is on intra-household spillovers following an index health shock.

### **3.1.2 Health Shock Definition**

We define health shocks as the first recorded diagnosis of acute cardiovascular conditions that are largely unanticipated in timing. Our primary definition includes first-time myocardial infarction (ICD-10: I21.x), stroke (I63.x, I64.x), and hospitalization for heart failure (I50.x). These events are selected for their clinical severity, low predictability, and

potential to trigger behavioural responses within the family.

### 3.2 Empirical Strategy

Our empirical strategy exploits the quasi-random timing of acute cardiovascular health shocks to identify their causal effect on other family members' healthcare utilization. The key identifying assumption is that the precise timing of such events is largely unpredictable and thus exogenous to the health-seeking behaviour or latent disease risk of other family members. While families naturally share genetic predispositions and environmental factors that influence baseline CVD risk, the sudden occurrence of an acute event, such as a heart attack or stroke, is largely unforeseeable. Our identification therefore relies not on the absence of shared risk factors, but on the timing variation of these acute events. This allows us to interpret downstream changes in diagnosis or screening as responses to the index case, rather than reflections of shared underlying trends.

To estimate the effect of an index case's CVD diagnosis on subsequent diagnoses within the household, we estimate the following [Equation 1](#):

$$Y_{ij} = \beta_0 + X_i' \beta + \sum_{k=1}^4 \beta_k \cdot 1(C_i = k) + \beta_5 GI_j + \tau_y + \gamma_m + \varepsilon_j \quad (1)$$

Here,  $Y_{ij}$  is a binary indicator equal to one if individual  $i$  in family  $j$  is diagnosed with CVD after the index case during the observation period. The vector  $X_i$  includes individual-level covariates: age, gender, and family size. The categorical variable  $C_i$  denotes the individual's relationship to the index case, with four main groups: spouse, child, parent, and sibling. The omitted category is the index case themselves. We also include an indicator  $GI_j$  for whether the index case is female, allowing us to examine heterogeneity in response by the gender of the initial case.

To control for time-varying confounders, we include month fixed effects  $\gamma_m$  to account

for seasonal variation in healthcare utilization and diagnosis, and year fixed effects  $\tau_y$  to absorb secular trends in medical technology, screening access, and treatment protocols. Standard errors are clustered at the family level to account for intra-household correlation in health behaviour and outcomes.

## 4 Results

### 4.1 Descriptive Statistics

Table 1 presents summary statistics for the analytic sample of 5.6 million individuals. The average age is 32.7 years (SD = 21.1), and 46.5% of the sample is female. The average family size is 3.3 members. In terms of relationship to the index case, 34.2% of observations pertain to the index individual, 24.4% to their spouse, and 41.4% to their children. Overall, 46.8% of individuals have ever been diagnosed with cardiovascular disease (CVD), and families have an average of 1.4 members diagnosed (SD = 0.6). Health screening is relatively limited: the average number of check-ups is 2.6 (SD = 3.9), and only 28.3% of individuals underwent a check-up after the index case's diagnosis. Among those screened after the index case, 39.0% had never previously been screened (i.e., had their health check-up for the first time after index case), indicating a potential behavioural response triggered by the family health event.

### 4.2 Spillover of CVD within Family

Figure 1 displays the time distribution of CVD diagnoses among family members relative to the index case, measured in years. The distribution is highly skewed towards the early years, with the largest share, 20% of diagnoses occurring within the same year following the index case. This is followed by 22% in the first year after the index diagnosis, and a sharp decline thereafter. By Year 5, the proportion falls below 5%, and stabilises at low levels in later years. This pattern suggests that a substantial portion of intra-household

CVD diagnoses are temporally clustered around the index event, likely reflecting a behavioural or informational response triggered by the initial diagnosis. Such clustering is consistent with mechanisms such as increased health awareness, medical check-ups, or shared exposure being activated by the index case.

To explore heterogeneity in spillover patterns, Panel A of [Figure 2](#) disaggregates the timing of diagnoses by relationship dyad. Spillovers between spouses exhibit the strongest and most immediate response: nearly 25% of spouses who eventually develop CVD are diagnosed in the same year as the index case, and an additional 23% are diagnosed in the following year. This sharp, front-loaded pattern suggests that shared lifestyle risk factors, co-residence, and spousal salience may play important roles in prompting early detection or diagnosis. In contrast, the parent-to-child dyad shows a more gradual but still elevated response. About 15% of children are diagnosed in the same year as the parent, rising to 21% in the subsequent year and remaining elevated through Year 5. This more dispersed profile likely reflects differences in baseline risk, access to screening, or salience of parental health events for younger adults.

[Figure 2](#) Panel B shows that spillovers from child to parent are also concentrated in the immediate aftermath of the index case. Roughly 14% of parents who eventually develop CVD are diagnosed in the same year as their child, rising sharply to 23% in the following year. This temporal clustering is similar in intensity to the spousal pattern, suggesting that parental health behaviour may be particularly responsive to children's health shocks, possibly due to increased concern, contact with the healthcare system, or changes in perceived familial risk. By contrast, sibling-to-sibling spillovers are more diffuse. While 13% are diagnosed in Year 0 and 22% in Year 1, the distribution remains relatively flat across subsequent years, with smaller declines over time. This flatter profile may reflect weaker co-residence or salience effects, but still points to a moderate behavioural response within sibling networks.

We use [Equation 1](#) to estimate the probability of a family member developing CVD conditional upon the index-case and controlling for individual and family characteristics. Results from estimation are presented in [Table 2](#). Column 1 includes only individual-level characteristics, while Column 2 adds indicators for the individual’s relationship to the index case. Column 3 introduces month fixed effects to account for seasonality in access to care and CVD diagnosis, as well as year fixed effects to absorb long-term trends in clinical practices and diagnostic criteria. Column 4, the fully saturated model, additionally includes family fixed effects and clusters standard errors at the family level to account for within-household correlation.

