WINPEC Working Paper Series No. E2525
November 2025

Beyond Average Effects: Heterogeneous Impacts of Health Checkup and Behavioral Guidance on Health Care

Yuichi Watanabe Haruko Noguchi

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Beyond Average Effects: Heterogeneous Impacts of Health Checkup and Behavioral Guidance on Health Care

Yuichi Watanabe[*], Haruko Noguchi[†]

December 2025

[*] Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO), Chiba, Japan, Email: yuichi_watanabe@ide.go.jp
[†] Faculty of Political Science and Economics, Waseda University, Tokyo, Japan, Email: h.noguchi@waseda.jp

Abstract

There exists a globally growing concern regarding the prevention and control of non-communicable diseases (NCDs), and Japan is no exception, as lifestyle-related NCDs have a significant impact on public health. To prevent the prevalence of metabolic syndrome and control rising healthcare costs, the Japanese government initiated a novel annual health checkup program in April 2008. We examine how organized prevention programs affect healthcare outcomes, separately identifying screening effects versus behavioral intervention effects while documenting substantial heterogeneity across demographic subgroups. Using comprehensive administrative data from Japan's National Health Insurance system (FY 2011–2016), we employ instrumental variable estimation exploiting peer participation rates to address selection bias in voluntary health checkups, and difference-in-differences estimation leveraging systematic assignment rules for behavioral guidance interventions. Health checkup participation generates minimal average effects but substantial heterogeneity: younger participants (40–64 years) reduce hospitalization, while elderly participants (65–74 years) increase outpatient care expenditures. Males experience higher inpatient care costs; females significantly reduce hospitalization. Income-based heterogeneity is absent, suggesting Japan's universal coverage successfully minimizes financial barriers. Strikingly, light-touch motivational support proves more effective than intensive six-month guidance at increasing outpatient care utilization, with effects concentrated among elderly, female, and lower-income populations. These findings reveal fundamental misalignment in current program design: resource-intensive interventions target populations least responsive to behavioral guidance while the most responsive populations receive minimal support. Our results challenge conventional dose-response assumptions and have important implications for optimal prevention program design in aging societies worldwide, suggesting substantial efficiency gains through reallocation toward targeted light-touch interventions.

Keywords: Health checkups, Behavioral guidance, Health care, Heterogeneous effects, Administrative data

JEL Classification Codes: I12, I13, I18, J14

1 Introduction

Non-communicable diseases (NCDs)—including cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes—have emerged as a critical global health challenge. According to the World Health Organization (WHO), NCDs accounted for over 60% of global mortality by the mid-2000s. The WHO projected a persistent rise in the NCD prevalence and mortality throughout the 2010s due to population aging and modifiable behavioral risk factors, including tobacco consumption, unhealthy diet, physical inactivity, and excessive alcohol intake (WHO, 2005).[1] Premature deaths arising from NCDs before age 70 have become pervasive, with over 80% of such occurrences occurring in low- and middle-income countries (WHO, 2022). The global economic burden of NCDs has been estimated at around USD 30 trillion in cumulative foregone output over 2011–2030, reaching nearly USD 47 trillion when including mental illness—approximately 5% of global gross domestic product (GDP) in 2010 (Bloom et al., 2011).

Japan, a high-income nation experiencing rapid population aging, faces similar challenges, with lifestyle-related NCDs having substantial public health and economic impacts. According to the Vital Statistics and Estimates of National Medical Care Expenditure for fiscal year (FY)[2] 2019, compiled by the Ministry of Health, Labour and Welfare (MHLW), lifestyle-related diseases were responsible for 53% of deaths in Japan. Furthermore, these ailments accounted for approximately 37% and 32% of national medical expenditures for inpatient and outpatient care, respectively. While Japan has maintained a relatively strong health profile among its population under universal health insurance coverage—achieved through regular health checkups provided by schools, workplaces, and local governments, as well as high standards of hygiene and well-balanced dietary intake (Ikeda et al., 2011), the Japanese government initiated a nationwide annual health checkup program targeting individuals aged 40–74 years in April 2008 to further reduce metabolic syndrome prevalence—a precursor to NCDs—and curb rising healthcare costs. This program, known as the Specific Health Checkups (SHC) and Specific Health Guidance (SHG), introduced standardized protocols across health insurers. The SHC provides comprehensive health checkups—including measurements of body mass index (BMI), abdominal circumference,

---

[1] By the early 2020s, NCDs accounted for 74% of all deaths worldwide (WHO, 2022).
[2] Japan's fiscal year runs from April 1 to March 31 of the following year.

blood pressure, and blood/urine tests—to identify individuals at risk of metabolic syndrome, while the SHG offers tailored behavioral interventions (ranging from one-off motivational support to six-month intensive guidance) to high-risk individuals identified through systematic criteria (see Section 2 for detailed program description). This study investigates (1) the causal relationship between SHC participation and healthcare utilization or expenditures, and (2) the causal effect of SHG support on healthcare among program participants, providing critical evidence on the heterogeneous impacts of organized prevention programs in an aging universal healthcare system. While our analysis focuses on National Health Insurance (NHI) enrollees in a single Japanese municipality, our findings offer important insights for prevention program design in aging societies worldwide, as we discuss in detail in Section 2.2.

The theoretical impact of preventive measures such as health checkups and screening tests on healthcare utilization remains ambiguous. Within the framework of the health production function, prevention is generally considered as an investment aimed at augmenting or preserving individual health capital, thereby promoting well-being or utility (Kenkel, 2000; Hall, 2011). Kenkel (2000) underscored the importance of policy interventions from a social perspective to address market failures resulting in inadequate prevention due to insufficient consumer awareness, as well as externalities and moral hazards of health insurance. Nevertheless, individuals' responses to healthcare vary depending on information acquired through health checkups or screening programs. Favorable outcomes, accompanied by the absence of health warnings, may increase self-confidence regarding health, potentially reducing utilization of healthcare services. Conversely, adverse findings and risk detection may provoke anxiety, resulting in increased use of healthcare services (Hall, 2011). This theoretical ambiguity underscores the importance of empirical evidence to guide policy decisions regarding large-scale prevention programs.

A growing body of empirical literature has examined the effectiveness of preventive interventions on healthcare and health outcomes, yet results remain inconsistent across contexts and methodologies. For example, Hackl et al. (2015) found that participation in a screening program in Austria led to immediate increases in both inpatient and outpatient care costs. However, they observed no long-run effects on healthcare costs or significant effects on overall health status. Jones et al. (2019) found no causal effects of workplace wellness programs in the U.S. on medical expenditures, health behaviors, or self-reported health status, despite higher lifetime health screening rates. Regarding direct exposure to risk information or signals from national health screening and checkups, limited causal impacts were noted on increased outpatient care utilization and marginal improvement in health outcomes for higher-risk individuals in Korea (Kim et al., 2019) and Japan (Iizuka et al., 2021). However, warning signals from mandatory health checkups modified smoking and drinking behaviors in a healthier direction over the long term among middle-aged employees in Japan (Hanaoka, 2023).

Several studies have focused on the effects of the SHC policy reform in Japan. Inui et al. (2017) found negligible impacts on individual health status, behavior, and medical expenses. Conversely, Kang et al. (2021) demonstrated that SHC participation was associated with healthier lifestyle choices and longer working hours among people with lifestyle-related diseases. Likewise, Oikawa (2024) highlighted significant improvements in health behaviors and outcomes among university graduates at higher risk of metabolic syndrome following the policy reform. At the municipality level, Oikawa et al. (2025) demonstrated that expansion of per capita expenditure on the SHC program following its introduction reduced the proportion of individuals diagnosed with lifestyle-related diseases, with effects concentrated among self-employed workers and homeowners, and found the municipal response to be cost-effective. Regarding the SHG support for individuals who underwent the SHC, Suzuki et al. (2015) found modest effects on subsequent checkup results, specifically for abdominal circumference and BMI measurements. Fukuma et al. (2020) also showed a decrease in obesity status only in the short run, with no significant change in cardiovascular risk factors due to the SHG intervention for Japanese men.

Despite this accumulating evidence, fundamental questions about prevention program design and effectiveness remain unresolved, with important implications for optimal policy design in aging societies worldwide. First, the causal effect of voluntary participation in organized screening programs on healthcare utilization remains poorly identified due to endogeneity concerns. Most existing studies analyze only program participants, creating sample selection bias that confounds treatment effects with participant characteristics (Suzuki et al., 2015; Kim et al., 2019; Fukuma et al., 2020; Iizuka et al., 2021). Even population-based studies often rely on methodologies—such as regression discontinuity designs exploiting age eligibility cutoffs (Inui et al., 2017) or difference-in-differences comparing pre- and post-reform periods (Oikawa, 2024; Oikawa et al., 2025)—that cannot fully address self-selection into participation. Instrumental variable approaches remain rare, with Hackl et al. (2015) representing a notable exception using physician recommendation as an instrument. However, their supplier-induced demand strategy may not apply in contexts with greater patient autonomy in screening decisions. Second, we lack rigorous evidence separately identifying the effects of health information provision (through screening) versus behavioral intervention (through guidance programs), making it difficult to assess the marginal value of resource-intensive follow-up support. Third, the heterogeneous nature of program impacts across demographic and socioeconomic subgroups—critical for understanding treatment effect mechanisms and designing targeted interventions—remains underexplored. The limited evidence on heterogeneity raises fundamental questions: Do prevention programs generate uniform benefits across populations, or do effects vary systematically by age, gender, and socioeconomic status in ways that suggest differential program design? Fourth, the degree to which increased healthcare utilization following screening represents appropriate detection and treatment of previously undiagnosed conditions versus overutilization remains theoretically ambiguous and empirically unresolved.

This study addresses these gaps through a comprehensive individual-level analysis that makes the following four key contributions to the literature on prevention program effectiveness. We examine the causal impacts of SHC participation and SHG support on healthcare expenditures and utilization for inpatient and outpatient care services using longitudinal administrative data covering all individuals insured under the NHI program in a Japanese municipality (referred to as 'City X') for FY 2011–2016. While Oikawa et al. (2025) examined municipality-level program expansion effects on health outcomes, our individual-level analysis focuses on participation decisions and behavioral guidance interventions, providing complementary micro-level evidence on program mechanisms. The characteristics of City X are described in detail in Section 2. City X represents a typical "future model" of Japanese regional cities, where population decline has already begun and is expected to accelerate further in conjunction with rapid population aging. Accordingly, amid tightening fiscal constraints and growing concerns over the sustainability of the current health insurance system, the municipality provides an exceptionally valuable empirical setting for examining the effectiveness of preventive health programs.

First, by employing instrumental variable (IV) estimation for health checkup participation and difference-in-differences (DID) estimation for guidance interventions—both within the same population sample, we provide the first rigorous decomposition of screening effects versus guidance intervention effects. This distinction is critical because policy debates often conflate these mechanisms, yet their optimal design may differ substantially. Our findings reveal that health checkup participation and behavioral guidance operate through different channels and affect different populations, with implications for resource allocation in prevention programs worldwide.

Second, we develop a novel IV strategy exploiting regional variation in peer effects to identify the causal effect of voluntary health checkup participation while including the entire eligible population—both participants and non-participants—in our analysis. This methodological innovation addresses selection bias inherent in participant-only samples and is applicable to other contexts where health checkup participation is voluntary but potentially influenced by social networks. Unlike physician recommendation instruments that rely on supplier-induced demand (Hackl et al., 2015), our peer-effects instrument is appropriate for settings with greater patient autonomy and can be implemented wherever geographic variation in participation rates exists.

Third, we document substantial and systematic heterogeneity in prevention program effects across demographic and socioeconomic subgroups, advancing our theoretical understanding of how individuals respond to health information and behavioral interventions. Younger individuals (40–64 years) experience reduced inpatient care utilization following health checkups, while elderly individuals (65–74 years) show increased outpatient care expenditures. Males exhibit increased inpatient costs, whereas females demonstrate decreased hospitalization. However, income-based heterogeneity is absent among peer-influenced individuals, suggesting low financial barriers to appropriate care-seeking across socioeconomic groups under Japan's universal coverage system. These heterogeneous effects speak to fundamental questions about information provision, behavioral responses, and optimal program targeting.

Fourth, we demonstrate that less intensive "light-touch" motivational support significantly increases outpatient care utilization, particularly among elderly, female, and lower-income populations, while more resource-intensive continuous guidance shows limited effects. These findings challenge conventional assumptions that more intensive interventions necessarily generate larger impacts and have important implications for cost-effective program design in resource-constrained settings globally. The effectiveness of light-touch interventions suggests that barriers to appropriate care-seeking may be overcome with minimal resource investment, whereas intensive support may face diminishing returns or target populations that are already motivated to seek care.

The remainder of the paper is organized as follows: Section 2 provides an institutional background on the historical context of health checkup programs in Japan and details of the SHC/SHG system. Section 3 describes the dataset used in this study and presents summary statistics. Section 4 outlines our empirical strategies to address endogeneity concerns regarding SHC participation and to identify the treatment effects of SHG interventions among SHC participants. Section 5 presents estimation results for the main analysis using the entire sample and subsample stratification analyses. Finally, Section 6 discusses the implications and limitations of this study, including considerations of external validity and generalizability to other contexts.

2 Institutional background

2.1 Japan's health checkup system and the SHC/SHG programs

In 1972, Japan launched routine health checkup programs primarily for middle-aged and older individuals as part of its health management and promotion policy.[3] The inception of these programs initially focused on salaried workers in companies with a workforce exceeding 50 employees. Subsequently, the coverage extended to incorporate employees in smaller-sized firms and organizations. By law, all employers are mandated to conduct annual health checkups for their employees. In contrast, health checkups for other types of workers, such as the self-employed, part-time workers, the unemployed and retirees, have been facilitated by local governments through the NHI system since the 1980s. Insured residents in municipalities have the option to undergo health checkups on a voluntary basis, given that local governments are legally obliged to exert their best efforts to implement these programs. Consequently, a noticeable disparity in participation rates of health checkups emerged between salaried workers whose checkups are provided by their employers and the self-employed who were covered by local governments: the

[3] The Japanese health checkup system prior to the policy reform in 2008 is thoroughly elaborated upon in Oikawa et al. (2025), which is based on the following website: https://www.mhlw.go.jp/shingi/2005/07/s0725-7b01.html. (accessed on October 30, 2025)

former consistently exhibited substantially higher participation rates than the latter (Oikawa, 2024).

Despite the extensive history of regular health checkups in Japan, there has been a notable increase in the number of individuals diagnosed with metabolic syndrome, such as visceral fat obesity and diabetes, often categorized as lifestyle-related diseases. This surge has been accompanied by a corresponding rise in medical expenditures related to these conditions. In an effort to counteract the prevalence of metabolic syndrome and mitigate the associated healthcare costs, the Japanese government introduced a standardized annual health checkup system in 2008 for all beneficiaries aged 40–74 years. This program, known as the SHC and SHG, was implemented as a uniform nationwide initiative, irrespective of the type of public health insurance an individual holds. The SHC and SHG were designed to screen for individuals who exhibit a high risk of developing lifestyle-related diseases, as opposed to the prior health checkups that predominantly focused on those already in the early stages of these diseases (Oikawa et al., 2025). The SHC consists of a comprehensive examination, including measurements of parameters such as BMI, abdominal circumference, blood pressure, and more. Additionally, it integrates blood and urine tests to analyze various markers such as blood lipids, blood sugar, liver function, and urinary constituents like sugar and proteins. Furthermore, lifestyle habits and medication histories are recorded through questionnaires. Based on the results of the SHC, participants are assessed for whether they have metabolic syndrome or are at high risk for developing it.[4]

Following the SHC, the SHG is extended to the participants who demonstrate multiple risk factors associated with metabolic syndrome (i.e., surpassing the established thresholds of blood sugar, blood lipids and blood pressure), accompanied by abdominal obesity or overweight as determined by BMI and smoking status (Table 1).[5] The SHG consists of two distinct types of health guidance tailored to the individual's risk profile: one-off motivational support and continuous active support provided over a duration of six months. The latter support, available only to individuals under the age of 65 years, is more intensively and regularly offered than the former based on personalized behavioral goals or plans. Both types of health guidance involve

---

[4] A summary of the SHC and SHG programs is available at https://www.mhlw.go.jp/english/wp/wp-hw3/dl/2-007.pdf. (accessed on October 30, 2025)

[5] Individuals currently undergoing treatments for diabetes, hypertension or dyslipidemia are exempt from participating in the SHG.

the engagement of healthcare professionals, including doctors and public health nurses. Their role is to motivate those at high risk to instigate lifestyle modifications and improve their health status. The SHG commences with an initial counseling session immediately following the SHC and concludes with a final evaluation after the six-month support period. It is essential to note that most of the administrative and operational costs associated with the SHC and SHG are borne by local governments or employers providing these programs. Consequently, participants generally incur minimal to no out-of-pocket expenses for their involvement.[6]

While continuous active support targets only working-age individuals (40–64 years) at highest risk, one-off motivational support is provided to both working-age and elderly participants (65–74 years) who meet moderate-risk criteria. This age-stratified program design reflects policy assumptions about optimal intervention intensity across life stages.

2.2 Study setting: City X characteristics and external validity considerations

This study analyzes comprehensive administrative data from a single Japanese municipality (referred to as 'City X' hereafter to maintain confidentiality, as required by the municipality). City X is located in northern Japan, with a population of approximately 35,000 during our study period (FY 2011–2016). While single-municipality studies face inherent questions about generalizability, City X provides a particularly valuable and policy-relevant empirical setting for three key reasons.

First, City X exemplifies the demographic and economic challenges facing a rapidly growing proportion of Japanese—and increasingly, global—communities. The municipality exhibits accelerated population aging and a declining working-age population. The local economy reflects structural transition, with substantially lower secondary industry employment (18.0% vs. 25.3% nationally) and correspondingly higher tertiary sector dependence (70.8% vs. 61.2% nationally), typical patterns characteristic of deindustrialized rural areas. Critically, City X faces substantial fiscal constraints, with limited capacity to finance healthcare and prevention programs from local

---

[6] The central government has provided monetary support to public health insurers since 2013 to incentivize better health outcomes among their enrollees. This financial support is tied to the implementation results of the SHC and SHG. The amount of financial aid allocated for the medical care system for the elderly, which public health insurers are obliged to pay in a lump sum, varies according to the results achieved through the SHC and SHG. These results encompass the participation rates, the percentage of participants identified with metabolic syndrome, and the rates of improvement observed through the SHG (Oikawa et al., 2025).

tax revenue alone (fiscal capacity index, measuring the ratio of standard fiscal revenue to standard fiscal needs: 0.38 vs. 0.50 for the national municipal average). While City X has not yet reached the threshold of a "marginal community" ("*genkai shuraku*" in Japanese),[7] the municipality experienced population decline of 6.2% during 2010–2015, nearly eight times the national municipal average of –0.8%, coupled with an aging rate of 35% during our observational period. These demographic trends suggest high potential for transitioning to marginal community status in the near future and position City X as particularly valuable for understanding prevention program effectiveness in communities approaching sustainability thresholds, where early intervention may be most crucial. Understanding whether prevention programs can effectively maintain health and control healthcare costs during this transitional phase has important implications for policy design in aging societies worldwide.

Second, and relatedly, examining program effectiveness in City X provides essential evidence for policy design in resource-constrained aging societies worldwide. Rather than reflecting limitations, these characteristics make City X an ideal test case for interventions that must function effectively where pressures on healthcare system sustainability are most acute. As emphasized in the 2024 proposal by the National Governors' Association of Japan regarding healthcare and long-term care provision through 2040,[8] understanding how to deliver effective prevention programs in municipalities facing demographic and fiscal challenges is critical for ensuring healthcare system sustainability. City X represents precisely the type of setting where prevention programs face their greatest implementation challenges yet are most urgently needed, making evidence from this context particularly policy-relevant for the growing proportion of communities worldwide facing similar transitions.

Third, our dataset offers unique analytical advantages unavailable in national databases such as the National Database (NDB) or Long-Term Care Insurance Database. The administrative data include not only health insurance claims and health checkup results but also household composition and individual income information from tax records. This comprehensive data

[7] "*Genkai shuraku*" is defined by the Japanese Cabinet Office as settlements where residents aged 65 and over comprise 50% or more of the population (Cabinet Office, 2011).

[8] This proposal is available at https://www.nga.gr.jp/conference/item/a361ea7aca5a054cd5c3ee92721fef78.pdf. (accessed on October 30, 2025)

linkage enables analyses accounting for socioeconomic characteristics that are crucial for understanding heterogeneous program effects—analyses that are typically impossible with nationally representative samples. The integration of household-level economic data is particularly valuable, given that neither the NDB nor national survey data can currently capture the intersection of health outcomes, healthcare utilization, and detailed socioeconomic status at the individual level.

City X exhibits several characteristics that distinguish it from national averages: higher aging rates (35% vs. 32% nationally), lower population density (286 vs. 1,198 per km² nationally), substantial population decline (–6.2% during 2010–2015 vs. –0.8% nationally), and constrained fiscal capacity. Our analysis focuses exclusively on individuals insured under the municipal NHI, which primarily covers self-employed workers, part-time workers, the unemployed, retirees, and their dependents. While these characteristics limit generalizability to urban or younger populations, they enhance policy relevance for aging, rural, resource-constrained communities where prevention programs face their greatest challenges. Comprehensive comparisons of City X's demographic, economic, healthcare infrastructure, and baseline health characteristics with national benchmarks, along with detailed discussion of external validity considerations, are provided in Appendix A.

3 Data

3.1 Description of data

This study uses comprehensive linked administrative data from the NHI in City X, which had a population of approximately 35,000 people during the study period (FY 2011–2016). Our dataset initially contains medical care claims records for each enrollee, documented on a monthly basis. These records represent all expenditures related to medical treatment, examination, medication and other relevant healthcare expenses.[9] They include out-of-pocket expenses

[9] The fee schedule codes in medical claims records are classified into various categories such as medical care (e.g., treatment and examination), dentistry, medication, bone-setting, and home care. However, it is important to note that long-term care expenses are not covered within these codes. This study focuses on the expenses for medical care and medication regarding metabolic syndrome.

incurred by the insured.[10] For the purpose of our analysis, we aggregate the monthly data over a span of 12 months for each individual, thereby creating a yearly medical claims dataset. We then integrate this dataset using a unique individual identifier with the master enrollment data for each year. The master data provide fundamental information on enrollees' demographics, such as date of birth, gender and the periods of enrollment and withdrawal from the NHI. This integration enables us to distinguish between enrollees who are eligible to undergo the SHC and non-enrollees in each FY, allowing us to construct yearly longitudinal data at the individual enrollee level.

The panel dataset is further linked using the same unique individual identifier to the annual SHC data, individual income records from the preceding calendar year, and the roster of citizens' addresses. The SHC data pertain solely to individuals who undergo health checkups in each FY and include detailed information, including results from physical examinations and blood/urine tests, lifestyle habits, and medication histories. The income data originate from income tax records and are collected separately from those of the NHI databases such as medical claims bills and SHC records. These income records contain annually aggregated pre-tax income, inclusive of salary, pension and business revenue for each individual. Additionally, they include a unique household identifier for each year, facilitating the calculation of total household income by summing the income and pension of each member within the household. Moreover, we obtain information on residential addresses of the insured in FY 2015.[11] This information allows for the construction of a regional indicator that proves instrumental in addressing the endogeneity issue related to SHC participation, a concern elaborated upon in detail in Section 4.1.

The compiled dataset covers six FYs from 2011 to 2016 and contains a comprehensive sample of 42,310 person-years of individuals who are enrolled in the NHI throughout each FY. However,

---

[10] In Japan, patients' cost-sharing for healthcare services varies based on their age. For instance, the copayment rate for children before the commencement of compulsory education (around 6-year-old) is set at 20% of the total medical care expenses. For individuals up to the age of 69 years, the copayment rate is 30%. Those in the age bracket of 70–74 years have a copayment rate of 20% (10% before April 2014), and the elderly aged 75 years or older have a copayment rate of 10%. However, those over 70-year-old with income comparable to the current workforce have a copayment rate of 30% (https://www.mhlw.go.jp/bunya/iryouhoken/iryouhoken01/dl/01_eng.pdf).

[11] Given that residential information from the address book is available for only a single year, for the purpose of this study, we assume that citizens do not change their residence within the municipality during the entire study period. Internal migration rates in rural Japanese municipalities are typically low, particularly among older populations, which comprise the majority of NHI enrollees, supporting the plausibility of this assumption.

some observations are excluded from our analysis based on predefined criteria. First, individuals who are not listed in the address book or reside outside of City X are also excluded (856 observations). Second, individuals lacking income and pension records, as well as household identifiers, are excluded due to the inability to calculate their household income (2,137 observations). Third, singleton individuals with only one observation during the study period are excluded (1,333 observations) due to the empirical strategies that require longitudinal variation described in Section 4. Following these exclusion criteria, a final sample of 37,984 person-years is obtained, representing 8,387 unique individuals for the entire study period, forming an unbalanced panel.

Our dataset offers several key advantages over existing studies using Japanese administrative data. Most critically, our dataset links health insurance claims, health checkup results, household income, and residential location at the individual level—a combination unavailable in national databases such as the NDB or in previous municipal-level studies (Ibuka et al., 2016; Imahori et al., 2019). This comprehensive linkage enables analyses accounting for socioeconomic characteristics and geographic peer effects that are typically impossible with nationally representative samples. As discussed in Section 2.2 and Appendix A, our focus on NHI enrollees (self-employed, part-time workers, unemployed, retirees, and their dependents) in a rural municipality has implications for external validity, though this sample represents precisely the population for whom municipal prevention programs are most policy-relevant.

3.2 Outcome measures and covariates

In this study, the health-related outcome measures of primary interest are healthcare expenditures and utilization for both inpatient and outpatient care services within a FY. For the analysis of SHC participation effects (Section 5.1), these healthcare outcomes are aggregated to annual measures per FY, consistent with the year-over-year identification strategy. On the other hand, for the analysis of SHG intervention effects (Section 5.2), outcomes are computed as monthly measures within event-time windows around the SHC participation date, consistent with the DID design using monthly event studies.

Healthcare expenses consist of medical care fees and drug dispensation fees, expressed in one thousand units, convertible to monetary values through the ratio of 1,000 units equating to 10,000

JPY (i.e., approximately USD 100 at average exchange rates during our study period). Healthcare utilization represents the duration of hospitalization for inpatient care and the frequency of consultations with physicians for outpatient care. These metrics are computed as an aggregate of medical service utilization in each claims bill episode within a FY. Specifically, we restrict the scope of these outcomes to those attributable to lifestyle-related diseases and metabolic syndrome—identified through primary or secondary diagnosis codes on claims records— notably diabetes, hypertension, dyslipidemia, ischemic heart disease (e.g., angina pectoris and myocardial infarction), cerebrovascular accident (stroke) and hyperuricemia.[12]

Covariates include demographic and socioeconomic characteristics, comprising the individual's age, gender, household size, household head status, cumulative number of prior SHC participations since FY 2008, and the level of annual household income per equivalent member from the preceding calendar year. The cumulative number of prior SHC participations captures differences in individual health consciousness and prior risk detection. The equivalized household income is calculated as total household income, inclusive of public pension and business revenue, divided by the square root of household size. The resulting equivalized income is then categorized into quartile groups for positive income and organized into two distinct groups for zero and negative income, respectively,[13] for each year. Subsequently, these income groups are transformed into binary variables, with the zero-income group serving as the reference category.

3.3 Summary statistics

The insured in City X exhibit a conspicuous demographic pattern skewed toward an aging population. A significant majority, surpassing 50% of the male cohort and 60% of the female cohort, exceeds the age threshold of 60 years. Furthermore, approximately 80% of the insured falls within the age bracket of 40 to 74 years, representing the targeted demographic for the SHC

---

[12] We identified these lifestyle-related diseases and metabolic syndrome based on the 10th revision of the International Classification of Diseases (ICD-10) listed as follows: diabetes (E10–E14), hypertension (I10–I15), dyslipidemia (E78), ischemic heart disease (I20–I25), cerebrovascular disease (I60–I69), and hyperuricemia (E79).

[13] Annual pre-tax income is calculated by subtracting gross expenses/costs or various tax deductions from gross earnings including pension income and business revenue for each year in a final income tax return. Negative or zero income occurs when gross expenses/costs, inclusive of tax deductions, are greater than or equal to gross earnings. Therefore, zero-income and negative-income groups do not necessarily represent the poorest individuals.

and SHG programs (see Figure B1 in Appendix B for the complete age distribution by gender).

Table 2 provides an overview of summary statistics for annual outcome measures and covariates throughout the entire study period. The mean annual healthcare expenditures for inpatient and outpatient care services stand at 116.8 and 85.8 thousand JPY, respectively. The average duration of hospitalization for inpatient care is 4.4 days, while the frequency of physician visits for outpatient care amounts to 8.6 times annually on average. In terms of demographics, the mean age is approximately 64 years for individuals eligible for the SHC, and the average household size is 2 individuals. Approximately 45% and 60% of the sample comprise males and household heads, respectively. The equivalized annual income averages about 2.36 million JPY, including households with zero and negative income. Notably, 30% of the individual sample received the SHC during the study period, and the cumulative number of prior SHC participations since FY 2008 is merely 0.85 times on average.

Additional summary statistics are provided in Appendix B, including temporal trends in SHC participation rates from FY 2008 to FY 2019 (Figure B2), age-stratified comparisons of characteristics by SHC participation (Tables B1 and B2), and detailed monthly outcome patterns before and after SHC participation categorized by SHG intervention type (Table B3). These supplementary statistics provide additional context for our main empirical analyses and foreshadow the heterogeneous effects documented in Section 5.

4 Empirical strategies

This section presents our empirical strategies for identifying the causal effects of health checkups and behavioral guidance interventions on healthcare outcomes. We address two complementary research questions using distinct identification approaches. First, we examine the effect of voluntary SHC participation on healthcare expenditures and utilization, employing an IV strategy that exploits regional variation in peer participation rates to address self-selection bias. Second, we analyze the incremental impact of behavioral guidance (SHG support) among SHC participants, using a DID framework that leverages quasi-experimental variation from systematic assignment rules based on objective health measurements. Together, these analyses decompose the total effect of prevention programs into screening effects versus guidance intervention effects, providing critical evidence for optimal program design.

4.1 Effects of the SHC participation

4.1.1 Baseline specification and endogeneity concerns

To estimate the causal impacts of SHC participation on healthcare expenditures and utilization, we begin with the following linear panel data model:

$$Y_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta}_{\mathbf{1}} + \delta_1 H_{it} + \alpha_{1i} + \gamma_{1t} + u_{it}, \tag{1}$$

where Y_{it} represents annual healthcare outcomes (expenditures and utilization for inpatient and outpatient care services) for individual i in FY t. $\boldsymbol{X}_{it}$ denotes a vector of time-varying covariates encompassing demographic and socioeconomic characteristics: age, gender, household size, household head status, cumulative number of prior SHC participations since FY 2008, equivalized household income categories, and a binary indicator for copayment rates that distinguishes between individuals aged 70–74 years (copayment rate of 20%, or 10% before FY 2014) and all others (30%). H_{it} is a binary indicator equal to 1 if individual i participated in the SHC during period t, and 0 otherwise. We include individual fixed-effects (α_{1i}) to control for time-invariant unobserved heterogeneity and FY fixed-effects (γ_{1t}) to account for common time trends. The random error term is denoted by u_{it}.

The parameter of primary interest, δ_1, captures the effect of SHC participation on healthcare outcomes. However, ordinary least squares (OLS) estimation of equation (1) faces a fundamental endogeneity problem. Participation decisions are likely correlated with time-varying unobservable factors affecting healthcare utilization, such as changes in health consciousness, risk perceptions, and unmeasured health shocks. If individuals in better health with higher health consciousness are more likely to participate in health checkups, OLS estimates would be biased downward, confounding the true causal effect of health checkups with selection effects. Conversely, if individuals with greater perceived health risks are more likely to participate, OLS estimates would be biased upward (Hackl et al., 2015). Table 3 demonstrates this concern empirically, showing that SHC participants have significantly lower inpatient care expenditures and utilization, associations that may reflect selection rather than causal effects.

4.1.2 IV strategy

To address this endogeneity, we employ an IV estimation strategy using regional variation in SHC participation rates within narrowly defined geographic communities. The first-stage equation is specified as:

$$H_{it} = \boldsymbol{X}'_{it}\boldsymbol{\varphi} + \theta Z_{j(i)t} + \alpha_{1i} + \gamma_{1t} + \varepsilon_{it}, \tag{2}$$

where $Z_{j(i)t}$ is the IV representing the leave-one-out mean participation rate in individual i's residential block j $(j = 1, \dots, 924)$ during FY t. We construct 924 discrete local communities at the granularity of residential block numbers within City X. For each individual i residing in block j during period t, the instrument is calculated as:

$$Z_{j(i)t} = \frac{\sum_{k \in j, k \neq i} H_{kt}}{N_{jt} - 1}, \tag{3}$$

where k represents individuals residing in block j excluding the individual i and N_{jt} denotes the total number of NHI-insured residents in j during t. This leave-one-out construction ensures that individual i's own participation decision does not mechanically influence the instrument, thereby avoiding reflection problems (Manski, 1993).[14] Approximately half of our sample resides in blocks with an annual population of 10 or fewer individuals, and 90% live in blocks with populations under 50,[15] ensuring that peer groups are geographically proximate and socially meaningful. The maximum block population in any single FY is 109 individuals.

---

[14] This local participation rate is similar to the individual participation rate on average, exhibiting smaller variation among local communities than across individuals (Table 2).

[15] In instances where a block has an annual population size of only one individual, we implement a strategy of merging these blocks with neighboring ones. By consolidating these sparsely populated blocks with adjacent ones, we create larger local communities comprising more than 2 people. This consolidation enables the computation of local participation rates in the SHC for these merged communities, ensuring a meaningful assessment of the SHC participation even in cases of very small initial populations.

4.1.3 Instrument validity

The validity of our IV approach rests on two critical assumptions: instrument relevance and the exclusion restriction. For instrument relevance, our instrument must be sufficiently correlated with individual participation. The identifying assumption is that individuals' SHC participation is influenced by the participation rates within their immediate residential communities through peer effects or social network mechanisms. If more neighbors or acquaintances participate in the SHC, social pressure, information diffusion, or normative influences may increase an individual's propensity to participate.

This assumption is strongly supported by the first-stage estimates presented in Section 5.1. The coefficient on the instrument of local participation rate is $\hat{\theta} = 0.131$ (standard error = 0.018, $p < 0.01$) for contemporaneous specifications and $\hat{\theta} = 0.129$ (standard error = 0.018, $p < 0.01$) for lagged specifications, indicating that a one-percentage-point increase in the local participation rate increases an individual's probability of participating by approximately 0.13 percentage points. The Cragg-Donald Wald F-statistics are 189.87 and 180.85 for contemporaneous and lagged specifications, respectively, substantially exceeding not only Stock and Yogo (2005) critical value of 16.38 for 10% maximal IV size with a single instrument but also more recent criteria proposed by Lee et al. (2022) and Windmeijer (2022). These diagnostics confirm that our instrument strongly satisfies the relevance condition and poses no weak instrument concerns.

Regarding the exclusion restriction, the instrument must affect individual healthcare outcomes only through its effect on individual participation, not through any direct channel. This assumption would be violated if local participation rates directly influenced healthcare utilization through mechanisms other than individual screening behavior. For example, if high-participation neighborhoods had systematically different healthcare access, provider availability, or health culture that directly affected utilization patterns, the exclusion restriction would not hold.

We provide several pieces of evidence supporting the exclusion restriction. First, we verify that local participation rates are not systematically correlated with individual demographic and socioeconomic characteristics that might independently determine healthcare utilization. We confirm that the correlation between the instrument and individual demographics (age, gender, household size, household head status, income level) is negligible ($|\rho| < 0.1$). Similarly, correlations between the instrument and community-level averages of these characteristics remain small ($|\rho| < 0.2$), suggesting that participation rates vary across communities for reasons largely orthogonal to observable determinants of healthcare demand.

Second, our specification includes individual fixed effects, which absorb all time-invariant community characteristics that might be correlated with both local participation rates and individual healthcare use. The exclusion restriction therefore requires only that time-varying, community-level unobservables affecting healthcare utilization are uncorrelated with changes in local participation rates—a substantially weaker assumption than would be required in a cross-sectional design.

Third, the institutional context supports exogeneity. The SHC program is nationally standardized with uniform implementation across all blocks within City X. There is no targeting of specific neighborhoods for enhanced outreach or differential program intensity that might create confounding variation in both participation rates and healthcare access. Variation in local participation rates arises primarily from idiosyncratic differences in social networks and community norms rather than from systematic program design features.

Taken together, the strong first-stage relationship, low correlations with observables, and uniform program implementation provide compelling support for the validity of our IV strategy.

4.1.4 Interpretation: Local average treatment effects

It is important to recognize that our IV estimates identify a local average treatment effect (LATE) rather than an average treatment effect (ATE) for the full population. Specifically, our estimates capture the causal effect of SHC participation for compliers—individuals whose participation decisions are influenced by the screening behavior of their neighbors. This complier population likely differs from always-takers (who participate regardless of peer behavior) and never-takers (who never participate regardless of peer influence). However, the LATE interpretation has important implications for policy. If social network effects are an important driver of SHC participation, then policies that leverage peer influence—such as community-based outreach, neighborhood health champions, or information campaigns targeting social networks—may be particularly effective at increasing the participation rate among this population.