Across all specifications, own age is a significant predictor of CVD onset, with each additional year associated with a 0.7 to 1.1 percentage point increase in probability. Female individuals are significantly less likely to be diagnosed with CVD post-index case, with marginal effects ranging from 2.0 to 2.6 percentage points lower than males. Larger families are associated with slightly higher risk: the coefficient on family size increases in magnitude across specifications and becomes 0.021 in the fully saturated model (column 4), suggesting that individuals in larger households may face greater cumulative exposure to shared risk or detection mechanisms. Finally, we note that having a female index case is associated with a small but significant increase (0.9 percentage points) in the probability that other family members are diagnosed with CVD. This may reflect differences in how female index cases affect family awareness, screening uptake, or household health behaviour.

We next examine the effect by relationship to the index case. Estimates from Column 4 are translated into predicted probabilities for ease of interpretation and presented in [Figure 3](#). The right panel shows that children and siblings of the index case have the highest likelihood of being diagnosed with CVD following the index event. The predicted probability is approximately 0.26 for children and 0.28 for siblings, with both effects

statistically significant at the 1% level. These elevated risks suggest substantial intra- and inter-generational clustering of CVD, potentially driven by shared genetic predispositions, common environments, or increased salience and behavioural responses to a family member’s diagnosis. In contrast, the left panel shows that the probability of diagnosis is considerably lower among spouses (0.10) and parents (0.07), pointing to generational differences in baseline risk profiles, health-seeking behaviour, or responsiveness to health information within households.

As a robustness check, we estimate a logistic regression model to account for the binary nature of the outcome variable and to assess the sensitivity of our findings to functional form assumptions. The results, presented in [Table 3](#), are consistent in both direction and statistical significance with those from the linear probability model reported in [Table 2](#) and visualised in [Figure 3](#). This consistency suggests that our findings are not driven by the choice of functional form and are robust to nonlinear specifications.

### **4.3 Timing of Spillover of CVD within Families**

We previously illustrated the temporal pattern of CVD spillovers within families in [Figures 1](#) and [2](#). While these figures provide descriptive insights, they are not adjusted for seasonal variation in diagnosis (e.g., across months), long-run changes in medical practice (e.g., across years), or observable individual- and household-level covariates. To formally assess the timing of CVD onset among family members after an index case, we re-estimate [Equation 1](#), changing the dependent variable to the time elapsed (in months) between the index case’s diagnosis and the subsequent CVD diagnosis of a family member. For this analysis, the sample is restricted to family members who were diagnosed with CVD after the index case.

The estimation results are presented in [Table 4](#). Column 4 includes the full set of controls:

individual-level covariates, month and year fixed effects, and family-level clustering to account for intra-household correlation. We find that age and family size are significant predictors of the time to diagnosis. A one-year increase in age is associated with a 0.25-month longer duration to diagnosis, while each additional family member adds approximately 5.5 months. The individual's gender is not statistically significant, but the gender of the index case matters: if the index case is female, the time to subsequent diagnosis among other family members is reduced by nearly two months, a statistically significant effect at the 1% level. This suggests that female index cases may trigger faster intra-household recognition or action in response to health risk.

Figure 4 visualises the average time lag in diagnosis across different relationship dyads. While Panel B of Figure 3 had shown that children of the index case exhibit the highest probability of developing CVD, the results here indicate they also experience the longest delay before diagnosis—on average, 41 months (approximately 3.5 years) after the index case. Siblings, who also had high incidence, are diagnosed after 36 months on average. In contrast, spouses and parents are diagnosed more quickly, at approximately 29–30 months post-index case. These results suggest that while spillovers exist across all relationship types, the timing of response varies systematically by familial proximity and potentially by caregiving or co-residence patterns.

These differences in timing suggest that while some family members, such as spouses, respond relatively quickly to a CVD diagnosis within the household, others experience substantial delays, particularly children and siblings. This raises the question of whether and to what extent the diagnosis of an index case motivates other family members to engage in preventive screening, a behavioural pathway we examine in the next section.

## 4.4 Spillover of CVD to Screening and its Protective Effect

We define intra-family spillover as the development of cardiovascular disease (CVD) by a family member subsequent to the index case. One potential mechanism to mitigate this negative spillover is through preventive health screening, which may enable early detection of risk factors such as pre-hypertension before the onset of CVD. The respondents in this JMDC dataset are those covered by employer-based insurance and companies in Japan (including small ones) are required to provide employees with health check-ups. Therefore, most people (around 70-80%) are expected to receive health check-ups mandatorily.

Table 5 reports the proportion of family members who underwent health screening following the index case’s diagnosis. Overall, 84% of family members underwent a health check-up after the index-case’s CVD diagnosis. The lowest uptake is notes in the small sample of secondary spouse (55%). In contrast, spouses and parents of index case appear more responsive, with 83.34% and 85.78% undergoing screening after the index case. Similarly, 78% of parents undergo screening if a child is diagnosed with CVD, but before their own diagnosis.

To assess whether such screenings are associated with a reduced likelihood of developing CVD, we estimate the following specification:

$$Y_{ij} = \beta_0 + \beta_1 \text{Check-Up}_i + \beta_2 X_i + \alpha_j + \gamma_m + \varepsilon_j \quad (2)$$

Here,  $Y_{ij}$  is a binary indicator for whether individual  $i$  in family  $j$  develops CVD within two years of the index case. The key regressor  $\text{CheckUp}_i$  is a binary variable equal to one if the individual underwent a health check-up after the index case but before their own diagnosis (if applicable), or at any point after the index case if they did not develop CVD. Individuals who never had a check-up are coded as zero. The vector  $X_i$  includes controls for gender, relationship to the index case, and birth year. Family fixed effects  $\alpha_j$  absorb time-invariant characteristics shared within families—such as genetic risk or

socioeconomic status—while month fixed effects  $\gamma_m$  capture seasonal variation in diagnosis patterns. Standard errors are clustered at the family level.

Table 6 presents the results and we find that undergoing a health check-up after the index case is associated with a 13 percentage point reduction in the probability of developing CVD within two years. This sizeable effect suggests that behavioural responses to health shocks within families, specifically through increased uptake of screening, can meaningfully reduce short-run CVD risk.