4.1.5 Extended specifications

To examine both immediate and delayed effects of SHC participation, we estimate equation (1) for outcomes measured in both the same FY t (contemporaneous effects) and the following FY $t+1$ (lagged effects). For the lagged specification, we replace Y_{it} with Y_{it+1} while maintaining H_{it} as the treatment variable with other covariates, allowing us to assess whether screening effects materialize with a one-year delay.

We also conduct extensive heterogeneity analysis by stratifying the sample along three dimensions. First, we examine age cohorts (40–64 years versus 65–74 years), recognizing that health checkups may have differential effects across the age distribution given different baseline health risks and copayment rates. Second, we analyze gender differences (male versus female) to assess whether men and women respond differently to health information from health checkups. Third, we investigate income heterogeneity (below versus above 50% of median equivalized income) to examine whether screening effects vary with socioeconomic status. For each subgroup analysis, we re-estimate both the first-stage and second-stage equations to ensure that the instrument remains valid within each subsample.

Given the highly skewed and zero-inflated distribution of healthcare expenditure and utilization (Table 2 shows substantial proportions of zero values), we estimate our models in levels rather than logarithms to avoid dropping zero-outcome observations. As a robustness check, we also conduct separate analyses restricted to individuals with positive healthcare outcomes, allowing us to distinguish between extensive margin effects (whether to use care services) and intensive margin effects (how much care services to use conditional on any utilization). Standard errors are clustered at two levels—the individual level and the local community (block) level—to account for both serial correlation in individual outcomes over time and spatial correlation in outcomes within communities. This two-way clustering ensures that our inference is robust to arbitrary correlation structures within individuals and geographic areas.

4.2 Impacts of the SHG support among the SHC participants

While Section 4.1 identifies the causal effect of voluntary SHC participation on healthcare outcomes, this section examines a complementary question: among those who participate in health checkups, what is the incremental impact of receiving behavioral guidance interventions

(SHG support)? This analysis addresses a critical policy question about the marginal value of resource-intensive follow-up support beyond health checkups alone.

4.2.1 Identification strategy: Quasi-experimental variation from systematic assignment

Unlike SHC participation, which is voluntary and subject to self-selection, eligibility for SHG support is determined by a systematic, rule-based assignment mechanism based on objective health checkup results. As detailed in Table 1, individuals are assigned to one of two guidance interventions—or to no guidance—according to predetermined clinical thresholds for metabolic syndrome risk factors (abdominal circumference, blood sugar, blood lipids, and blood pressure) combined with age and smoking status. Critically, these assignment rules are applied uniformly to all SHC participants based solely on their checkup results, creating quasi-experimental variation conditional on observable health measures.

This systematic assignment mechanism offers several key advantages for causal inference. First, conditional on the measured risk factors that determine eligibility, assignment to guidance support is as-good-as-random—that is, any remaining variation in assignment is orthogonal to unobserved determinants of future healthcare outcomes. Second, individuals cannot manipulate their assignment status, as eligibility is determined by biological measurements rather than self-reported information or discretionary clinical judgment. Third, the assignment rules are transparent, uniform, and consistently applied across all participants, eliminating concerns about targeting bias or differential selection into treatment.

We exploit this quasi-experimental variation using a DID framework that compares changes in healthcare utilization before and after the SHC participation between those eligible for SHG support (treatment group) and those not eligible (control group). The control group consists of SHC participants whose checkup results did not meet the clinical thresholds for SHG eligibility—specifically, individuals classified as low-risk or who met fewer than the required number of risk factor criteria. Crucially, those are individuals who underwent health checkups but did not qualify for guidance interventions. By restricting both groups to SHC participants, we isolate the incremental effect of behavioral guidance while holding constant the effect of health checkups itself.

4.2.2 Baseline DID specification

Our DID specification is presented as follows:

$$Y_{is} = \boldsymbol{X}'_{\boldsymbol{is}}\boldsymbol{\beta}_{\boldsymbol{2}} + \lambda Treat_i + \sigma Post_s + \delta_2(Treat_i \times Post_s) + \alpha_{2i} + \gamma_{2t} + \eta_m + \tau_c + \epsilon_{is}, \quad (4)$$

where Y_{is} represents monthly healthcare outcomes (expenditures and utilization for inpatient and outpatient care services) for individual i in event time s, measured in months relative to the SHC participation date. We use s to denote monthly event time (where $s = 0$ corresponds to the month of SHC participation, $s < 0$ indicates pre-participation months, and $s > 0$ signifies post-participation months), distinct from t, which denotes FY in Section 4.1.

$\boldsymbol{X}_{\boldsymbol{is}}$ represents the same vector of demographic and socioeconomic covariates as in equation (1). $Treat_i$ is a binary indicator equal to 1 if individual i is eligible for SHG support (either continuous active support or one-off motivational support, analyzed separately), and 0 otherwise. $Post_s$ is a binary indicator equal to 1 for all months after the SHC participation ($s > 0$), and 0 for all months before or during the participation month ($s \leq 0$).

We include a comprehensive set of fixed-effects to absorb various sources of variation: α_{2i} denotes individual fixed-effects, controlling for all time-invariant individual characteristics; γ_{2t} denotes FY fixed-effects, controlling for common trends affecting all individuals in a given FY; η_m denotes calendar month fixed-effects, controlling for seasonal patterns in healthcare utilization (e.g., winter flu season); τ_c denotes fixed-effects for each unique SHC participation event, which is particularly important because individuals may participate in the SHC multiple times during our study period, and each participation event generates a separate treatment episode with its own pre-post observation window.[16] ϵ_{is} denotes the random error term.

The parameter of primary interest, δ_2, captures the DID treatment effect—the differential change in healthcare outcomes after vs. before SHC participation for those eligible for SHG support relative to those not eligible. Under the identifying assumptions discussed below, this coefficient represents the causal effect of SHG eligibility on healthcare outcomes.

[16] There are 11,970 individual SHC participation events in total during our study period.

4.2.3 Identification assumptions

The validity of our DID approach rests on two key assumptions: parallel trends and no anticipation.

Regarding the parallel trends assumption, the fundamental identifying assumption is that, in the absence of SHG interventions, the treatment and control groups would have experienced common trends in healthcare outcomes. Formally, this requires that:

$$\begin{aligned} & E\left[Y_{is}^{0} \middle| Treat_i = 1, s > 0\right] - E\left[Y_{is}^{0} \middle| Treat_i = 1, s < 0\right] \\ & = E\left[Y_{is}^{0} \middle| Treat_i = 0, s > 0\right] - E\left[Y_{is}^{0} \middle| Treat_i = 0, s < 0\right], \end{aligned} \tag{5}$$

where Y_{is}^{0} denotes potential outcomes in the absence of SHG interventions. Intuitively, this assumption requires that individuals eligible and ineligible for SHG support would have followed similar healthcare trajectories had neither group received guidance.

This assumption is plausible in our context for several reasons. First, assignment to SHG support is based on predetermined clinical rules applied to objective biological measurements, not on anticipated future healthcare needs or unobservable characteristics that might predict differential trends. Second, we condition on the same set of demographic and socioeconomic covariates that determine both SHC participation and baseline healthcare utilization, substantially reducing the scope for confounding. Third, both treatment and control groups consist of SHC participants who have already demonstrated comparable health consciousness by voluntarily undergoing health checkups.

Nevertheless, the parallel trends assumption is untestable for the post-treatment period. We therefore conduct extensive pre-treatment validation using event study models (described in Section 4.2.4) to verify that treatment and control groups exhibited common trends in all observable pre-treatment months. Section 5.2 presents comprehensive validation evidence including Figures 4 and 5, which provide visual and statistical assessment of the parallel trends assumption.

For no anticipation, the second key assumption is that individuals do not anticipate their future SHG eligibility status and adjust their healthcare behavior in pre-treatment periods based on this knowledge. This assumption is highly plausible because individuals do not learn their SHG eligibility until they receive their checkup results, which occurs at or after $s = 0$. Prior to participation, individuals cannot know whether their biological measurements will exceed the clinical thresholds for guidance eligibility. This feature of our research design effectively rules out anticipation effects, strengthening the causal interpretation.

4.2.4 Event study specification and parallel trends validation

To formally test the parallel trends assumption and to examine the dynamic evolution of treatment effects, we estimate the following event study model:

$$Y_{is} = \boldsymbol{X}'_{is}\boldsymbol{\beta}_2 + \sum_{s \neq 0}^{T} \delta_{2s}(Treat_i \times D_s) + \phi_s + \alpha_{2i} + \gamma_{2t} + \eta_m + \tau_c + \epsilon_{is}, \qquad (6)$$

where D_s refers to a set of indicator variables for each monthly event time s relative to the SHC participation month ($s = 0$, which serves as the reference period), and ϕ_s represents fixed-effects for event time, capturing common patterns in healthcare outcomes relative to health checkup dates. The coefficients δ_{2s} trace out the difference in healthcare outcomes between treatment and control groups at each event time relative to the reference period. The parallel trends assumption requires that $\delta_{2s} = 0$ for all pre-treatment periods ($s < 0$). Visually, this corresponds to point estimates centered around zero with confidence intervals that consistently encompass zero throughout the pre-treatment period. If this pattern holds, then any systematic divergence in point estimates during post-treatment periods ($s > 0$) can be attributed to the causal effect of SHG interventions rather than pre-existing differential trends between treatment and control groups.

We present event study estimates graphically in Figures 4 and 5 for continuous active support and one-off motivational support, respectively. Each figure contains four panels (inpatient care expenditures, inpatient care utilization, outpatient care expenditures, and outpatient care utilization), with point estimates of δ_{2s} and 95% confidence intervals plotted for all monthly event times from $s = -12$ to $s = 12$. The vertical dashed line at $s = 0$ marks the SHC participation month, and the horizontal line at zero provides a visual reference for assessing deviations from the parallel trends assumption. Figures B3 and B4 in Appendix B provide complementary evidence by plotting raw mean outcomes over event time separately for treatment and control groups, allowing visual inspection of whether the two groups exhibit parallel trajectories in levels before the intervention.

The event study serves two purposes: (1) it validates the parallel trends assumption by testing whether pre-treatment coefficients are jointly and individually insignificant, and (2) it reveals the timing and persistence of treatment effects by showing how post-treatment coefficients evolve over the 12-month follow-up period. Based on these validation results, presented in Section 5.2, we determine the appropriate scope of inference for our DID estimates—specifically, whether the identifying assumptions are satisfied for each type of healthcare outcomes.

5 Estimation results

5.1 Effects of SHC participation

5.1.1 Main results for the entire sample

All outcome measures in this subsection—healthcare expenditures and utilization for both inpatient and outpatient care—are measured on an annual basis (per FY). Table 3 presents both OLS and IV estimation results for healthcare outcomes across the full sample. The OLS estimates indicate a statistically significant negative association between SHC participation and healthcare outcomes, primarily within the same FY. Specifically, SHC participation correlates with reduced inpatient and outpatient care expenditures (-29.94 and -5.98 thousand JPY, respectively, $p < 0.01$) and decreased inpatient utilization (-0.902 days, $p < 0.01$) in the contemporaneous period. However, these associations are not statistically significant for outpatient care utilization.

In contrast, the IV estimates reveal substantially different patterns that address endogeneity concerns inherent in voluntary participation. As anticipated in Section 4.1.3, our instrument demonstrates strong relevance: the local participation rate positively and significantly predicts individual SHC participation in the first-stage estimation (coefficients ≈ 0.13, $p < 0.01$), and the Cragg-Donald Wald F-statistics (189.87 and 180.85 for contemporaneous and lagged specifications, respectively) confirm the strength of our instrument.

The second-stage IV results show that SHC participation does not exhibit statistically significant effects on healthcare outcomes in either the same or the subsequent FY for the full sample, with one notable exception: inpatient care utilization decreases by approximately 9 days

($p < 0.1$) in the year following SHC participation. While statistically marginal, this effect is economically meaningful—representing a reduction of more than twice the sample mean of annual length of hospital stay (4.4 days in Table 2)—and suggests that for these peer-influenced individuals, the SHC may partially prevent them from developing severe conditions associated with metabolic syndrome that would necessitate hospital admission. The substantial divergence between OLS and IV estimates underscores the importance of addressing selection bias. The negative OLS associations likely reflect positive selection—healthier individuals or those with greater health consciousness are more likely both to participate in health checkups and to have lower baseline healthcare utilization. Once we instrument for SHC participation using exogenous variation in peer behavior, these selection-driven associations largely disappear, revealing that the causal effect of health checkups on compliers generates minimal impacts on average healthcare utilization in the short run.[17]

5.1.2 Heterogeneous effects across demographic and socioeconomic groups

These aggregate results, however, mask substantial heterogeneity across demographic and socioeconomic subgroups. We examine heterogeneity along three dimensions: age, gender, and household income. Table 4 summarizes the key findings across all subgroups, while Tables C1–C6 in Appendix C present detailed results for each dimension.

Age heterogeneity: Figure 1 shows the IV estimation results stratified by age group (detailed results including OLS estimates are provided in Tables C1 and C2 in Appendix C). The analysis reveals substantial heterogeneity across age cohorts that has important implications for program targeting and design. For individuals aged 40–64 years, while the OLS estimates suggest a positive association between SHC participation and outpatient care services in the same FY, the IV estimates indicate this association is not causal. However, we observe a statistically significant reduction in annual inpatient care utilization (–15.7 days, $p < 0.1$) in the subsequent FY, with no corresponding change in inpatient care expenditures. This pattern—reduced utilization without corresponding expenditure decreases—suggests that screening-induced behavioral changes

[17] Table D1 in Appendix D shows that separate IV estimation for the sample with positive healthcare outcomes demonstrates no significant effects of SHC participation on healthcare expenses and use. This result is consistent with the age cohort, gender and income group stratification analysis.

among working-age compliers prevent relatively low-cost hospital episodes rather than high-intensity acute care, and/or that limited statistical power in highly skewed expenditure data prevents detection of modest cost savings in this relatively healthy working-age population.

In contrast, for individuals aged 65–74 years, the IV estimates reveal a distinct pattern: SHC participation is associated with increased annual outpatient care expenditures (181.89 thousand JPY, $p < 0.1$) in the subsequent FY without a significant change in visit frequency. This finding is consistent with detection of previously undiagnosed conditions requiring more intensive—and costly—outpatient management. The absence of corresponding increases in utilization frequency, combined with higher per-visit expenditures, suggests that elderly participants may receive more complex diagnostic workups, specialist consultations, or pharmaceutical interventions following risk detection through health checkups. These age-stratified results highlight a fundamental tension in prevention program design: health checkups appear to shift healthcare utilization patterns differently across the age distribution, with younger participants experiencing reduced hospitalization (suggesting successful prevention of acute events) and older participants experiencing increased outpatient expenditures (suggesting detection and management of chronic conditions).

Gender heterogeneity: Gender differences prove equally striking. Figure 2 presents IV estimation results stratified by gender (detailed results including OLS estimates are provided in Tables C3 and C4 in Appendix C), revealing substantial gender differences in how individuals respond to health checkup information. For males, the IV estimates indicate that among male compliers whose participation is influenced by peer behavior, SHC participation leads to a statistically significant increase in contemporaneous annual inpatient care expenditures (314.55 thousand JPY, $p < 0.1$) without corresponding changes in utilization. This pattern—higher costs without longer hospital stay—suggests that these peer-influenced male participants may receive more intensive inpatient treatments or procedures rather than simply being hospitalized more frequently. One interpretation is that males with lifestyle-related disease risk factors detected through SHC may be channeled directly into inpatient care for acute interventions, bypassing the typical outpatient care pathway. This finding raises important questions about whether immediate hospitalization represents optimal care delivery or whether some of these inpatient episodes could be prevented through enhanced outpatient management following health checkups.

Conversely, for female compliers, the IV estimates demonstrate significantly reduced annual inpatient care utilization both contemporaneously (–17.2 days, $p < 0.05$) and in the subsequent FY (–18.4 days, $p < 0.05$), with no statistically significant changes in inpatient care expenditures. This reduction represents approximately one-third of mean annual hospital days for females in our sample, indicating substantial prevention of hospitalization among peer-influenced female participants. The absence of expenditure effects alongside reduced utilization suggests that for these female compliers, health checkups successfully prevent marginal hospitalization among females—hospital episodes that health checkup information helps individuals avoid through behavior modification or enhanced outpatient disease management. The gender-stratified findings suggest fundamentally different mechanisms through which health checkups affect healthcare outcomes for men and women, potentially reflecting gender differences in health-seeking behavior, provider practice patterns, or baseline disease severity. These results underscore the necessity of gender-specific follow-up protocols that account for these differential response patterns.

Income heterogeneity: Figure 3 presents IV estimation results stratified by household income group relative to the 50% median income threshold,[18] with detailed results provided in Tables C5 and C6 in Appendix C. Notably, the IV estimates demonstrate no statistically significant effects of SHC participation on healthcare expenditures or utilization among compliers for either income group, given sufficiently strong instruments (Cragg-Donald F-statistics exceed conventional thresholds in both subsamples). This absence of income-based heterogeneity contrasts with the substantial age and gender differences documented above, suggesting that under Japan's universal NHI system with relatively low copayment rates, financial barriers do not substantially modify screening program effects. The finding that lower-income and higher-income participants exhibit similar healthcare responses to health checkup information is consistent with Japan's success in achieving relatively equitable healthcare access across socioeconomic groups, at least among this self-employed and non-employed population covered by municipal NHI.

Synthesis of SHC participation effects: Synthesizing these findings, we draw three main conclusions regarding the causal effects of voluntary SHC participation on healthcare outcomes

---

[18] The poverty line ranges from 1,069,105 JPY to 1,151,318 JPY for each year during our study period.

among compliers—individuals whose health checkup decisions are influenced by peer participation rates in their local communities. First, for these peer-influenced individuals as a whole, voluntary SHC participation generates minimal short-run impacts on healthcare expenditures or utilization, suggesting limited immediate cost savings from the health checkup program alone for this population segment. Second, substantial heterogeneity exists across demographic subgroups of compliers: health checkup reduces hospitalization among working-age and female compliers while increasing outpatient care expenditures among elderly compliers and inpatient care costs among male compliers. Third, income-based heterogeneity is absent among compliers, suggesting that for peer-influenced individuals, Japan's universal coverage system successfully mitigates financial barriers to appropriate post-screening care-seeking across socioeconomic groups.

These patterns carry important implications for optimal program design. The heterogeneous effects suggest that uniform health checkup protocols may be suboptimal; instead, age- and gender-tailored follow-up interventions could better align health checkup information with subsequent healthcare utilization patterns. Moreover, the finding that health checkups increase certain types of healthcare outcomes (outpatient spending among elderly, inpatient costs among males) while decreasing others (hospitalization among younger individuals and females) complicates cost-effectiveness calculations—the net welfare effects depend critically on whether increased utilization reflects appropriate detection and treatment versus potentially excessive care.

These findings specifically characterize the treatment effects among compliers—a population segment that may differ systematically from always-takers (individuals who participate in health checkups regardless of peer behavior) or never-takers (individuals who never participate regardless of social influence). Compliers represent individuals who are responsive to social norms and community health behaviors, suggesting that peer-based interventions or community health promotion strategies may be particularly effective for this group. The heterogeneous effects across demographic subgroups of compliers indicate that even within this peer-responsive population, health checkup generate different healthcare behavioral responses depending on age and gender. Understanding these differential responses is crucial for designing targeted interventions that leverage social influence mechanisms.

5.2 Impacts of SHG support among SHC participants

5.2.1 Validation of identification assumptions

Before interpreting our DID estimates of behavioral guidance interventions, we must verify that the identifying assumption—parallel trends in healthcare outcomes between treatment and control groups in the absence of the intervention—holds in the pre-treatment period. Figures 4 and 5 present event study results for continuous active support and one-off motivational support, respectively, plotting estimated treatment-control differences for each month relative to SHC participation. For outpatient care outcomes (expenditures and utilization), all pre-treatment point estimates are statistically indistinguishable from zero for both types of SHG support. This pattern—flat pre-trends with no systematic divergence between treatment and control groups—provides strong evidence supporting the parallel trends assumption.

In contrast, for inpatient care outcomes, some pre-treatment coefficients achieve statistical significance, indicating that treatment and control groups exhibit differential pre-existing trends in hospitalization patterns. This violation of parallel trends for inpatient care outcomes precludes causal interpretation of DID estimates for them. The differential pre-trends may reflect that individuals meeting SHG eligibility criteria—who by definition have more metabolic risk factors—were already on different hospitalization trajectories prior to health checkups, independent of the guidance intervention itself. Consequently, we focus our analysis exclusively on outpatient care outcomes, where the identification assumption is credibly satisfied.

5.2.2 Main results: differential effectiveness by intervention intensity

With the parallel trends assumption validated for outpatient care outcomes, we turn to the main DID results. All outcome measures in this subsection—outpatient care expenditures and utilization—are measured on a monthly basis (per month during the observation window). Table 5 reports DID estimation results for both continuous active support (six-month intensive guidance for high-risk individuals under age 65) and one-off motivational support (brief counseling for moderate-risk individuals), examining effects at both 6-month and 12-month horizons. A striking finding emerges, one-off motivational support generates statistically significant increases in outpatient care outcomes, while more resource-intensive continuous active support shows minimal effects. Specifically, one-off motivational support increases monthly outpatient expenditures by 550 JPY and visit frequency by 0.081 consultations ($p < 0.01$, respectively) during the 6-month post-intervention period, with effects attenuating somewhat to 370 JPY ($p < 0.05$) and 0.06 visits ($p < 0.01$) per month over the full 12-month horizon. While these effect sizes appear modest in absolute terms—roughly one additional physician visit every 12–17 months—they represent meaningful increases relative to baseline utilization rates and suggest that even minimal behavioral guidance can influence healthcare-seeking patterns.

The ineffectiveness of continuous active support—despite its substantially greater resource intensity—presents a puzzle. One explanation emerges from examining heterogeneity across demographic groups: continuous active support targets exclusively individuals under age 65 deemed at highest risk, while one-off motivational support covers a broader age range including individuals over 65. As we demonstrate in the next subsection, the effectiveness of behavioral guidance varies dramatically across age groups, with elderly participants showing substantially greater responsiveness than younger adults. This misalignment between program intensity and population responsiveness may explain why resource-intensive interventions fail to generate measurable effects in the aggregate.

5.2.3 Heterogeneous effects across demographic groups

The aggregate results mask considerable heterogeneity in treatment responses across demographic subgroups. We examine heterogeneity along the same three dimensions as in Section 5.1.2: age, gender, and household income. Table 4 summarizes the key findings, while Tables C7–C9 present detailed results in Appendix C.

Age heterogeneity: Table C7 presents age-stratified results for one-off motivational support, revealing that the intervention's effectiveness concentrates entirely among elderly participants aged 65 years and over. Among individuals under age 65, the intervention generates almost no statistically significant effects on outpatient care. In sharp contrast, among those aged 65–74 years, one-off motivational support increases monthly outpatient expenditures by 640 JPY (6-month horizon, $p < 0.05$) and 430 JPY (12-month horizon, $p < 0.1$), with corresponding increases in visit frequency of 0.087 (6-month horizon, $p < 0.01$) and 0.06 (12-month horizon, $p < 0.01$) consultations per month. These age-specific findings help explain why continuous active support—which excludes individuals over age 65 by program design—shows minimal effectiveness: the intervention targets precisely the age group least responsive to behavioral guidance while excluding the age group most responsive. This misalignment between program intensity and population responsiveness suggests substantial scope for improved resource allocation.

Gender heterogeneity: Table C8 presents gender-stratified DID results, revealing that the effectiveness of one-off motivational support is driven primarily by female participants. For males, one-off motivational support increases physician visit frequency by 0.06 consultations per month over the 6-month horizon ($p < 0.05$) without corresponding expenditure increases. Moreover, this effect does not persist to the 12-month horizon. In contrast, for females, one-off motivational support generates substantially larger and more persistent effects: monthly outpatient expenditures increase by 1,070 JPY (6-month horizon, $p < 0.01$) and 940 JPY (12-month horizon, $p < 0.01$), with visit frequency rising by 0.125 (6-month horizon, $p < 0.01$) and 0.117 (12-month horizon, $p < 0.01$) consultations per month. Remarkably, even continuous active support—ineffective in the pooled sample and other subgroups—shows significant positive effects among females, increasing 12-month outpatient utilization by 1,020 JPY per month ($p < 0.05$) and 0.118 visits per month ($p < 0.05$).

These gender differences are substantial: female participants exhibit approximately twice the magnitude of utilization response compared to males for comparable interventions. Several mechanisms could generate this pattern. Women may be more responsive to health information or behavioral guidance, possibly due to gender differences in health literacy, risk perception, or social norms around preventive care-seeking. Alternatively, females may face fewer time or scheduling constraints in accessing outpatient care, particularly among the self-employed and non-employed NHI population studied here. Understanding the drivers of these gender differences represents an important avenue for future research with implications for intervention targeting.

Income heterogeneity: Table C9 reports income-stratified results, comparing individuals below versus above 50% of median equivalized household income. Among lower-income participants (below 50% of median), one-off motivational support significantly increases physician visit frequency by 0.116 (6-month horizon, $p < 0.05$) and 0.141 (12-month horizon, $p < 0.05$) consultations per month, with no corresponding expenditure increases. This pattern—

increased utilization without increased spending—suggests that lower-income participants may increase visits for relatively minor symptoms or preventive consultations that do not generate substantial per-visit costs. For higher-income participants (above 50% of median), one-off motivational support generates effect sizes comparable to the full-sample estimate: approximately 0.05–0.08 additional visits per month and 360–540 JPY in additional monthly expenditures. The larger utilization response among lower-income individuals, despite no expenditure effects, could reflect differential care-seeking thresholds: lower-income participants may have previously faced greater barriers (psychological or logistical rather than financial ones, given universal coverage) to accessing care services for less acute symptoms, and motivational support helps overcome these barriers. Importantly, the precision of our estimates is comparable across both income subgroups, ruling out differential statistical power as an explanation for these patterns.

Synthesis of SHG intervention effects: Several key findings emerge from this behavioral guidance analysis. First, "light-touch" one-off motivational support proves more effective at increasing outpatient care services than resource-intensive continuous active support, challenging conventional assumptions that intervention intensity monotonically increases effectiveness. Second, responsiveness to behavioral guidance concentrates among elderly participants (age 65–74), females, and somewhat among lower-income individuals, while younger adults and males show minimal responses. Third, the concentration of intervention effects among population segments excluded from or unresponsive to intensive guidance reveals fundamental misalignment in current program design.

These findings carry important policy implications. Current resource allocation—which directs intensive six-month support to high-risk individuals under age 65 while providing only brief counseling to moderate-risk elderly individuals—appears to invert the responsiveness gradient. Reallocating resources toward enhanced guidance for elderly participants, females, and potentially lower-income individuals could substantially improve program cost-effectiveness. Moreover, the effectiveness of brief motivational support suggests that behavioral interventions can shift healthcare utilization patterns without requiring intensive resource commitments, though whether increased utilization represents appropriate detection and management of previously undertreated conditions versus unnecessary care remains an open question requiring further investigation.

6 Discussion

This study examines the causal impacts of both SHC participation and SHG support on healthcare expenditures and utilization for inpatient and outpatient care services. We employ an IV strategy that exploits regional variation in peer effects and a DID estimation that leverages the quasi-experimental allocation of individuals to behavioral guidance interventions, using unique Japanese administrative data from the NHI at the individual enrollee level. Our analysis addresses two complementary research questions: (1) how does voluntary health checkup participation affect healthcare outcomes among peer-influenced individuals (compliers), and (2) what is the incremental impact of behavioral guidance interventions among health checkup participants? By examining both questions within the same population, we provide the first rigorous decomposition of screening effects versus behavioral guidance effects, while documenting substantial heterogeneity across demographic and socioeconomic subgroups that has important implications for optimal program targeting and resource allocation. We summarize our key empirical findings in the following four main contributions, discuss their implications for policy and theory, contextualize our results within the broader literature, and acknowledge important limitations.

First contribution: Decomposition of screening effects versus guidance effects

Beginning with screening participation effects, our IV estimation reveals that among compliers—individuals whose health checkup decisions are influenced by peer participation rates in their local communities—SHC participation generates minimal average effects on healthcare expenditures and utilization in both the same and the subsequent FYs for inpatient and outpatient care at the aggregate level. We find only marginal evidence that health checkups may reduce inpatient care utilization in the subsequent FY. This minimal average effect, however, masks substantial and policy-relevant heterogeneity across demographic subgroups of compliers (discussed as our second contribution below). For peer-influenced individuals as a whole, voluntary health checkup participation generates limited short-run cost savings, suggesting that implementing the SHC in City X may not be cost-effective on average for this population segment.

However, the heterogeneous impacts documented below indicate that targeted health checkup strategies focused on specific demographic groups could yield meaningful impacts. This finding prompts a reevaluation of promotion strategies from a public health policy perspective, especially considering the relatively lower participation rate in City X.

Turning to behavioral guidance interventions among health checkup participants, our DID estimation reveals a striking inversion of conventional dose-response assumptions: one-off motivational support proves effective at increasing outpatient care services, while more resource-intensive continuous active support shows minimal effects. The light-touch intervention generates statistically significant increases in both outpatient care expenditures and physician visit frequency, with effects persisting over both 6-month and 12-month horizons.

The ineffectiveness of continuous active support—despite substantially greater resource intensity—presents a puzzle that our heterogeneity analysis helps resolve. As documented in our third contribution below, continuous active support targets exclusively individuals under age 65 deemed at highest risk, while one-off motivational support covers a broader age range including individuals over 65. The misalignment between program intensity and population responsiveness (elderly participants show substantially greater responsiveness than younger adults) explains why resource-intensive interventions fail to generate measurable aggregate effects. This decomposition reveals that screening and guidance operate through fundamentally different channels and affect different populations, with critical implications for optimal program design.

Second contribution: Systematic heterogeneity across demographic subgroups

Our subsample stratification analysis uncovers distinct patterns that illuminate how health checkup information affects healthcare outcomes through different pathways across demographic groups of compliers. Individuals under the age of 65 years reduce their inpatient care utilization in the subsequent FY, while the elderly over 65 years of age increase their annual outpatient care expenditures in the year following SHC participation. This age gradient suggests fundamentally different responses to health information: working-age compliers appear to modify behaviors in ways that prevent acute care episodes, while elderly compliers translate health checkup information into intensified chronic disease management through outpatient care services.

Additionally, males tend to increase their annual inpatient care expenditures after participating in the SHC, suggesting they receive more intensive treatments rather than simply longer hospital stays. In contrast, females substantially reduce their inpatient care utilization both contemporaneously and in the subsequent FY through SHC participation. These gender differences may reflect differential healthcare-seeking patterns, varying baseline disease severity, or distinct responses to risk information between male and female compliers.[19]

The absence of income-based heterogeneity among compliers—in contrast to the substantial age and gender differences—provides evidence that Japan's universal coverage system successfully minimizes financial barriers to post-screening healthcare access, at least for peer-influenced individuals. This finding suggests that heterogeneous utilization patterns observed across age and gender reflect genuine differences in health needs, information processing, or behavioral responses rather than differential affordability constraints.

These findings emphasize the necessity of providing tailored follow-up care to SHC participants, beyond the current SHG support, considering the heterogeneous causal effects across different demographic groups. Moreover, differential intervention approaches according to their frequency and pattern of SHC participation may also be worth considering, as different patterns of participation are usually observed across demographic groups.

Third contribution: More effective light-touch interventions than intensive guidance

Examining heterogeneity in behavioral guidance effects reveals that the effectiveness of one-off motivational support concentrates among the elderly over 65 years of age, females, and the higher-income group. For females, even continuous active support generates positive effects on outpatient care services. In contrast, we find increased physician visits without corresponding increase in expenditures due to one-off motivational support among males and the lower-income group, suggesting these groups may increase visits for relatively minor symptoms or preventive consultations that do not generate substantial per-visit costs.

These patterns reveal a fundamental misalignment in current program design: the most resource-intensive intervention (continuous active support) targets precisely the population least responsive to behavioral guidance (working-age, higher-risk individuals), while the most

[19] For example, our sample shows a higher cumulative number of previous SHC participations for females than males on average, with a statistically significant difference.

responsive populations (elderly, females, and lower-income individuals) receive only minimal intervention. This inversion of the responsiveness gradient suggests substantial scope for efficiency gains through resource reallocation.

The effectiveness of light-touch interventions challenges conventional assumptions that intervention intensity monotonically translates into larger behavioral impacts. One interpretation is that motivational support addresses primarily informational and psychological barriers to care-seeking, which can be overcome with minimal resource investment, whereas continuous active support may encounter diminishing returns or target individuals who already possess high health motivation. Alternatively, the null effects of intensive guidance may reflect implementation challenges or insufficient statistical power to detect modest effects given the smaller eligible population.

These findings emphasize the necessity of providing targeted and strategically designed care to the population eligible for SHG support, including support for appropriate healthcare service utilization, accounting for heterogeneous impacts across demographic groups. Moreover, the delivery approach for continuous active support requires reexamination, as more intensive intervention for high-risk individuals may not alter their healthcare utilization patterns.

Fourth contribution: Methodological innovation addresses selection bias

Our findings contribute to and extend the growing literature on the effectiveness of preventive health interventions in several important ways. By employing a peer-effect IV strategy, we address selection bias inherent in voluntary participation while including the entire eligible population—both participants and non-participants—in the analysis. This methodological innovation is applicable to other contexts where health checkup participation is voluntary but potentially influenced by social networks.

Consistencies and differences with previous literature

Examining consistencies and differences with prior studies illuminates the mechanisms underlying the impacts of health checkup programs and clarifies the conditions under which prevention programs generate meaningful healthcare effects. Our findings regarding health checkup participation effects align broadly with recent quasi-experimental and experimental studies showing minimal average impacts. Kim et al. (2019), exploiting South Korea's National Health Screening Program, similarly found limited causal impacts on outpatient care utilization and only marginal improvements in health outcomes concentrated among higher-risk individuals. Likewise, Iizuka et al. (2021), using Japanese administrative data with regression discontinuity methods, documented minimal effects of health signal exposure on subsequent healthcare utilization. Jones et al. (2019), evaluating a randomized workplace wellness program in the United States (US), found no causal effects on medical expenditures, health behaviors, or self-reported health status over 18 months. The consistency across these diverse contexts—Korea, Japan, and the US—using different identification strategies suggests that voluntary screening programs alone generate limited short-run impacts on average healthcare utilization in populations with existing healthcare access.

However, our findings diverge from prior evidence in documenting substantial heterogeneity that reveals differentiated mechanisms. While Kim et al. (2019) found effects concentrated among higher-risk individuals, we demonstrate that heterogeneity operates primarily along demographic dimensions (age and gender) rather than baseline risk alone, with elderly compliers increasing outpatient care expenditures while younger compliers reduce hospitalization, and female compliers reducing inpatient care utilization while male compliers increase inpatient care costs. This demographic heterogeneity suggests that responses to health checkup information reflect not only baseline health risk but also differential healthcare-seeking patterns, information processing, and behavioral modification capacity across demographic subgroups.

Our results also provide important nuance to findings from regression discontinuity designs exploiting age eligibility thresholds. Inui et al. (2017) found negligible impacts of the SHC policy reform on individual health status, behavior, and medical expenses using age-discontinuity identification. In contrast, Oikawa (2024) demonstrated significant improvements in health behaviors and outcomes among university graduates at higher risk of metabolic syndrome following the policy reform, also using a DID design. These apparently conflicting results may reflect population heterogeneity: Oikawa (2024) focuses on a highly educated, younger population with strong labor market attachment, while Inui et al. (2017) and our study examine broader populations including non-employed and elderly individuals. Our finding that screening effects vary systematically by age and education-related characteristics (proxied by income level in our NHI population) reconciles these divergent results and suggests that health checkup program effectiveness depends critically on participants' characteristics.

The apparent inconsistency between our individual-level findings and the municipality-level cost-effectiveness documented by Oikawa et al. (2025) merits careful interpretation. Oikawa et al. (2025) demonstrate that expansion of per capita expenditure on the SHC program following its 2008 introduction reduced the proportion of individuals diagnosed with lifestyle-related diseases at the population level, with effects concentrated among self-employed workers and homemakers, and found the municipal response to be cost-effective. These complementary findings examine fundamentally different policy margins: Oikawa et al. (2025) identify how expanding program availability and intensity affects population health outcomes (extensive margin), while we identify how individual participation affects healthcare utilization among compliers responsive to peer influence (intensive margin). The municipal-level cost-effectiveness may reflect several mechanisms not captured in our individual participation effects: (1) general equilibrium effects of widespread program availability on health awareness and behavior even among non-participants; (2) targeting and outreach strategies that successfully recruit high-benefit individuals beyond the peer-responsive compliers we identify; (3) longer-term health outcome improvements that reduce subsequent healthcare needs; or (4) spillover effects on family members and social networks. Together, these complementary findings suggest that screening programs may generate population-level benefits through channels beyond individual participation effects, highlighting the importance of examining prevention programs at multiple levels of analysis.