#### 4.4.1 Heterogeneity Analysis

We next account for heterogeneity in screening timing. In particular, some individuals may undergo preventive screening prior to the index case’s diagnosis due to other motivations, such as personal health concerns, employment requirements, or exposure to health campaigns. To examine this, we reclassify  $\text{CheckUp}_i$  as a categorical variable with the following groups: (i) never screened; (ii) screened before the index case; and (iii) screened after the index case but before own diagnosis (if applicable), or at any time after the index case (if no diagnosis occurred).

Table 7 reports the results and we find that individuals who were screened *before* the index case already exhibit a significantly lower CVD risk, suggesting a protective effect from proactive screening. Compared to those never screened, early screeners are 5.8 percentage points less likely to develop CVD within two years. Individuals who undergo screening *after* the index case face an even larger reduction of 17 percentage points. These results underscore both the direct benefits of screening and the behavioural responsiveness to health events within the family.

Figure 5 visualises the predicted probabilities from this regression. The top panel shows that individuals who were never screened face a baseline CVD probability of 42%, which

declines to 36% for those screened before the index case, and further to 25% for those screened after the index case. This pattern illustrates the strong behavioural effect induced by family health shocks and the potential for preventive care to mitigate disease risk.

#### **4.4.2 Robustness Check: 5-Year CVD Risk**

While the 2-year window serves as our primary time frame of interest, capturing the behavioural response most immediately following a family health shock, we also estimate the model using a 5-year horizon to assess the persistence and stability of the screening effect.

This longer time window allows us to evaluate whether the protective association between health check-ups and CVD development persists over a broader period, or whether it attenuates as individuals become exposed to other risk factors or health shocks unrelated to the index case.

As shown in Table 6 column (2), the magnitude of the effect declines slightly when the outcome is defined over a 5-year window. Individuals who underwent a health check-up after the index case experienced an 8 percentage point reduction in CVD risk, compared to a 13 percentage point reduction over two years. This attenuation is expected, as the longer horizon increases the likelihood of CVD onset due to unrelated causes.

We then repeat the analysis using the categorical specification of screening timing. Table 7 columns (4)-(6) show the regression estimates using the same categorical specification of check-up timing. The results indicate that the protective effects remain statistically significant but are more muted: screening before the index case reduces CVD probability by 4 percentage points, while screening after the index case reduces it by 10 percentage points.

Figure 5 presents the predicted probabilities. The baseline 5-year CVD risk among those never screened is 48%, falling to 44% among early screeners and to 38% among those screened after the index case. Although the magnitudes are smaller relative to the 2-year results, the direction and significance of the effects remain consistent, reinforcing the robustness of our findings.

Together, the main and robustness results suggest that preventive health check-ups—especially when timed in response to a family member’s diagnosis—are strongly associated with a reduced likelihood of developing CVD. These findings highlight the importance of spillover-responsive screening as a potentially cost-effective strategy for CVD prevention.

## 5 Limitations

Several limitations should be acknowledged when interpreting our findings. First, while our identification strategy leverages the quasi-random timing of acute CVD events, we cannot fully eliminate concerns about shared family characteristics such as genetic predispositions, dietary habits, and home environments. Although the sharp temporal clustering we observe is consistent with behavioral responses, residual confounding from unobserved family-level factors may remain. In our future work, we will employ family fixed effects and event study designs with pre-trend analysis to separate behavioral responses from shared risk factors. We will also conduct analysis separately individuals who undergo regular annual screening (As mandatorily provided by employers) as compared to those who may forego regular checkups for more comprehensive medical checkup (ningen dock). Additionally, we will conduct heterogeneity analysis by co-residence status to assess if spillovers are stronger among co-residing family members due to behavioral and environmental channels driving the effects.

Second, our analysis focuses exclusively on cardiovascular disease within Japan’s employer-based health insurance system, potentially limiting generalizability to other conditions or healthcare contexts. Third, the absence of socioeconomic data restricts our ability to examine heterogeneity across population subgroups, though Japan’s universal coverage may mitigate some access-related confounding. Fourth, while we document strong associations between post-shock screening and reduced CVD risk, our observational design cannot completely rule out selection bias in screening behaviors. Finally, our data do not capture important behavioral mechanisms such as lifestyle modifications, medication adherence, or psychological stress responses that may mediate the observed spillovers.

## 6 Conclusion

This study provides the first comprehensive analysis of intra-household spillover effects following cardiovascular disease diagnoses using population-scale administrative data from Japan. Our findings demonstrate substantial clustering of CVD diagnoses within families, with 42% of family member diagnoses occurring within the first year following an index case. The spillover patterns vary significantly by family relationship: spouses exhibit the most immediate responses with nearly 25% diagnosed in the same year as the index case, while children and siblings show higher overall probabilities of eventual diagnosis (26% and 28% respectively) but longer delays to detection. Importantly, we identify preventive screening as a key mechanism through which families can mitigate these negative spillovers. Family members who undergo health check-ups after the index case experience a 13 percentage point reduction in two-year CVD risk, suggesting that health shocks serve as powerful behavioral nudges that can be leveraged for disease prevention.

Overall our findings have important implications for healthcare policy and clinical practice. Healthcare cost projections that ignore household-level spillovers may systematically underestimate total expenditure following major health events, as the ripple effects

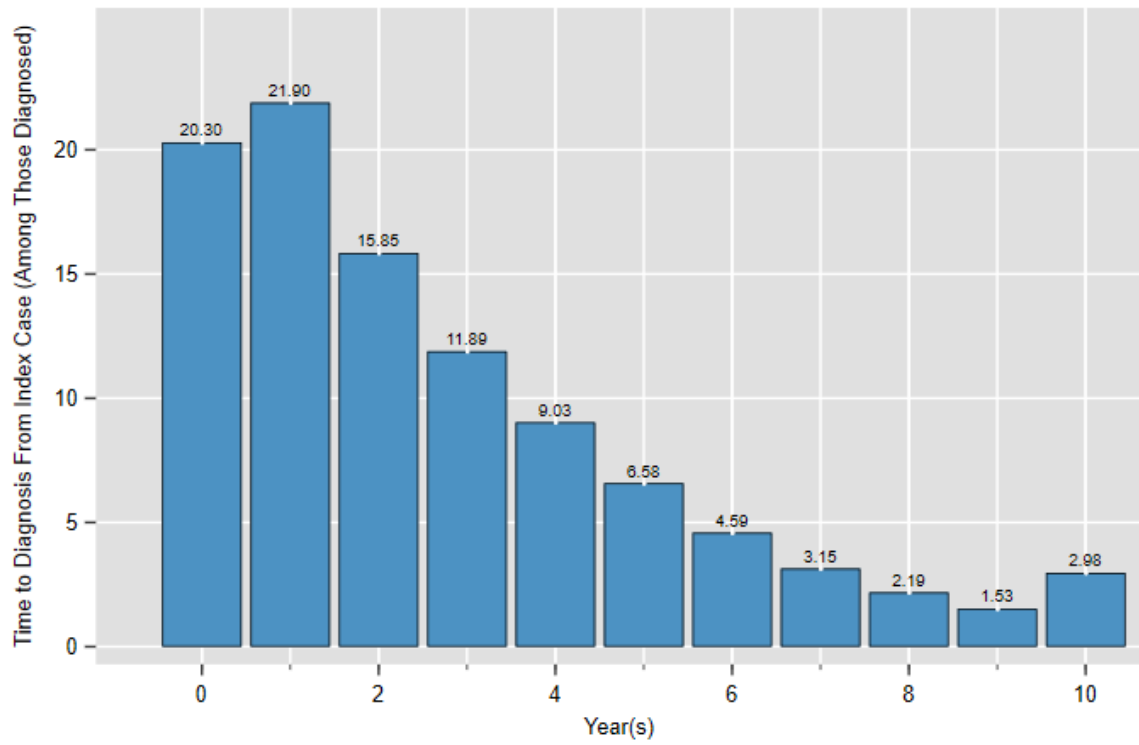
extend well beyond the index patient. Prevention programs should consider targeting entire families rather than individuals, leveraging the natural information channels and heightened health awareness that follow family health crises. Our results suggest that family-centered screening interventions could be particularly cost-effective, capitalizing on the increased receptivity to preventive care that follows a relative's diagnosis. For health systems like Japan's, where family-based care responsibilities are culturally embedded, understanding and utilizing these spillover dynamics could enhance both population health outcomes and healthcare efficiency. Future research should explore whether similar patterns exist for other chronic diseases and examine the long-term persistence of these behavioral changes to inform optimal intervention timing and design.

## References

- Banerjee, Abhijit V, Esther Duflo, Rachel Glennerster, and Dhruva Kothari, “Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives,” *British Medical Journal*, 2010, *340*, c2220.
- Card, David, Carlos Dobkin, and Nicole Maestas, “The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare,” *American Economic Review*, 2008, *98* (5), 2242–2258.
- Christakis, Nicholas A and James H Fowler, “The spread of obesity in a large social network over 32 years,” *New England Journal of Medicine*, 2007, *357* (4), 370–379.
- Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo, “The economic consequences of hospital admissions,” *American Economic Review*, 2018, *108* (2), 308–352.
- Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite, “Is more information better? The effects of ‘report cards’ on health care providers,” *Journal of Political Economy*, 2003, *111* (3), 555–588.
- Einav, Liran and Amy Finkelstein, “Moral hazard in health insurance: what we know and how we know it,” *Journal of the European Economic Association*, 2013, *11* (5), 957–982.
- Fadlon, Itzik and Torben Heien Nielsen, “Family health behaviors,” *American Economic Review*, 2019, *109* (9), 3162–3191.
- Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group, “The Oregon health insurance experiment: evidence from the first year,” *The Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.
- Grossman, Michael, “On the concept of health capital and the demand for health,” *Journal of Political Economy*, 1972, *80* (2), 223–255.
- Jacobsen, Linda K, “The family as producer of health—an extended Grossman model,” *Journal of Health Economics*, 2000, *19* (4), 611–637.
- Kremer, Michael and Dan M Levy, “Peer effects and alcohol use among college students,” *Journal of Economic Perspectives*, 2001, *15* (4), 189–206.
- Ma, Zhuoru and Akshar Saxena, “Identifying joint spatial risk factors for Hypertension and Diabetes in China: A bivariate joint spatial modeling approach with CHARLS Wave 1 and Wave 4,” 2025.
- Manski, Charles F, “Economic analysis of social interactions,” *Journal of Economic Perspectives*, 2000, *14* (3), 115–136.
- Munshi, Kaivan, “Networks in the modern economy: Mexican migrants in the US labor market,” *The Quarterly Journal of Economics*, 2003, *118* (2), 549–599.
- Polyakova, Maria, “Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D,” *American Economic Journal: Applied Economics*, 2016, *8* (3), 165–195.
- Saxena, Akshar and Emily Mendenhall, “Syndemic thinking in large-scale studies: Case studies of disability, hypertension, and diabetes across income groups in India and China,” *Soc. Sci. Med.*, February 2022, *295* (113503), 113503.
- , Renne Phay, Jianbang Chiang, Fy Wong, Diana Ishak, Siqin Zhou, Sara Tasnim, and Joanne Ngeow, “Does financial subsidy equalize cancer genetic testing uptake across socioeconomic groups? A retrospective observational study in Singapore,” *BMJ Open*, 2025.
- World Health Organization, “Cardiovascular diseases (CVDs),” [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) 2021. Accessed: 2024.