Turning to behavioral guidance interventions, our finding that light-touch motivational support proves more effective than intensive continuous active support contrasts sharply with conventional dose-response assumptions. Suzuki et al. (2015) found modest effects of SHG support on subsequent checkup results (specifically abdominal circumference and BMI), while Fukuma et al. (2020) demonstrated decreased obesity status only in the short run with no significant changes in cardiovascular risk factors for Japanese men receiving intensive guidance. Our finding that intensive guidance shows minimal effects on healthcare outcomes is consistent with their evidence of limited health outcome improvements, suggesting that resource-intensive interventions may encounter diminishing returns or face implementation challenges that undermine theoretical effectiveness. The greater effectiveness of brief motivational support, particularly among elderly and female populations, suggests that barriers to appropriate healthcare utilization are often informational and psychological rather than requiring intensive behavioral modification support.

Our findings on heterogeneous guidance effects also complement Hanaoka (2023), who documented that warning signals from mandatory health checkups modified smoking and drinking behaviors in a healthier direction over the long term among middle-aged employees in Japan. The long-term behavioral modifications documented by Hanaoka (2023) among employed, working-age individuals contrast with our finding that intensive guidance is least effective for younger participants. This apparent inconsistency may reflect differences between mandatory employer-provided checkups (creating stronger institutional pressure for behavior change) versus voluntary municipal health checkup programs, or may indicate that behavioral effects manifest over longer time horizons than our 6–12 month observation window captures. The concentration of guidance effectiveness among elderly participants in our study—who face different health risks and healthcare needs than working-age employees—highlights how program responsiveness varies systematically across life stages and employment contexts.

Several methodological innovations distinguish our study from prior work and enable these new insights. First, our peer-effect IV strategy addresses selection bias while including the entire eligible population—both participants and non-participants—in the analysis, avoiding the sample selection problems inherent in participant-only studies (Suzuki et al., 2015; Kim et al., 2019; Fukuma et al., 2020; Iizuka et al., 2021). Second, we provide the first rigorous decomposition of screening effects versus behavioral guidance effects within the same population, revealing that these mechanisms operate through different channels and affect different demographic groups. Third, our comprehensive examination of heterogeneity across age, gender, and income dimensions illuminates the conditions under which prevention programs generate meaningful impacts, moving beyond average treatment effects to understand the underlying mechanisms. Fourth, our quasi-experimental identification of guidance effects exploits systematic assignment rules based on objective health measurements, providing internally valid estimates while avoiding the ethical and practical challenges of randomizing access to behavioral interventions.

Limitation and future directions

Several limitations of this study suggest important avenues for future research. First, the study focuses on a single municipality in Japan, implying that the empirical results may not be universally applicable to the entire Japanese population due to variation in demographic characteristics. As detailed in Section 2.2 and Appendix A, City X represents a particularly policy-relevant setting—an aging, fiscally constrained regional municipality where prevention programs face their greatest implementation challenges yet are most urgently needed. While this limits generalizability to younger, urban, or more affluent populations, it enhances relevance for understanding program effectiveness in precisely the contexts where healthcare sustainability concerns are most acute. Future research should examine whether our findings—particularly the age and gender heterogeneity patterns—replicate in more diverse geographic and demographic settings.

Second, the study primarily covers the self-employed, part-time workers, the unemployed, retirees, and their dependents, potentially yielding different results for those under employer-based health insurance. Our findings specifically characterize prevention program effects among the NHI population, who typically have lower health checkup participation rates and may face different barriers to healthcare access compared to employee-insured individuals. Extending this analysis to employer-based insurance populations would provide important evidence on whether the heterogeneous treatment effects we document are universal or specific to this population segment.

Third, we do not examine health outcomes themselves, as in Suzuki et al. (2015) and Fukuma et al. (2020) due to the unavailability of repeated data on health checkup results. This limitation prevents us from directly assessing whether increased healthcare utilization following health checkups reflects appropriate detection and treatment of previously undiagnosed conditions (welfare-enhancing). Without health outcome data, we cannot definitively determine whether the increased outpatient utilization we observe among elderly and female participants represents welfare-enhancing appropriate care or welfare-reducing overutilization induced by health checkups. Future research linking health checkup participation to longitudinal health outcomes—including metabolic syndrome indicators, cardiovascular events, and mortality—would provide critical evidence for welfare analysis and cost-effectiveness calculations. Among the limitations discussed here, this absence of health outcome data represents the most critical gap for policy evaluation; linking our utilization findings to longitudinal health outcomes should be the highest priority for future research, as this would enable definitive welfare conclusions.

Fourth, the analysis is confined to short-term effects due to constraints on data availability. Future research should extend the study period to explore medium-term to long-term effects on healthcare utilization, health outcomes, and medical and long-term care costs. The prevention benefits of health checkups and behavioral guidance may materialize primarily over longer time horizons as lifestyle modifications accumulate and chronic disease progression is delayed. Multi-year follow-up would also enable examination of whether the heterogeneous effects we document persist, attenuate, or amplify over time.

Despite these limitations, our study makes several important contributions. Methodologically, we provide the first application of peer-effect IV to identify screening participation effects while including the entire eligible population—both participants and non-participants—in the analysis. This approach addresses selection bias inherent in participant-only samples and is applicable to other contexts where participation is voluntary but potentially influenced by social networks. Substantively, we provide the first rigorous decomposition of screening effects versus behavioral guidance effects within the same population, revealing that these mechanisms operate through different channels and affect different demographic groups. The substantial heterogeneity we document—particularly the finding that light-touch interventions prove more effective than intensive guidance, and that program responsiveness is greatest among populations receiving minimal support—has important implications for optimal prevention program design in aging societies worldwide.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) (15H03365) and (C) (20K02230) to Mr. Nobuyuki Izumida of the National Institute of Population and Social Security Research (IPSS). We would like to sincerely express our appreciation to Mr. Nobuyuki Izumida of the IPSS, who allowed us to utilize comprehensive valuable data on the administrative records of City X. We also gratefully acknowledge local administrative staff in City X for their understanding and cooperation. Anonymity of the provided data has been strictly maintained by de-identification, and the Waseda University Institutional Review Board has approved our study (2015-063). We express our gratitude to Dr. Yukiko Ito of Keio University for discussing our paper at the 19th Annual Conference of the Japan Health Economics Association. Our special thanks go to Dr. Koryu Sato of Keio University and those who participated in the 13th International Health Economics Association (IHEA) World Congress and the European Health Economics Association (EuHEA) Conference 2024 for their helpful comments. We take full responsibility for any errors.

References

Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A. Z., Weinstein, C. (2011). The Global Economic Burden of Noncommunicable Diseases. World Economic Forum, Geneva.

Cabinet Office (2011). White Paper on Aging Society [Kōrei shakai hakusho]. Government of Japan, Tokyo. [in Japanese]

Fukuma, S., Iizuka, T., Ikenoue, T., Tsugawa, Y. (2020). Association of the National Health Guidance Intervention for Obesity and Cardiovascular Risks With Health Outcomes Among Japanese Men. *JAMA International Medicine*, 180(12), 1630–1637.

Hackl, F., Halla, M., Hummer, M., Pruckner, G.J. (2015). The Effectiveness of Health Screening. *Health Economics*, 24(8), 913–935.

Hall, J. (2011). Disease Prevention, Health Care, and Economics. In: Glied, S., Smith, P.C. (eds.) *The Oxford Handbook of Health Economics*. Oxford University Press, Oxford, 555–577.

Hanaoka, C. (2023). The effect of warning signals from health check-ups on modifiable lifestyle risk factors: evidence from mandatory health check-ups for employees in Japan. *Applied Economics Letters*, DOI: 10.1080/13504851.2023.2288031.

Ibuka, Y., Chen, S.H., Ohtsu, Y., Izumida, N. (2016). Medical Spending in Japan: An Analysis Using Administrative Data from a Citizen's Health Insurance Plan. *Fiscal Studies*, 37(3-4), 561–592.

Iizuka, T., Nishiyama, K., Chen, B., Eggleston, K. (2021). False alarm? Estimating the marginal value of health signals. *Journal of Public Economics*, 195, 104368.

Ikeda, N., Saito, E., Kondo, N., Inoue, M., Ikeda, S., Satoh, T., Wada, K., Stickley, A., Katanoda, K., Mizoue, T., Noda, M., Iso, H., Fujino, Y., Sobue, T., Tsugane, S., Naghavi, M., Ezzati, M., Shibuya, K. (2011). What has made the population of Japan healthy? *Lancet*, 378(9796), 1094–1105.

Imahori, M., Kurihara, T., Noguchi, H. (2019). Does income affect medical care and long-term care expenditures for the elderly? Evidence from claims data of a municipality in Japan. *Japanese Journal of Health Economics and Policy*, 31(1), 27–46. [in Japanese]

Inui, T., Ito, Y., Kawakami, A., Ma, X.X., Nagashima, M., Zhao, M. (2017). Empirical Study on the Utilization and Effects of Health Checkups in Japan. RIETI Discussion Paper Series 17-E-082.

Jones, D., Molitor, D., Reif, J. (2019). What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study. *Quarterly Journal of Economics*, 134(4), 1747–1791.

Kang, C., Kawamura, A., Noguchi, H. (2021). Benefits of knowing own health status: effects of health check-ups on health behaviours and labour participation. *Applied Economics Letters*, 28(11), 926–931.

Kenkel, D.S. (2000). Prevention. In: Culyer, A.J., Newhouse, J.P. (eds.) *Handbook of Health Economics*. Elsevier, Amsterdam, 1675–1720.

Kim, H.B., Lee, S.A., Lim, W. (2019). Knowing is not half the battle: Impacts of information from the National Health Screening Program in Korea. *Journal of Health Economics*, 65, 1–14.

Lee, D.S., McCrary, J., Moreira, M.J., Porter, J. (2022). Valid *t*-Ratio Inference for IV. *American Economic Review*, 112(10), 3260–3290.

Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3), 531–542.

Oikawa, M. (2024). The role of education in health policy reform outcomes: evidence from Japan. *European Journal of Health Economics*, 25 (1), 49–76.

Oikawa, M., Otake, T., Awatani, T., Noguchi, H., Kawamura, A. (2025). Impacts of health checkup programs standardization on working-age self-employed and unemployed: Insights from Japan's local government response to national policy. *Journal of Health Economics*, 103, 103046.

Stock, J.H., Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In: Andrews, D.W.K., Stock, J.H. (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, Cambridge, 80–108.

Suzuki, W., Iwamoto, Y., Yuda, M., Morozumi, R. (2015). Measuring the effect of specific health checkups and specific maintenance guidance in Japan: Evidence from program evaluation approaches. *Japanese Journal of Health Economics and Policy*, 27(1), 2–39. [in Japanese]

WHO (2005). Preventing chronic diseases: a vital investment: WHO global report. World Health

Organization, Geneva.

WHO (2022). Noncommunicable diseases progress monitor 2022. World Health Organization, Geneva.

Windmeijer, F. (2022). Weak Instruments, First-Stage Heteroskedasticity, the Robust F-Test and a GMM Estimator with the Weight Matrix Based on First-Stage Residuals. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2208.01967

Table 1 Assignment criteria for Specific Health Guidance (SHG) interventions following Specific Health Checkups (SHC)

<table>
<tr><th rowspan="2">Abdominal circumference</th><th>Additional risk factors</th><th rowspan="2">Smoking</th><th colspan="2">Types of support</th></tr>
<tr><th>Blood sugar, Blood lipids, Blood pressure</th><th>Ages 40 to 64</th><th>Ages 65 to 74</th></tr>
<tr><td rowspan="3">≥85cm (male)
≥90cm (female)</td><td>More than two factors</td><td>N/A</td><td rowspan="2">Continuous active support</td><td rowspan="3">One-off motivative support</td></tr>
<tr><td rowspan="2">One factor</td><td>Yes</td></tr>
<tr><td>No</td><td></td></tr>
<tr><td rowspan="4">Other than above
BMI≥25</td><td>Three factors</td><td>N/A</td><td rowspan="2">Continuous active support</td><td rowspan="4">One-off motivative support</td></tr>
<tr><td rowspan="2">Two factors</td><td>Yes</td></tr>
<tr><td>No</td><td rowspan="2"></td></tr>
<tr><td>One factor</td><td>N/A</td></tr>
</table>

Source: Ministry of Health, Labour and Welfare, Japan. Standards for Specific Health Checkups and Specific Health Guidance. Available at: https://www.mhlw.go.jp/file/05-Shingikai-12401000-HokenkyokuSoumuka/0000099071.pdf. (accessed October 30, 2025)

Notes: This table presents the systematic assignment rules for behavioral guidance interventions following participation in Specific Health Checkups (SHC). Assignment is determined by objective health measurements and demographic characteristics, creating quasi-experimental variation for the difference-in-differences analysis presented in Section 4.2. Additional risk factor thresholds are defined as follows: (1) Blood sugar: fasting blood sugar ≥100 mg/dl or HbA1c ≥5.2% (or ≥5.6% depending on measurement method); (2) Blood lipids: neutral lipid ≥150 mg/dl or HDL cholesterol <40 mg/dl; (3) Blood pressure: systolic ≥130 mmHg or diastolic ≥85 mmHg. Individuals currently receiving medical treatment for diabetes, hypertension, or dyslipidemia are excluded from SHG eligibility. Continuous active support involves six-month intensive guidance with personalized behavioral goals; one-off motivational support provides brief counseling. N/A = Not applicable (no guidance provided).

Table 2 Descriptive statistics:
Healthcare outcomes and individual characteristics on an annual basis (FY 2011-2016)

	N	Mean	Std. Dev.	Min	Max
Exp. for inpatient care (1K units)	37,984	11.68	63.97	0	2,075.8
Exp. for outpatient care (1K units)	37,984	8.58	29.81	0	884.44
Length of stay (inpatient)	37,984	4.36	31.91	0	369
Frequency of visits (outpatient)	37,984	8.59	14.51	0	242
Age	37,984	64.00	8.65	40	75
Gender (male=1)	37,984	0.45	0.50	0	1
Household head (=1)	37,984	0.61	0.49	0	1
Household members	37,984	2.02	0.88	0	8
Eq. income (1K JPY)	37,984	2,363.4	2,231.2	-12,832.8	69,571.5
Negative income	370	-1,025.1	1,944.9	-12,832.8	-0.21
Zero income	3,204	0	0	0	0
1st quartile	8,472	666.9	337.9	0.003	1,292.3
2nd quartile	8,617	1,871.4	353.5	1,168.7	2,506.6
3rd quartile	8,680	2,923.5	288.5	2,389.6	3,509.1
4th quartile	8,641	4,976.0	2,938.4	3,391.4	69,571.5
Num. of SHC participation	37,984	0.85	1.78	0	8
Health checkups (=1)	37,984	0.30	0.46	0	1
Local participation rate (IV)	37,984	0.30	0.23	0	1

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

Notes: Summary statistics are based on 37,984 person-years representing 8,387 unique individuals enrolled in the NHI throughout each fiscal year during the study period. Healthcare expenditures and utilization are restricted to lifestyle-related diseases and metabolic syndrome (diabetes, hypertension, dyslipidemia, ischemic heart disease, cerebrovascular accident, and hyperuricemia). Expenditures are expressed in one thousand units, convertible to monetary values at the ratio of 1,000 units = 10,000 JPY (approximately 100 USD at average exchange rates during the study period). Equivalized household income is calculated as total household income (including pension and business revenue) divided by the square root of household size, then categorized into quartile groups for positive income and two separate groups for zero and negative income. Negative or zero income occurs when gross expenses/costs (including tax deductions) are greater than or equal to gross earnings in final income tax returns; these groups do not necessarily represent the poorest individuals. "Num. of SHC participation" represents cumulative number of SHC participations since FY 2008. "Local participation rate (IV)" is the leave-one-out mean participation rate in the individual's residential block, used as the instrumental variable in Section 4.1.

Table 3 Effects of SHC participation on healthcare expenditures and utilization: OLS and IV estimates

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.994***		2.501		-0.598**		2.102	
	(0.829)		(13.748)		(0.254)		(4.711)	
Health checkup (*t*-1)		2.131*		-20.357		-0.235		4.064
		(1.135)		(17.374)		(0.370)		(6.118)
Local participation rate			0.131***	0.129***			0.131***	0.129***
(First-stage)			(0.018)	(0.018)			(0.018)	(0.018)
N of obs.	37,984	36,648	37,984	36,648	37,984	36,648	37,984	36,648
Adj. / Centered R2	0.505	0.463	0.614	0.577	0.679	0.647	0.749	0.724
Cragg-Donald Wald F stat.	–	–	189.87	180.85	–	–	189.87	180.85

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.902***		-4.351		0.074		-0.547	
	(0.298)		(4.518)		(0.150)		(1.990)	
Health checkup (*t*-1)		0.031		-9.115*		0.074		1.987
		(0.387)		(5.199)		(0.166)		(2.634)
Local participation rate			0.131***	0.129***			0.131***	0.129***
(First-stage)			(0.018)	(0.018)			(0.018)	(0.018)
N of obs.	37,984	36,648	37,984	36,648	37,984	36,648	37,984	36,648
Adj. / Centered R2	0.769	0.745	0.820	0.798	0.752	0.723	0.807	0.784
Cragg-Donald Wald F stat.	–	–	189.87	180.85	–	–	189.87	180.85

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: This table presents ordinary least squares (OLS) and instrumental variable (IV) estimation results for the causal effects of SHC participation on annual healthcare expenditures and utilization, corresponding to equations (1) and (2) in Section 4.1. The dependent variables are measured either in the same fiscal year as participation (t) or in the subsequent fiscal year (t+1). Healthcare outcomes are restricted to lifestyle-related diseases and metabolic syndrome. The IV strategy uses local participation rates (leave-one-out mean participation rate in the individual's residential block) as an instrument for individual SHC participation to address endogeneity concerns. First-stage results show that the instrument strongly predicts individual participation (coefficient ≈ 0.13, $p<0.01$). Cragg-Donald Wald F-statistics substantially exceed conventional thresholds (Stock and Yogo, 2005) and recent criteria (Lee et al., 2022; Windmeijer, 2022), confirming no weak instrument concerns. All specifications include individual fixed effects, fiscal year fixed effects, and time-varying covariates (age, gender, household size, household head status, cumulative number of prior SHC participations, equivalized household income categories, and copayment rate indicator). Robust standard errors reported in parentheses are clustered at both the individual and local community (block) levels to account for serial correlation and spatial correlation. Statistical significance: ***$p<0.01$, **$p<0.05$, *$p<0.1$.*

Table 4 Summary of heterogeneous treatment effects across demographic and socioeconomic subgroup

<table>
<tr><th colspan="2"></th><th>SHC participation effects
(IV estimates)</th><th>SHG intervention effects
(DID estimates:
one-off-motivational support)</th></tr>
<tr><td rowspan="2">Age</td><td>40-64 years</td><td>Inpatient utilization: −15.7 days/year in subsequent FY*
Outpatient care services: No significant effects</td><td>Outpatient care services: Almost no significant effects</td></tr>
<tr><td>65-74 years</td><td>Outpatient expenditures: +182 thousand JPY/year (+1,820 USD) in subsequent FY*
Outpatient utilization: No significant change
Inpatient care services: No significant effects</td><td>Outpatient expenditures: +640 JPY/month (6-month)**, +430 JPY/month (12-month)*
Outpatient utilization: +0.087 visits/month (6-month)***, +0.06 visits/month (12-month)***</td></tr>
<tr><td rowspan="2">Gender</td><td>Males</td><td>Inpatient expenditures: +315 thousand JPY/year (+3,150 USD) in same FY*
Inpatient utilization: No significant change
Outpatient care services: No significant effects</td><td>Outpatient utilization: +0.06 visits/month (6-month horizon)**
Outpatient expenditures: No significant change
Continuous active support: Almost no significant effects</td></tr>
<tr><td>Females</td><td>Inpatient utilization: −17.2 days/year (same FY); −18.4 days/year (subsequent FY)**
Inpatient expenditures: No significant change
Outpatient care services: No significant effects</td><td>Outpatient expenditures: +1,070 JPY/month (6-month)***, +940 JPY/month (12-month)***
Outpatient utilization: +0.13 visits/month (6-month)***, +0.12 visit/month (12-month)***
Continuous active support (12-month): +1,020 JPY/month**, +0.12 visits/month**</td></tr>
<tr><td rowspan="2">Income</td><td>Below 50% median income</td><td>No significant effects on any healthcare outcomes</td><td>Outpatient utilization: +0.12 visits/month (6-month)**, +0.14 visit/month (12-month)**
Outpatient expenditures: No significant change</td></tr>
<tr><td>Above 50% median income</td><td>No significant effects on any healthcare outcomes</td><td>Outpatient expenditures: +540 JPY/month (6-month)**, +360 JPY/month (12-month)*
Outpatient utilization: +0.08 visits/month (6-month)***, +0.05 visit/month (12-month)**</td></tr>
<tr><td colspan="2">Full sample (Reference)</td><td>Inpatient utilization: −9.1 days/year in subsequent FY*
All other outcomes: No significant effects</td><td>Outpatient expenditures: +550 JPY/month (6-month)***, +370 JPY/month (12-month)**
Outpatient utilization: +0.08 visits/month (6-month)***, +0.06 visits/month (12-month)***
Continuous active support: Minimal effects</td></tr>
</table>

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: This table summarizes key statistically significant findings from IV estimation (for SHC participation effects) and DID estimation (for SHG intervention effects). SHC effects are reported as annual changes in healthcare outcomes; SHG effects are reported as monthly changes. Statistical significance: ***p<0.01, **p<0.05, *p<0.1. All estimates control for individual fixed effects, fiscal year fixed effects, and demographic/socioeconomic covariates. For SHC participation, IV estimates use local participation rates as instruments (Cragg-Donald F-statistics exceed 180 for full sample and most subgroups). For SHG interventions, DID estimates compare eligible versus ineligible SHC participants; parallel trends assumption validated for outpatient outcomes (Figures 4 and 5) but not for inpatient outcomes (results for inpatient care therefore not reported). One-off motivational support results presented; continuous active support showed significant effects only for females. Currency conversions use approximate rate of 1,000 JPY ≈ 10 USD during the study period (2011–2016).*

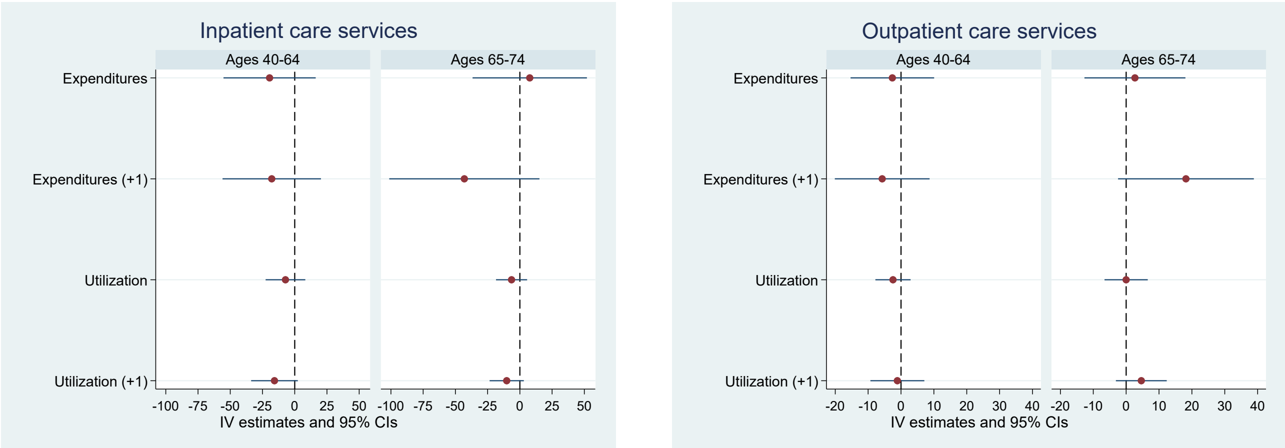
Table 5 Effects of behavioral guidance interventions on monthly healthcare outcomes: Difference-in-differences estimates

Full sample	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
Continuous active support	0.548*	0.371	0.092	0.095	0.046*	0.038*	0.056*	0.035
(under 65 years of age)	(0.327)	(0.314)	(0.056)	(0.083)	(0.027)	(0.020)	(0.030)	(0.025)
N of obs.	49,401	88,136	49,401	88,136	49,401	88,136	49,401	88,136
Adj. R2	0.020	0.034	0.174	0.189	0.546	0.544	0.665	0.648
One-off motivatioal support	0.186	0.047	0.025	0.010	0.055***	0.037**	0.081***	0.060***
	(0.196)	(0.153)	(0.032)	(0.028)	(0.021)	(0.018)	(0.020)	(0.019)
N of obs.	138,934	248,522	138,934	248,522	138,934	248,522	138,934	248,522
Adj. R2	0.014	0.019	0.128	0.118	0.340	0.346	0.539	0.521

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

*Notes: This table presents difference-in-differences (DID) estimation results for the causal effects of behavioral guidance interventions on monthly outpatient care expenditures and utilization, corresponding to equation (4) in Section 4.2. The analysis compares changes in healthcare outcomes before and after SHC participation between those eligible for SHG support (treatment group) and those not eligible (control group), both restricted to SHC participants. Two types of interventions are examined: (1) continuous active support—six-month intensive guidance for high-risk individuals under age 65; (2) one-off motivational support—brief counseling for moderate-risk individuals of all ages. Results are reported for both 6-month and 12-month post-intervention horizons. Inpatient care results are not reported because the parallel trends assumption is violated for hospitalization outcomes (see Figures 4 and 5 for validation), while parallel trends hold for outpatient care outcomes throughout the pre-treatment period. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and time-varying covariates (age, gender, household size, household head status, cumulative number of prior SHC participations, equivalized household income categories, and copayment rate indicator). Robust standard errors reported in parentheses are clustered at the individual level. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Figure 1 Heterogeneous effects of SHC participation on healthcare outcomes by age group

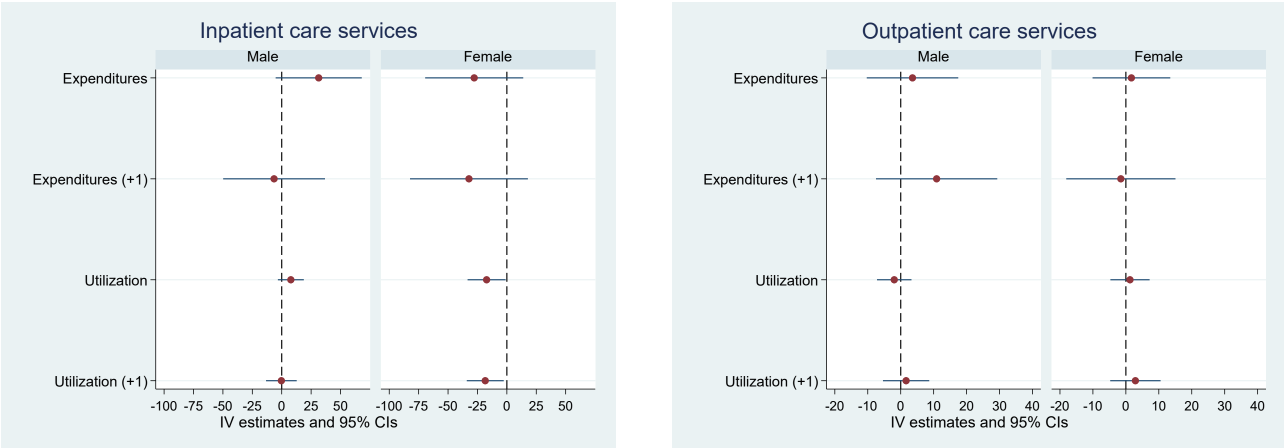


Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

Notes: This figure presents instrumental variable (IV) estimation results for the causal effects of SHC participation on annual healthcare expenditures (thousand units) and utilization (hospital days for inpatient care; number of visits for outpatient care), stratified by age group. "Ages 40–64" represents working-age participants; "Ages 65–74" represents elderly participants. "Expenditures" and "Utilization" refer to outcomes measured in the same fiscal year as SHC participation (t); "Expenditures (+1)" and "Utilization (+1)" refer to outcomes measured in the subsequent fiscal year (t+1). Point estimates are shown with 95% confidence intervals. The IV strategy uses local participation rates (leave-one-out mean participation rate in the individual's residential block) as an instrument for individual SHC participation. First-stage F-statistics exceed conventional thresholds in both age subgroups (see Tables C1 and C2 in Appendix C for detailed results). All specifications include individual fixed effects, fiscal year fixed effects, and time-varying covariates. Standard errors are clustered at both the individual and local community levels. Healthcare outcomes are restricted to lifestyle-related diseases and metabolic syndrome.

Key findings: Working-age participants (40–64 years) show reduced inpatient care utilization (–15.7 days, $p<0.1$) in the subsequent year, while elderly participants (65–74 years) exhibit increased outpatient care expenditures (+181.9 thousand JPY, $p<0.1$) in the subsequent year. Detailed discussion is provided in Section 5.1.2.

Figure 2 Heterogeneous effects of SHC participation on healthcare outcomes by gender

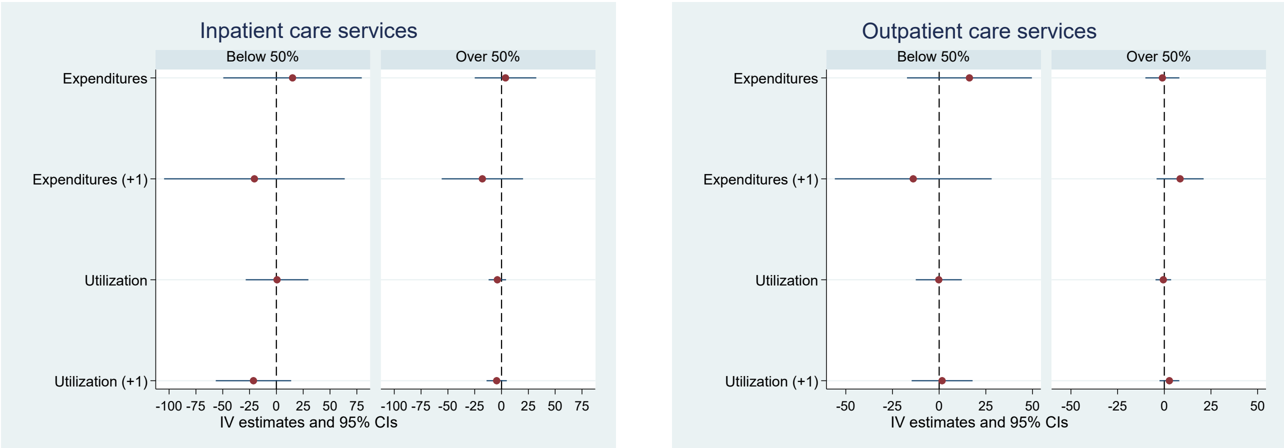


Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

Notes: This figure presents instrumental variable (IV) estimation results for the causal effects of SHC participation on annual healthcare expenditures (thousand units) and utilization (hospital days for inpatient care; number of visits for outpatient care), stratified by gender. "Expenditures" and "Utilization" refer to outcomes measured in the same fiscal year as SHC participation (t); "Expenditures (+1)" and "Utilization (+1)" refer to outcomes measured in the subsequent fiscal year (t+1). Point estimates are shown with 95% confidence intervals. The IV strategy uses local participation rates (leave-one-out mean participation rate in the individual's residential block) as an instrument for individual SHC participation. First-stage F-statistics exceed conventional thresholds in both age subgroups (see Tables C3 and C4 in Appendix C for detailed results). All specifications include individual fixed effects, fiscal year fixed effects, and time-varying covariates. Standard errors are clustered at both the individual and local community levels. Healthcare outcomes are restricted to lifestyle-related diseases and metabolic syndrome.

Key findings: Males exhibit increased inpatient care expenditures (+314.6 thousand JPY, $p<0.1$) in the same year without corresponding changes in utilization, suggesting more intensive treatments rather than longer hospital stays. Females demonstrate significantly reduced inpatient care utilization both in the same year (−17.2 days, $p<0.05$) and subsequent year (−18.4 days, $p<0.05$), without significant expenditure changes. These gender-stratified results reveal fundamentally different mechanisms through which health checkups affect healthcare utilization for men and women. Detailed discussion is provided in Section 5.1.2.

Figure 3 Heterogeneous effects of SHC participation on healthcare outcomes by income group

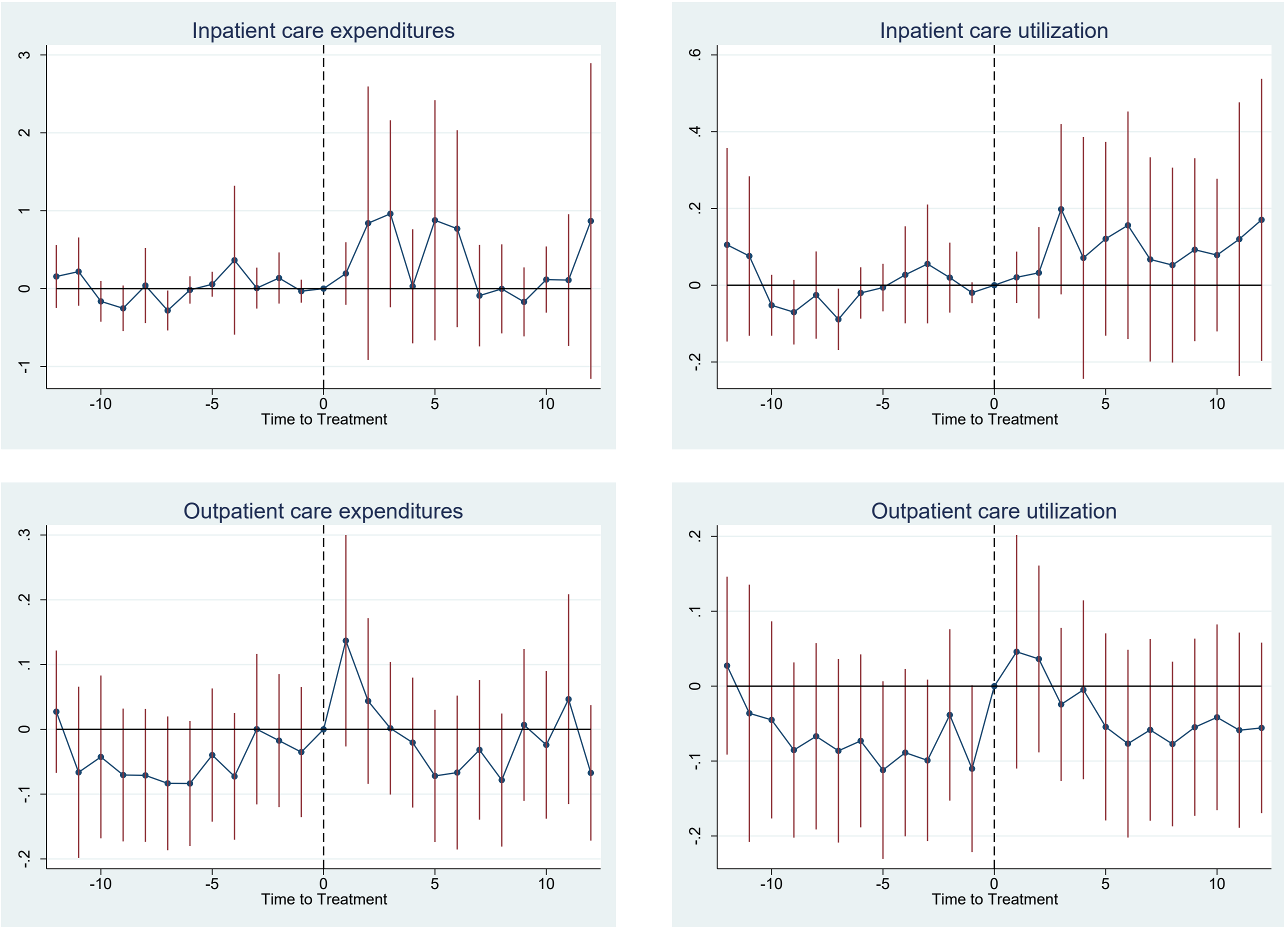


Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

Notes: This figure presents instrumental variable (IV) estimation results for the causal effects of SHC participation on annual healthcare expenditures (thousand units) and utilization (hospital days for inpatient care; number of visits for outpatient care), stratified by household income relative to the 50% median threshold. "Below 50%" represents individuals with equivalized household income below 50% of the median; "Over 50%" represents those above this threshold. "Expenditures" and "Utilization" refer to outcomes measured in the same fiscal year as SHC participation (t); "Expenditures (+1)" and "Utilization (+1)" refer to outcomes measured in the subsequent fiscal year (t+1). Point estimates are shown with 95% confidence intervals. The IV strategy uses local participation rates (leave-one-out mean participation rate in the individual's residential block) as an instrument for individual SHC participation. First-stage F-statistics exceed conventional thresholds in both age subgroups (see Tables C5 and C6 in Appendix C for detailed results). All specifications include individual fixed effects, fiscal year fixed effects, and time-varying covariates. Standard errors are clustered at both the individual and local community levels. Healthcare outcomes are restricted to lifestyle-related diseases and metabolic syndrome.