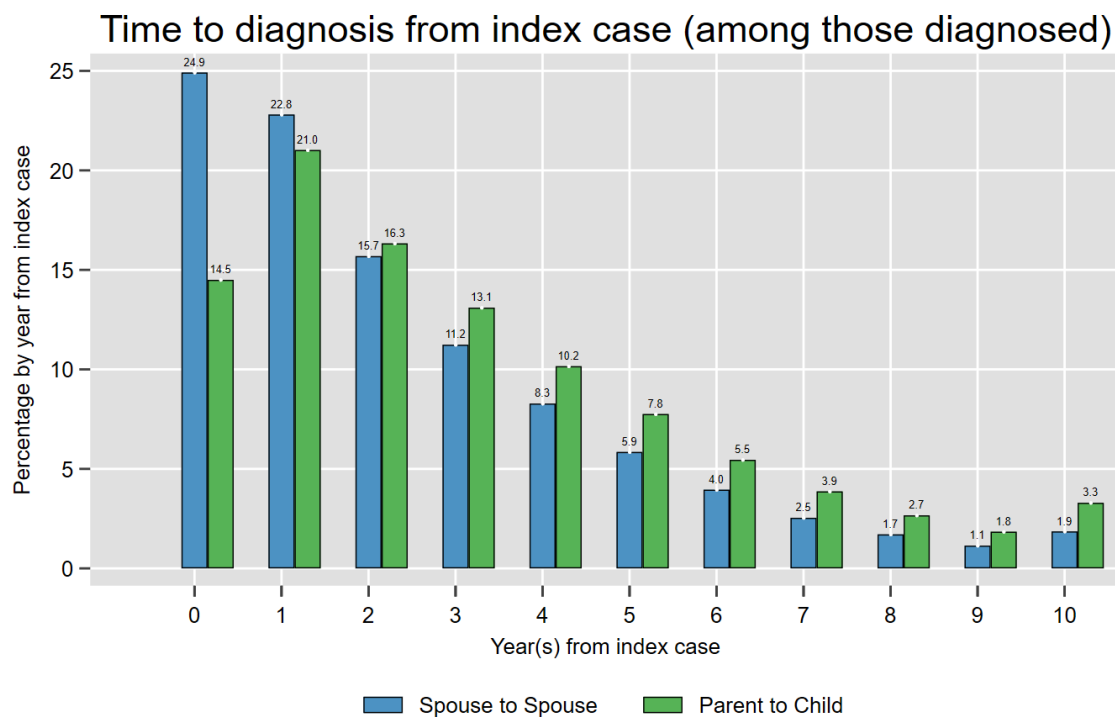
## Figures and Tables

Figure 1: Timing of CVD Diagnoses Among Family Members Relative to Index Case

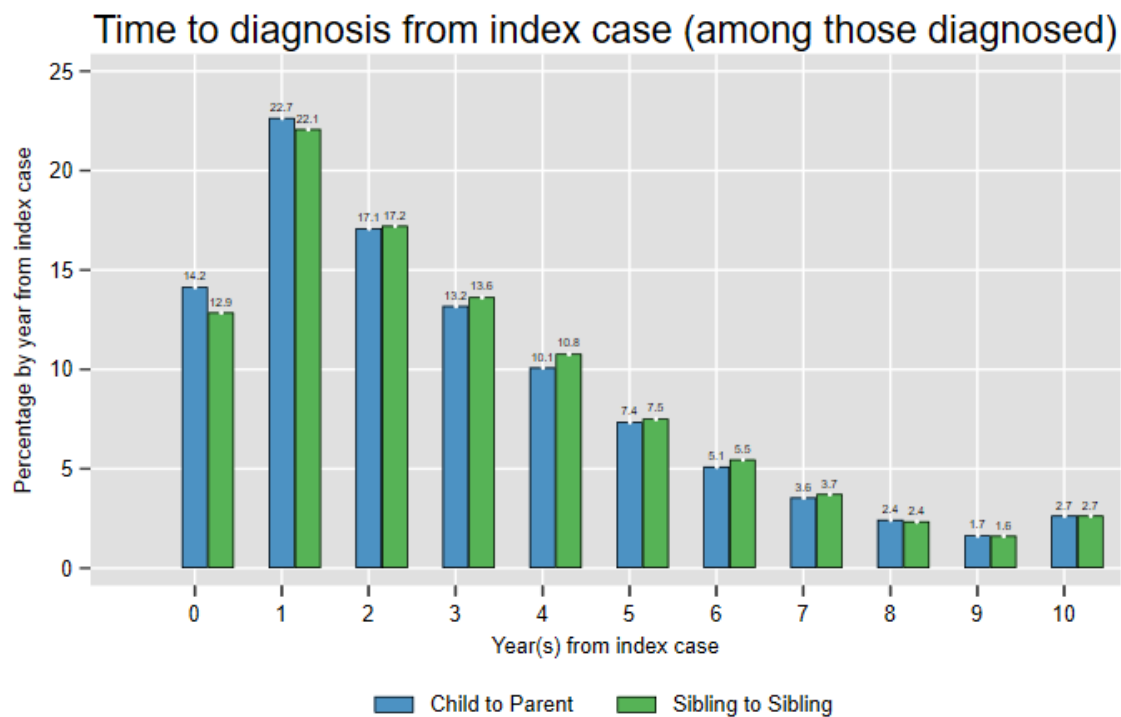


*Notes:* The figures show the coefficients and corresponding 95% confidence intervals for the association of health checkup on being diagnosed with CVD within 2 years of index case within family. The comparison of proportions by period is conducted only among families within an affected member.

Figure 2: Time Trend of CVD Onset Among Family Members Following the Index Case, by Dyad Type



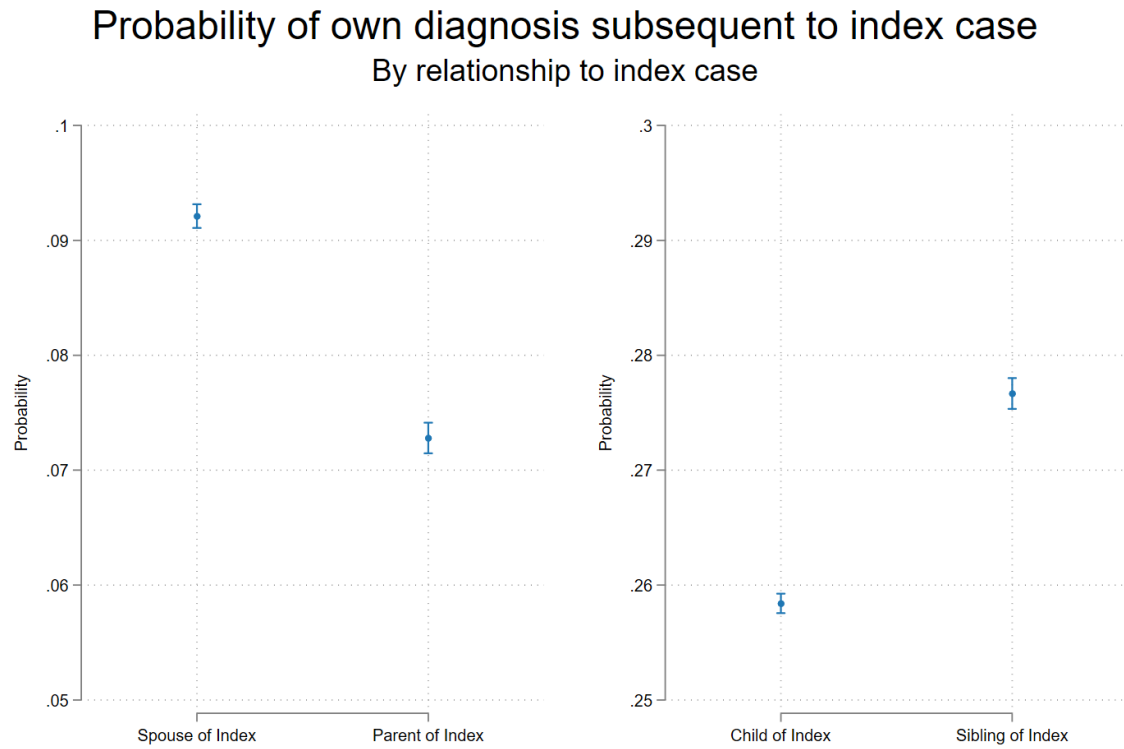
(a) Spillovers from Spouse to Spouse and Parent to Child



(b) Spillovers from Child to Parent and Sibling to Sibling

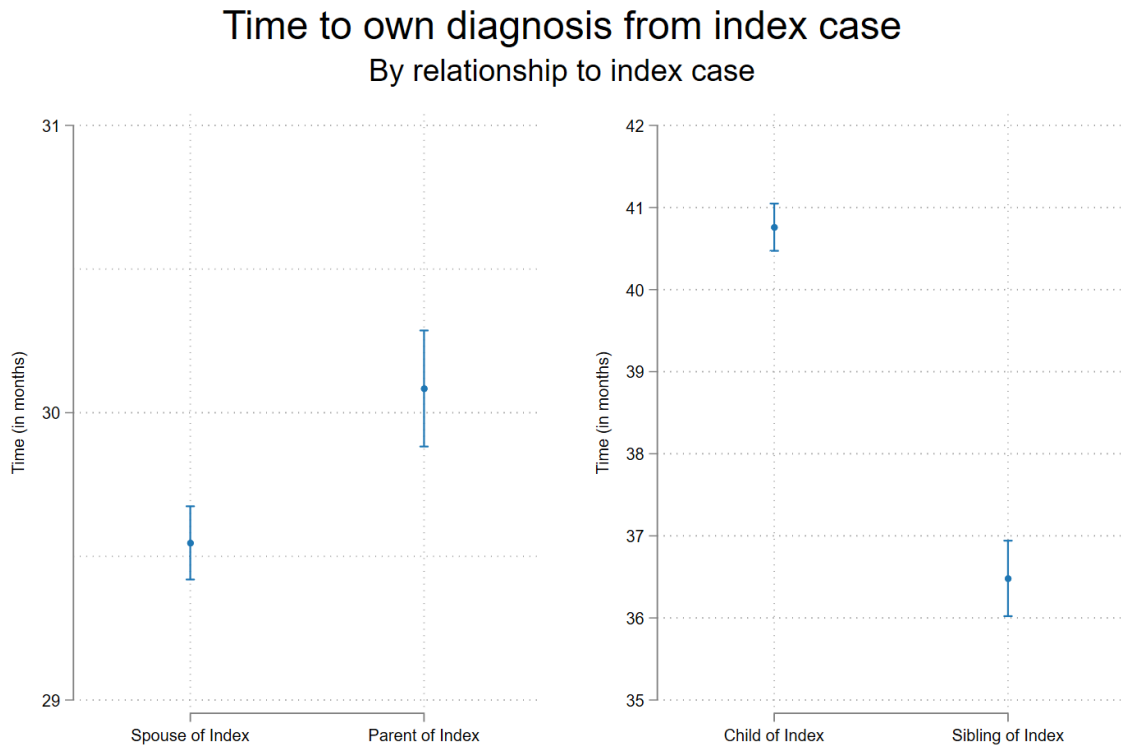
Notes: The figure plots the temporal pattern of CVD onset among family members following the index case's diagnosis, stratified by the relationship between the index case and the affected individual. Panel A displays spillovers from spouse to spouse (blue bars) and from parent to child (green bars). Panel B shows spillovers from child to parent (blue bars) and from sibling to sibling (green bars).