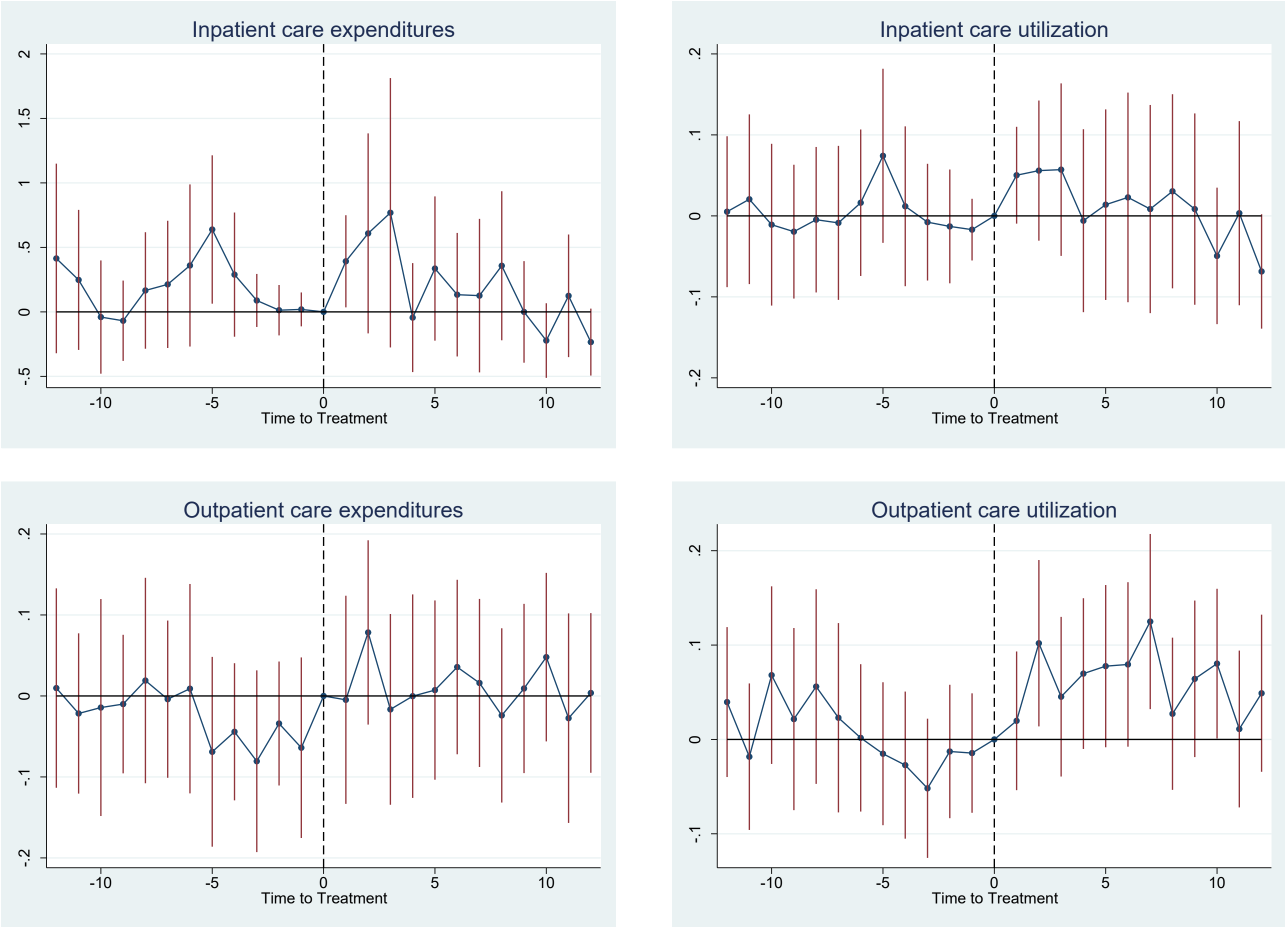
Key findings: Unlike the substantial age and gender heterogeneity documented in Figures 1 and 2, the IV estimates demonstrate no statistically significant effects of SHC participation on healthcare expenditures or utilization among compliers for either income group. This absence of income-based heterogeneity suggests that under Japan's universal NHI system with relatively low copayment rates, financial barriers do not substantially modify health checkup program effects for peer-influenced individuals. Detailed discussion is provided in Section 5.1.2.

Figure 4 Parallel trends validation for continuous active support:
Event study estimates of monthly healthcare outcomes



Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
Notes: This figure presents event study estimation results to validate the parallel trends assumption for the difference-in-differences (DID) analysis of continuous active support effects, corresponding to equation (6) in Section 4.2.4. The horizontal axis represents monthly event time relative to SHC participation (Time to Treatment = 0); the vertical axis shows point estimates of treatment-control differences with 95% confidence intervals. Continuous active support provides six-month intensive guidance to high-risk individuals under age 65. Each panel plots coefficients from separate regressions for: (1) inpatient care expenditures (thousand units per month), (2) inpatient care utilization (hospital days per month), (3) outpatient care expenditures (thousand units per month), and (4) outpatient care utilization (number of visits per month). The dashed vertical line marks the SHC participation month (event time = 0).
Critical findings: For outpatient care outcomes (bottom two panels), all pre-treatment point estimates are statistically indistinguishable from zero, providing strong support for the parallel trends assumption. In contrast, for inpatient care outcomes (top two panels), some pre-treatment coefficients achieve statistical significance, indicating violation of parallel trends. Consequently, the DID analysis in Table 5 focuses exclusively on outpatient care outcomes where the identification assumption is credibly satisfied. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and time-varying covariates. Standard errors are clustered at both the individual and SHC participation event levels. Detailed discussion of identification assumptions and their validation is provided in Sections 4.2.3 and 5.2.1.

Figure 5 Parallel trends validation for one-off motivational support: Event study estimates of monthly healthcare outcomes



Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.

Notes: This figure presents event study estimation results to validate the parallel trends assumption for the difference-in-differences (DID) analysis of one-off motivational support effects, corresponding to equation (6) in Section 4.2.4. The horizontal axis represents monthly event time relative to SHC participation (Time to Treatment = 0); the vertical axis shows point estimates of treatment-control differences with 95% confidence intervals. One-off motivational support provides brief counseling to moderate-risk individuals of all ages (40–74 years), in contrast to continuous active support which targets only high-risk individuals under age 65. Each panel plots coefficients from separate regressions for: (1) inpatient care expenditures (thousand units per month), (2) inpatient care utilization (hospital days per month), (3) outpatient care expenditures (thousand units per month), and (4) outpatient care utilization (number of visits per month). The dashed vertical line marks the SHC participation month (event time = 0).

Critical findings: For outpatient care outcomes (bottom two panels), all pre-treatment point estimates are statistically indistinguishable from zero, providing strong support for the parallel trends assumption. In contrast, for inpatient care outcomes (top two panels), some pre-treatment coefficients achieve statistical significance, indicating violation of parallel trends. Consequently, the DID analysis in Table 5 focuses exclusively on outpatient care outcomes where the identification assumption is credibly satisfied. Notably, one-off motivational support shows significant positive effects on outpatient care utilization, while continuous active support shows minimal effects in the aggregate. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and time-varying covariates. Standard errors are clustered at both the individual and SHC participation event levels. Detailed discussion of identification assumptions and their validation is provided in Sections 4.2.3 and 5.2.1.

APPENDIX A: City X's characteristics and external validity assessment

This appendix documents City X's characteristics and assesses the external validity of our findings by comparing the municipality across demographic, economic, healthcare infrastructure, and health status dimensions with national averages. This analysis clarifies the contexts to which our findings most appropriately apply and identifies the proportion of Japanese municipalities sharing City X's profile.

City X is a rural municipality with a population of approximately 33,000 in 2015—less than half the national municipal average of 74,000. The municipality experienced substantial population decline between 2010 and 2015 (–6.2%), nearly eight times the national average (–0.8%), reflecting accelerated depopulation characteristic of rural Japanese communities. Population density (286 per km²) is approximately one-quarter of the national average (1,198 per km²), with habitable land area (122.5 km²) nearly double the national average, creating particular challenges for healthcare delivery and prevention program implementation across geographically dispersed populations.

The aging rate reached 34.6% in 2015, compared to 31.9% nationally, positioning City X in the upper third of municipalities by age structure. While not yet meeting the 50% threshold for "marginal community" ("*genkai shuraku*" in Japanese) status, given rapid population decline and continuing outmigration of younger residents, City X faces high probability of transitioning to this status within the next decade. This makes City X particularly valuable for examining prevention program effectiveness during the critical transitional phase before maintaining basic social functions becomes extremely difficult. Life expectancy closely matches national averages (males: 80.2 vs. 80.6 years; females: 86.7 vs. 87.0 years), indicating that despite demographic and economic challenges, baseline longevity outcomes remain comparable nationally.

The fiscal capacity index of 0.38 compared to the national average of 0.50 demonstrates significantly limited ability to finance administrative expenses, including healthcare and prevention programs, from local tax revenue. This index measures the ratio of a municipality's standard fiscal revenue to its standard fiscal demand under Japan's Local Allocation Tax system, with values below 1.0 indicating dependence on central government transfers. City X's index places it in the lower quartile of Japanese municipalities, indicating substantial resource constraints that directly affect prevention program investment capacity.

Average annual income (2.63 million yen) is approximately 5% lower than the national average (2.77 million yen), unemployment is slightly higher (4.5% vs. 4.0%), and home ownership rates are lower (66.8% vs. 73.8%). The industrial composition reveals that primary industry employment (10.3%) is similar to the national average (11.1%), contrary to common perceptions of rural municipalities as overwhelmingly agriculture-dependent. However, secondary industry employment is substantially lower (18.0% vs. 25.3%), while tertiary sector employment is higher (70.8% vs. 61.2%), suggesting the economy has undergone structural transition away from manufacturing toward services, a pattern common in depopulating rural areas. Interestingly, public assistance recipients (11.4%) are fewer than the national average (16.7%), suggesting that despite fiscal constraints, the municipality does not exhibit exceptionally high poverty rate.

City X exhibits unexpected healthcare infrastructure patterns that merit careful interpretation. The number of general hospitals (5.7 per 100,000 population) and clinics (77.2 per 100,000) are similar to national averages (6.7 and 78.5 per 100,000, respectively). However, hospital bed capacity is dramatically higher: 2,535 beds per 100,000 population compared to the national average of 1,095 beds per 100,000—more than twice the national rate. Similarly, the number of physicians (210.2 per 100,000 population) substantially exceed the national average (157.0 per 100,000).

These seemingly paradoxical findings—higher healthcare capacity in a fiscally constrained rural municipality—likely reflect several factors: historical maintenance of relatively high hospital bed capacity due to geographic dispersion and harsh winter conditions requiring extensive inpatient care capacity; the denominator effect as population declines while hospital infrastructure remains relatively stable; and the potential role serving as a regional healthcare hub for surrounding smaller municipalities. These patterns have important implications: City X residents likely face fewer barriers to healthcare access than might be expected in a typical rural municipality, suggesting our results may not fully capture challenges of prevention program implementation in areas with genuinely limited healthcare infrastructure. This is an important caveat: our findings may represent a relatively favorable scenario for rural prevention program delivery.

Table A1 reveals striking differences in cause-specific mortality providing critical context for interpreting prevention program effectiveness. Diabetes mortality (6.1 per 100,000 population) is substantially lower than the national average (10.7 per 100,000), suggesting relatively better diabetes management or lower prevalence at baseline. However, cardiovascular disease mortality (258.9 vs. 157.8 nationally) and cerebrovascular disease mortality (137.1 vs. 90.1 nationally) are markedly higher—64% and 52% above national averages, respectively. Cancer mortality is also elevated (411.3 vs. 298.0 nationally), 38% above the national rate.

These higher mortality rates for cardiovascular and cerebrovascular diseases are particularly relevant, as these conditions are primary targets of the SHC/SHG programs designed to identify and modify metabolic syndrome risk factors. The high baseline mortality suggests: (1) City X residents face elevated cardiovascular risk, making prevention programs potentially more impactful if effective; (2) the population entering SHC/SHG programs may have more advanced disease or higher risk profiles than national averages; and (3) any observed effects on healthcare utilization may partly reflect detection and treatment of previously undiagnosed conditions rather than solely prevention of new disease.

To what proportion of Japanese municipalities is City X representative? Based on 2015 National Census data (Statistics Bureau of Japan, 2015), approximately 45% have aging rates exceeding 30%; roughly 40% have populations below 50,000; about 35% experienced population decline exceeding 3% during 2010–2015 (Statistics Bureau of Japan, 2017); and municipalities with fiscal capacity indices below 0.40 represent approximately 30% of all Japanese municipalities (Ministry of Internal Affairs and Communications, 2015). Combining these criteria, City X's profile—aging rate above 30%, population below 50,000, significant population decline, and constrained fiscal capacity—is shared by approximately 20–25% of Japanese municipalities, representing roughly 350–450 municipalities nationwide encompassing several million residents. Thus, while City X is not nationally representative in a statistical sense, it represents a substantial and growing segment of Japanese society facing acute challenges in healthcare system sustainability.

The critical policy insight is that City X represents not an outlier but a harbinger: the demographic, economic, and fiscal challenges City X faces today are challenges that an increasing proportion of municipalities in Japan and other developed countries will face in coming decades. Understanding what prevention programs can and cannot achieve in this context is essential for designing sustainable healthcare policy for aging societies.

Importantly, City X's characteristics represent a relatively favorable scenario among municipalities facing similar challenges: healthcare infrastructure remains reasonably maintained, mortality rates—while higher for some conditions—are not dramatically worse than national averages, and fiscal constraints—while substantial—are not catastrophic. If prevention programs show limited effectiveness even in this relatively favorable rural context, this suggests even greater challenges in municipalities with more severe resource constraints or healthcare access barriers.

Table A1 Characteristics of City X

Characteristics	Year	National average per municipality 1)	City X
1. Demographic indicators			
Population	2015	73,935	32,826
Change in population 2010–2015 (%)	2010–2015	–0.8	–6.2
Habitable land area (km²)	2015	71.1	122.5
Population density (per km²)	2015	1,197.5	285.7
Aging rate (65+, %)	2015	31.9	34.6
Male life expectancy	2015	80.6	80.2
Female life expectancy	2015	87.0	86.7
2. Economic indicators			
Fiscal capacity index 2)	2015	0.50	0.38
Average annual income (1,000 JYP)	2015	2,774.5	2,627.6
Ratio of primary industry workers (%)	2015	11.1	10.3
Ratio of secondary industry workers (%)	2015	25.3	18.0
Ratio of tertiary industry workers (%)	2015	61.2	70.8
Unemployment ratio (%)	2015	4.0	4.5
Public assistance recipients (%) 3)	2015	16.7	11.4
Home ownership rate (%)	2013	73.8	66.8
3. Healthcare infrastructure per 100,000 population			
General hospitals	2015	6.7	5.7
Clinics	2015	78.5	77.2
Hospital beds	2015	1,095.1	2,534.6
Physicians	2014	157.0	210.2
4. Cause-specific mortality rate per 100,000 population 4)			
Diabetes	2015	10.7	6.1
Cardiovascular disease	2015	157.8	258.9
Cerebrovascular disease	2015	90.1	137.1
Cancer	2015	298.0	411.3

Source: Statistics Bureau of Japan. (2023). Social and Demographic Statistics: Profiles of Prefectures and Municipalities. Tokyo: Ministry of Internal Affairs and Communications. https://www.e-stat.go.jp/regional-statistics/ssdsview. (accessed on October 30, 2025)

Notes:

1) *Statistics are averages of all 1,741 municipalities.*
2) *Fiscal Capacity Index: An index that measures a municipality's ability to finance its administrative expenses using its own tax revenue. It is calculated as the ratio of the municipality's standard fiscal revenue to its standard fiscal demand, as defined under Japan's Local Allocation Tax system.*
3) *Public assistance recipients: For the national average, data are from Ministry of Health, Labour and Welfare, Social Welfare and War Victims' Relief Bureau, Public Assistance Division. (2015). Survey on Recipients of Public Assistance (Basic Survey and Individual Survey). Tokyo: Ministry of Health, Labour and Welfare; https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450312&tstat=000001089175 (accessed on October 30, 2025). For City X, we utilize data from the municipal survey. The public assistance rate is defined as the number of public assistance recipients divided by the resident population based on the Basic Resident Register.*
4) *Cause-specific mortality rates: Data are from Ministry of Health, Labour and Welfare. (2015). Vital Statistics of Japan: Vital Statistics Survey. https://www.e-stat.go.jp/stat-search/database?page=1&layout=normal&toukei=00450011&tstat=000001028897&survey=%E4%BA%BA%E5%8F%A3%E5%8B%95%E6%85%8B%E8%AA%BF%E6%9F%BB&result_page=1&metadata=1&data=1. (accessed on October 10, 2025)*

APPENDIX B: Supplementary data characteristics

B.1 Overview

This appendix provides supplementary summary statistics that complement the main data description in Section 3. We present: (1) demographic patterns skewed toward an aging population, (2) temporal trends in SHC participation rates over an extended period beyond our main study window, (3) more detailed age-stratified comparisons of sample characteristics, and (4) descriptive patterns of monthly healthcare outcomes before and after SHC participation for individuals eligible for different types of SHG interventions. These supplementary statistics provide additional context for understanding our sample composition, selection patterns, and the evolution of outcomes around the SHC participation date.

B.1 A Skewed Pattern toward an Aging Population

Figure B1 shows the age distribution of the insured in City X, distinguishing between male and female cohorts during the study period. Notably, we find a conspicuous demographic pattern skewed toward an aging population within the municipal national health insurance system. A significant majority, surpassing 50% of the male cohort and 60% of the female cohort, exceeds the age threshold of 60 years. Furthermore, a predominant segment of enrollees, constituting approximately 80% of the insured, falls within the age bracket of 40 to 74 years, representing the targeted demographic for the SHC and SHG initiatives.

B.2 Temporal Trends in SHC Participation

Figure B2 shows the progression of SHC participation rates in City X from FY 2008 to FY 2019, stratified by age group. The figure covers a longer period than our main analysis (FY 2011–2016) to provide context on the evolution of the program since its nationwide introduction in April 2008.

Several notable patterns emerge from this extended time series. The participation rate among the age cohort under 65 has exhibited a consistent upward trajectory since 2011, owing to a sharp decline in the population targeted by these programs. Consequently, this participation rate has surpassed the national average in recent years. Nevertheless, the participation rate among individuals aged 65 and older has also shown an increasing trend but has consistently lagged behind the national average. It is noteworthy to emphasize that the overall participation rate within City X has remained below the national average, albeit exceeding the average within its respective prefecture.

These temporal patterns have several implications for interpreting our findings. First, the gradual increase in participation rates during our study period (2011–2016) indicates that the program was still evolving and that participation decisions were not yet stabilized. This temporal trend provides useful variation for our analysis but also suggests that our results may reflect program effects during a transitional implementation phase rather than a mature, stable program. Second, the persistently lower participation rates among the elderly (65–74 age group) compared to national averages suggest particular challenges in reaching this demographic in rural, resource-constrained settings like City X. This pattern motivates our age-stratified analyses in Section 5.1. Third, the fact that participation rates in City X consistently lag national averages but exceed the prefectural average suggests that our findings may be most representative of rural municipalities within the prefecture or similar rural regions outside of the prefecture rather than urban areas or the national average.