Figure 3: Probability of Own CVD Diagnosis Ever by Relationship to Index Case



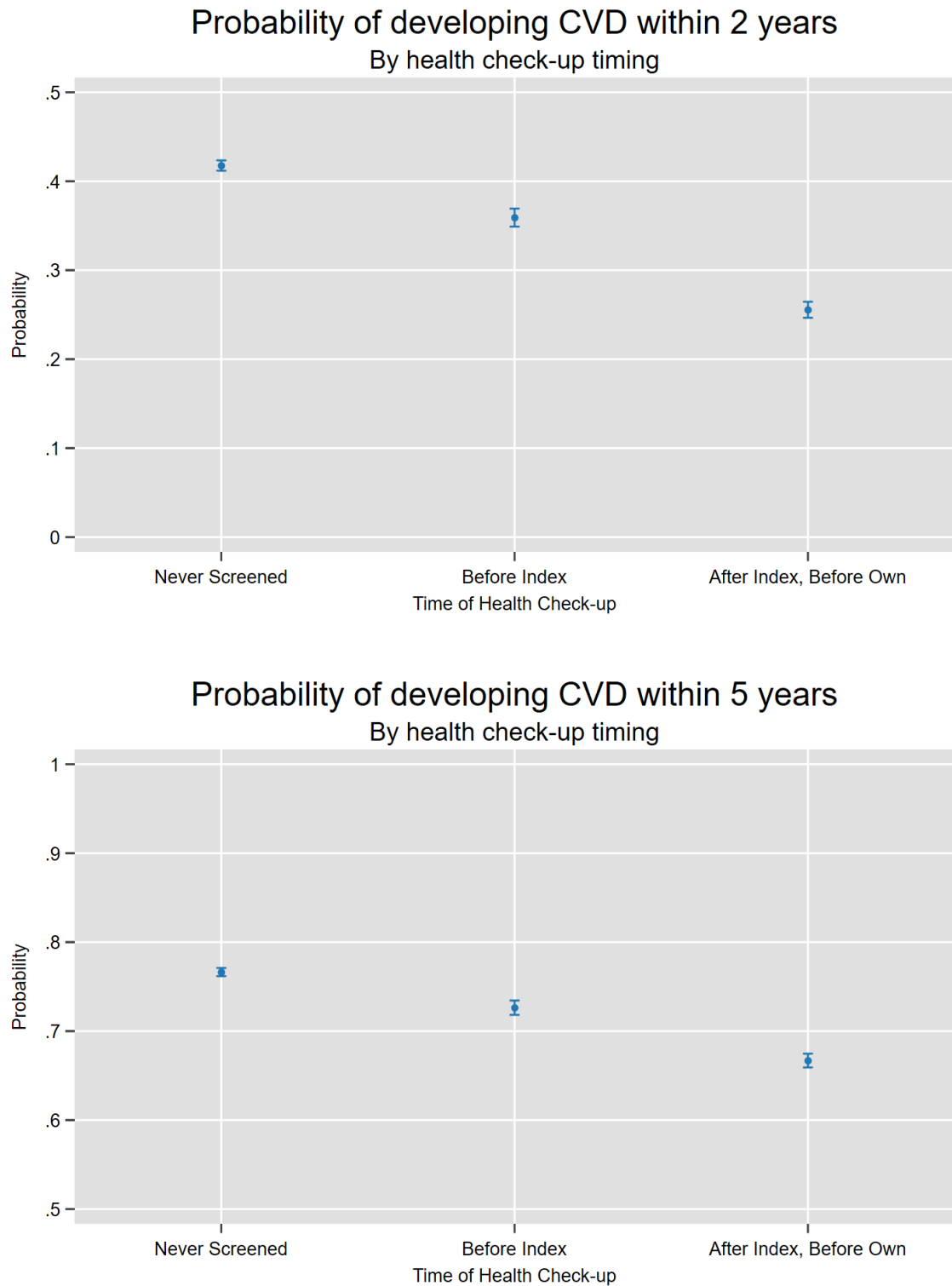
*Notes:* The figures show the coefficients and corresponding 95% confidence intervals for the probability of developing CVD based on relationship to index case.

Figure 4: Time to Own CVD Diagnosis by Relationship to Index Case



*Notes:* The figures show the coefficients and corresponding 95% confidence intervals for the probability of developing CVD based on relationship to index case.

Figure 5: Probability of developing CVD after index-case with respect to timing of own health checkup



*Notes:* The figures show the coefficients and corresponding 95% confidence intervals for the association of health checkup on being diagnosed with CVD within 2 years (Panel A) and within 5 years (Panel B) of index case within family. The comparison of proportions by period is conducted only among families within an affected member.

Table 1: Descriptive Statistics

<b>Factor</b>	<b>Value</b>
N	5,632,875
Age, mean (SD)	32.7 (21.1)
Female (1=Female, 0=Male)	46.5%
Family size, mean (SD)	3.3 (1.1)
Relationship to Index Case:	
Self	34.7%
Spouse of Index	21.4%
Child of Index	32.0%
Parent of Index	7.1%
Sibling of Index	4.8%
Other Spouse	<1%
CVD (Ever) (1=Yes, 0=No)	46.8%
Number of Family members with CVD, mean (SD)	1.4 (0.6)
Index is Female (1=Female, 0=Male)	37.2%
Number of Health Checks, mean (SD)	2.6 (3.9)
Had a health-check after index	28.3%
Health-check after index was first check	39.0%

*Notes:* This table presents the descriptive statistics for the sample which comprises of family that ever had any member diagnosed with CVD over the time-period 2005-2023. The relationship categories are defined relative to the index case (i.e., the first person in the family who develops the CVD) rather than the categories as noted in the insurance information.

Table 2: Association of Developing CVD Ever Due to Family Characteristics

	CVD Diagnosis Post Index Case			
	(1)	(2)	(3)	(4)
Age	0.007*** (0.000)	0.011*** (0.000)	0.010*** (0.000)	0.010*** (0.000)
Female	-0.026*** (0.000)	-0.020*** (0.000)	-0.020*** (0.000)	-0.020*** (0.000)
Familysize	0.006*** (0.000)	0.014*** (0.000)	0.021*** (0.000)	0.021*** (0.000)
<b>Relationship to Index:</b>				
Spouse of Index		-0.004 (0.009)	0.036*** (0.009)	0.036*** (0.009)
Child of Index		0.179*** (0.009)	0.202*** (0.009)	0.202*** (0.009)
Parent of Index		-0.015 (0.009)	0.017* (0.009)	0.017* (0.009)
Sibling of Index		0.199*** (0.009)	0.220*** (0.009)	0.220*** (0.009)
Index is Female		0.011*** (0.000)	0.009*** (0.000)	0.009*** (0.000)
Constant	-0.015*** (0.001)	-0.262*** (0.009)	-0.288*** (0.009)	-0.288*** (0.010)
Observations	3,678,047	3,678,047	3,678,047	3,678,047
Clustering	Robust	Robust	Robust	Family
Month FE	No	No	Yes	Yes
Family FE	No	No	Yes	Yes

*Notes:* This table presents estimates of developing own CVD based on the estimation of Equation 1 in Section 3 using the restricted sample of families which had any case of CVD. The dependent variable is whether the family member develops CVD after the index member in the family being diagnosed of CVD. The robust standard errors are shown in parentheses for columns (1)-(3). The standard errors are clustered at the family-level for column (4). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3: Association of Developing CVD Ever Due to Family Characteristics

	CVD Diagnosis Post Index Case			
	(1)	(2)	(3)	(4)
Age	1.055*** [1.055,1.055]	1.089*** [1.089,1.090]	1.085*** [1.084,1.085]	1.085*** [1.084,1.085]
Female	0.824*** [0.819,0.829]	0.875*** [0.869,0.881]	0.858*** [0.852,0.864]	0.858*** [0.852,0.864]
<b>Relationship to Index:</b>				
Family size	1.114*** [1.111,1.117]	1.227*** [1.223,1.231]	1.237*** [1.233,1.241]	1.237*** [1.232,1.241]
Spouse of Index		0.983 [0.880,1.098]	1.309*** [1.145,1.495]	1.309*** [1.145,1.496]
Child of Index		4.055*** [3.629,4.532]	4.728*** [4.136,5.406]	4.728*** [4.134,5.409]
Parent of Index		1.045 [0.935,1.167]	1.251*** [1.094,1.429]	1.251*** [1.094,1.430]
Sibling of Index		4.909*** [4.389,5.489]	5.660*** [4.948,6.475]	5.660*** [4.945,6.478]
Index is Female		1.043*** [1.036,1.050]	1.037*** [1.030,1.045]	1.037*** [1.029,1.046]
Observations	3,678,047	3,678,047	3,678,047	3,678,047
Clustering	Robust	Robust	Robust	Family
Month FE	No	No	Yes	Yes
Family FE	No	No	Yes	Yes

*Notes:* This table presents the Odds-Ratio (OR) and their 95% CI of developing own CVD based on logistic regression of model in Equation 1 in Section 3 using the restricted sample of families which had any case of CVD. The dependent variable is whether the family member develops CVD after the index member in the family being diagnosed of CVD. The robust standard errors are shown in parentheses for columns (1)-(3). The standard errors are clustered at the family-level for column (4). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4: Time (in months) to Developing CVD Ever Due to Family Characteristics

	Time to Own CVD Diagnosis from Index			
	(1)	(2)	(3)	(4)
Age	-0.011*** (0.002)	0.266*** (0.005)	0.247*** (0.004)	0.247*** (0.005)
Female	0.662*** (0.073)	0.109 (0.095)	-0.102 (0.091)	-0.102 (0.090)
Family size	5.507*** (0.042)	6.123*** (0.046)	5.534*** (0.043)	5.534*** (0.047)
<b>Relationship to Index:</b>				
Spouse of Index		-30.175*** (1.811)	-27.945*** (1.685)	-27.945*** (1.697)
Child of Index		-16.698*** (1.821)	-16.731*** (1.695)	-16.731*** (1.708)
Parent of Index		-27.789*** (1.813)	-27.407*** (1.687)	-27.407*** (1.701)
Sibling of Index		-20.552*** (1.830)	-21.010*** (1.704)	-21.010*** (1.718)
Index is Female		-1.812*** (0.095)	-1.954*** (0.091)	-1.954*** (0.098)
Constant	14.739*** (0.214)	28.244*** (1.851)	29.804*** (1.725)	29.804*** (1.739)
Observations	680,228	680,228	680,228	680,228
Clustering	Robust	Robust	Robust	Family
Month FE	No	No	Yes	Yes
Family FE	No	No	Yes	Yes

*Notes:* This table presents estimates of developing own CVD based on the modified [Equation 1](#) in Section 3 using the restricted sample of families which had any case of CVD. The dependent variable is the time (in months) from the diagnosis of index case to the diagnosis of the individual family member. The robust standard errors are shown in parentheses for columns (1)-(3). The standard errors are clustered at the family-level for column (4). \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 5: First health-check is after index case

Relationship to Index	Had other check (ever)	First check after Index's diagnosis	Total
Spouse of Index	16.66	83.34	100.00
Child of Index	21.17	78.83	100.00
Parent of Index	14.22	85.78	100.00
Sibling of Index	24.63	75.37	100.00
Other Spouse	44.66	55.34	100.00
Total	16.03	83.97	100.00

Table 6: Association of Developing CVD within years of index case by own health check timing after index case

	(1) 2-Year	(2) 5-Year
Health Check-up After Index	-0.133*** (0.007)	-0.081*** (0.006)
Observations	125,235	125,235
Individual Controls	Yes	Yes
Month FE	Yes	Yes
Family FE	Yes	Yes

*Notes:* This table presents estimates of the timing of health check-up on developing own CVD based on the estimation of Equation 2 in Section 4 using the restricted sample of families which had any case of CVD. The dependent variable for columns (1) and (2) is whether the family member develops CVD within 2-years or 5-years of index member in the family being diagnosed of CVD, respectively. The individual controls include birth-year, gender, and relationship to the index case. The robust standard errors are shown in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 7: Association of Developing CVD within years of index case by own health check timing

	CVD Within 2-year			CVD Within 5-year		
	(1)	(2)	(3)	(4)	(5)	(6)
Before Index	0.011*** (0.003)	-0.104*** (0.006)	-0.058*** (0.008)	0.024*** (0.003)	-0.055*** (0.005)	-0.040*** (0.006)
After Index Before Own	-0.126*** (0.003)	-0.238*** (0.005)	-0.162*** (0.007)	-0.099*** (0.003)	-0.176*** (0.005)	-0.100*** (0.006)
Observations	150383	150383	150383	150383	150383	150383
Individual Controls	No	Yes	Yes	No	Yes	Yes
Month FE	No	No	Yes	No	No	Yes
Family FE	No	No	Yes	No	No	Yes

*Notes:* This table presents estimates of the timing of health check-up on developing own CVD based using the restricted sample of families which had any case of CVD. The dependent variable for columns (1)-(3) is whether the family member develops CVD within 2-years of index member in the family being diagnosed of CVD. The time-window is extended to within 5-years for columns (4)-(6). The individual controls include birth-year, gender, and relationship to the index case. The robust standard errors are shown in parentheses. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01.