B.3 Age-Stratified Summary Statistics by SHC Participation

Tables B1 and B2 present mean values and standard deviations for annual outcome measures and covariates stratified by both SHC participation status and age group (40–64 years vs. 65–74 years). These more detailed stratification complements Table 2 in Section 3.3, which presents comparisons by SHC participation for the entire sample.

For the younger age cohort (40–64 years), we observe several distinctive patterns. SHC participants in this age group show significantly lower inpatient care expenditures and shorter hospital stays compared to the non-participants, with statistically significant differences. However, outpatient care expenditures and utilization show no significant differences between participants and non-participants. This pattern suggests that among working-age adults, those who participate in the SHC may have lower risk of severe conditions requiring hospitalization, possibly reflecting both health consciousness and selection effects. Additionally, younger SHC participants tend to have higher incomes and larger household sizes than younger non-participants.

For the older age cohort (65–74 years), patterns differ notably. While SHC participants still exhibit lower inpatient care expenditures and utilization on average, the differences are smaller in magnitude and statistical significance compared to the younger cohort. Interestingly, no statistically significant disparity is observed in the frequency of physician visits between elderly participants and non-participants, despite significant differences in other outcomes. These patterns suggest that selection into SHC participation may operate differently across age groups, potentially reflecting different motivations (health consciousness among younger adults vs. physician recommendation or social factors among the elderly) and different baseline health status.

These age-stratified patterns strongly foreshadow the heterogeneous effects we document in our main analysis (Section 5.1). The distinct characteristics of SHC participants across age groups suggest that the causal effects of SHC participation may differ systematically by age, motivating our age-stratified instrumental variable analyses. The fact that selection patterns differ by age also underscores the importance of our instrumental variable (IV) strategy: simple comparisons between participants and non-participants would confound causal effects with these complex selection patterns.

B.4 Pre/Post SHC Patterns for SHG Analysis

Table B3 provides average values with standard deviations for monthly outcome measures over the period of one year before and after SHC participation, categorized by SHG intervention type (continuous active support vs. one-off motivational support). Figures B3 and B4 present graphical representations of these monthly outcome trends for inpatient and outpatient care services, respectively.

Sample composition: Our SHC data contain 4,417 unique participants during the entire study period, including 285 and 753 unique individuals eligible for continuous active support and one-off motivational support, respectively. These two groups are mutually exclusive—individuals can only be eligible for one type of SHG interventions based on systematic criteria related to their SHC results (see Table 1 in Section 2.1 for detailed eligibility criteria).

Patterns for one-off motivational support: SHC participants who are eligible for one-off motivational support show significantly lower outpatient care expenses and utilization on average compared to those not eligible for any SHG intervention, both before and after SHC participation. However, no statistically significant differences in inpatient care services are found between these groups. The relatively stable differences before and after SHC suggest that individuals selected for motivational support differ systematically in their baseline healthcare utilization patterns, which we account for in our difference-in-differences (DID) analysis in Section 5.2.

Patterns for continuous active support: SHC participants eligible for continuous active support (available only to those under age 65 with higher risk profiles) exhibit a more complex pattern. Before SHC participation, this group shows significantly lower inpatient care expenditures and utilization on average compared to non-eligible participants. However, after SHC participation, they show significantly higher inpatient care expenditures and utilization. Importantly, no statistically significant differences are obtained when comparing the entire periods before and after SHC participation, suggesting that the changes may reflect detection and treatment of previously undiagnosed conditions rather than deterioration in health status.

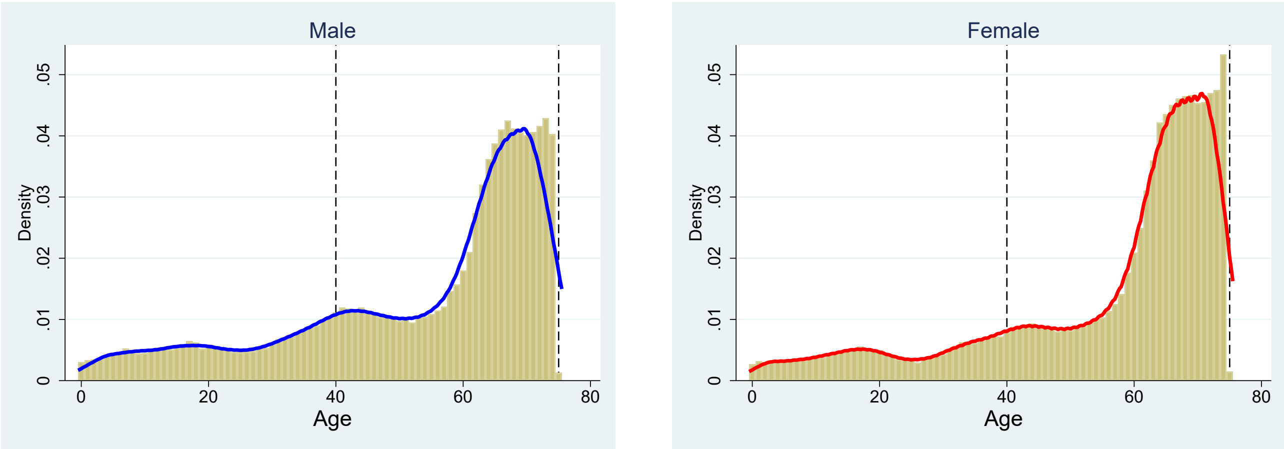
Common pattern across all groups: A critical observation from Table B3 and Figures B3 and B4 is that all average values of monthly outcome measures increase after SHC participation for both treatment groups (eligible for SHG) and control groups (not eligible for SHG), regardless of the type of SHG support. This universal increase in healthcare outcomes following SHC participation likely reflects several mechanisms: (1)

detection of previously undiagnosed conditions prompting medical consultation, (2) increased health awareness leading to care-seeking, and (3) physician recommendations following checkup results. This common trend emphasizes the importance of our DID approach, which compares the change in outcomes between treatment and control groups rather than relying on simple pre/post comparisons.

Implications for parallel trends assumption: The graphical evidence in Figures B3 and B4 provides preliminary assessment of whether treatment and control groups exhibit similar trends before SHC participation—a key identifying assumption for our DID analysis. Visual inspection suggests broadly similar pre-SHC trends for outpatient care outcomes, supporting the parallel trends assumption for these outcomes. However, some divergence in pre-trends is visible for inpatient care outcomes, particularly for continuous active support. We formally test the parallel trends assumption using event study models in Section 5.2 (Figures 4 and 5), which allow for statistical assessment of whether pre-treatment differences between two groups are significant. As discussed in Section 5.2, we find that the parallel trends assumption is reasonably satisfied for outpatient care outcomes but may be violated for inpatient care outcomes, leading us to focus our interpretation of SHG effects primarily on outpatient care services.

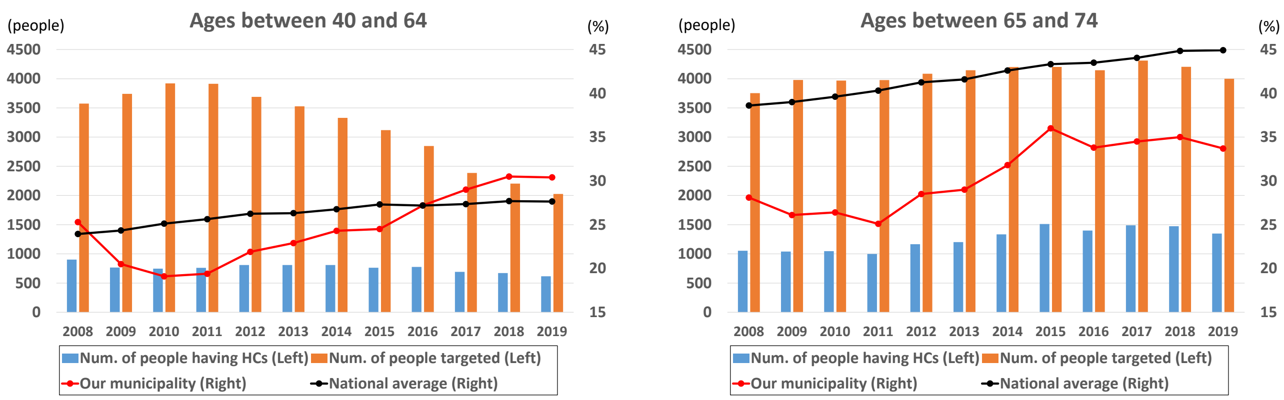
Connection to main results: These descriptive patterns provide important context for interpreting our DID estimates in Section 5.2. The fact that both treatment and control groups experience increases in healthcare outcomes after SHC participation highlights why a DID approach is necessary—simple pre/post comparisons would incorrectly attribute all outcome increases to SHG interventions. The distinct patterns across SHG types (motivational support vs. active support) also foreshadow our main finding that less intensive motivational support appears more effective than more intensive active support, possibly because the populations selected for each type differ systematically in ways that affect their responsiveness to interventions.

Figure B1 Age distribution by gender among the insured in City X



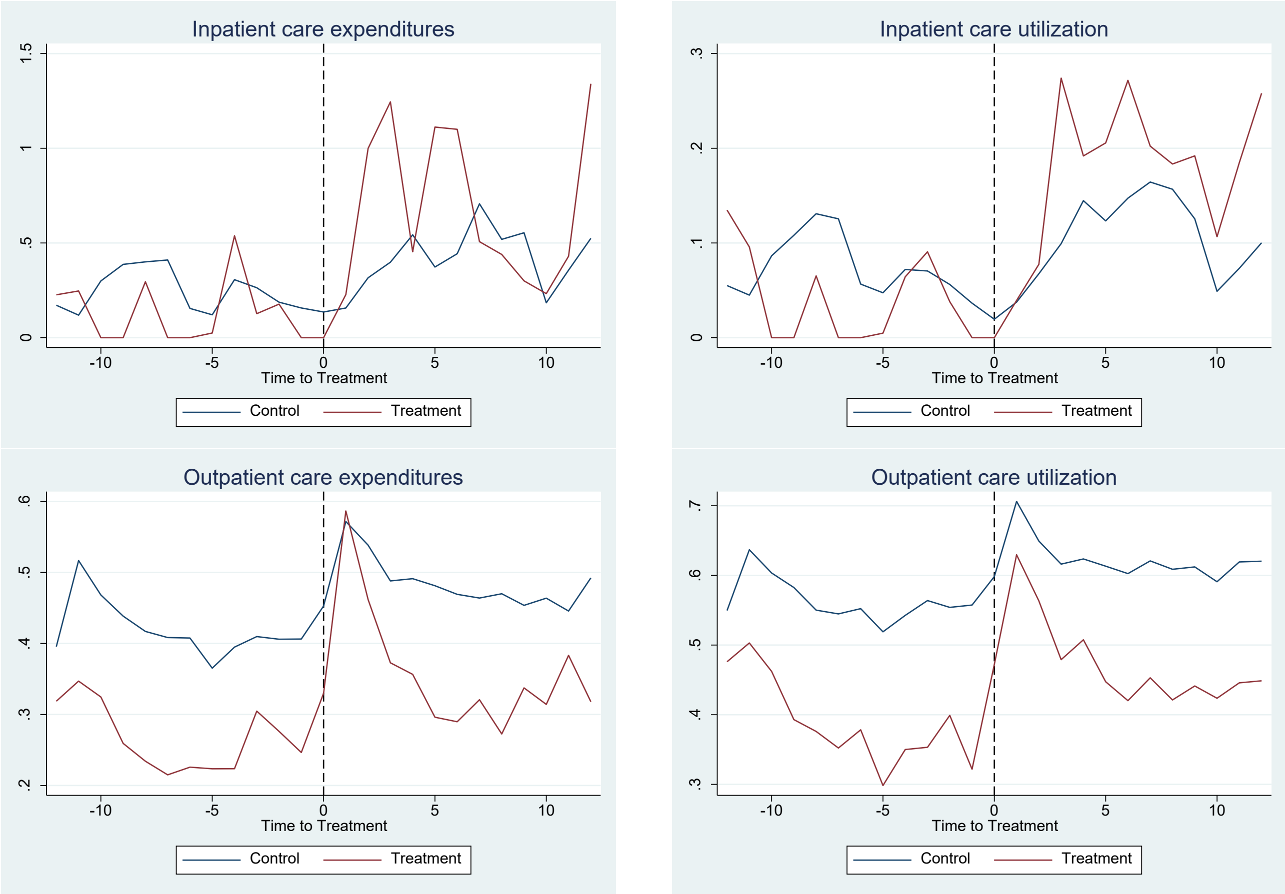
Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
Notes: Figure B1 shows the age distribution of the insured in City X by gender during the study period. The figure reveals a conspicuous demographic pattern skewed toward an aging population within the municipal NHI system. A significant majority, surpassing 50% of the male cohort and 60% of the female cohort, exceeds the age threshold of 60 years. Furthermore, approximately 80% of the insured falls within the age bracket of 40 to 74 years, representing the targeted demographic for the SHC and SHG initiatives described in Section 2.1.

Figure B2 Participation rates in the SHC in City X by age cohort



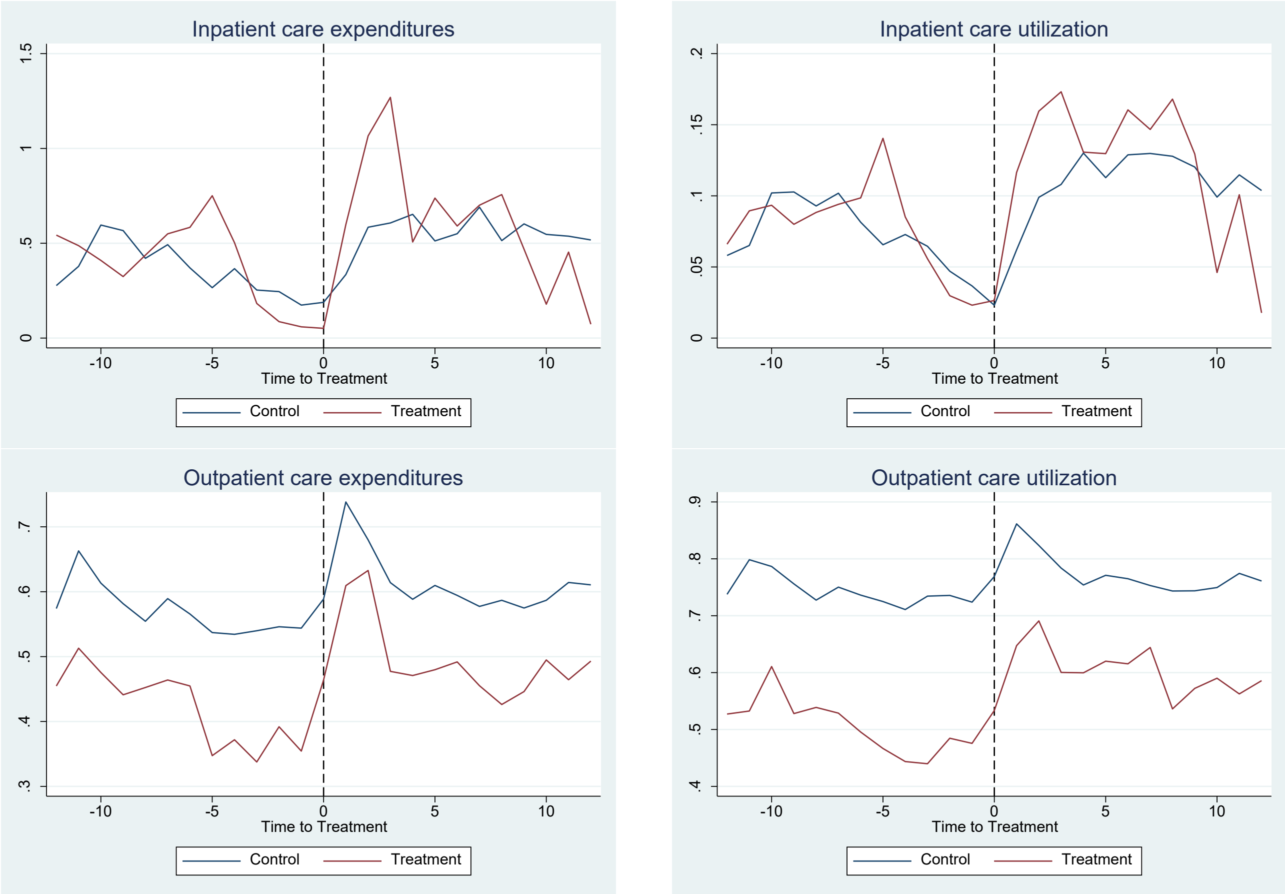
Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2008–2019.
Notes: Figure B2 shows the temporal trends in SHC participation rates in City X by age cohort from FY 2008 to FY 2019. The left panel displays individuals aged 40–64 years, and the right panel shows individuals aged 65–74 years. The left axis represents the number of people, while the right axis represents the participation rate (%). Orange bars indicate the number of people targeted for the SHC, blue bars indicate the number of participants, the red line shows City X's participation rate, and the black line shows the national average participation rate. Despite the introduction of the standardized SHC program in April 2008, participation rates in City X have remained relatively stable over time and consistently lower than the national average for both age cohorts, as discussed in Section 2.2.

Figure B3 Changes in monthly outcome measures by the treatment and control group: Continuous active support



Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
Notes: Figure B3 shows the temporal trends in raw mean healthcare outcomes over event time separately for treatment groups (SHC participants eligible for continuous active support) and control groups (those not eligible for continuous active support): (1) inpatient care expenditures (thousand units per month), (2) inpatient care utilization (hospital days per month), (3) outpatient care expenditures (thousand units per month), and (4) outpatient care utilization (number of visits per month). Continuous active support provides six-month intensive guidance to high-risk individuals under age 65. The horizontal axis represents monthly event time relative to SHC participation (Time to Treatment = 0). The dashed vertical line marks the SHC participation month (event time = 0).

Figure B4 Changes in monthly outcome measures by the treatment and control group: One-off motivational support



Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
Notes: Figure B4 shows the temporal trends in raw mean healthcare outcomes over event time separately for treatment groups (SHC participants eligible for one-off motivational support) and control groups (those not eligible for one-off motivational support): (1) inpatient care expenditures (thousand units per month), (2) inpatient care utilization (hospital days per month), (3) outpatient care expenditures (thousand units per month), and (4) outpatient care utilization (number of visits per month). One-off motivational support provides brief counseling to moderate-risk individuals of all ages (40–74 years), in contrast to continuous active support which targets only high-risk individuals under age 65. The horizontal axis represents monthly event time relative to SHC participation (Time to Treatment = 0). The dashed vertical line marks the SHC participation month (event time = 0).

Table B1 Summary statistics for annual outcome measures by SHC participation and age cohort

Ages 40-74	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Exp. for inpatient care	11,429	5.24 (31.72)	26,555	14.45 (73.44)	***
Exp. for outpatient care	11,429	6.77 (10.39)	26,555	9.36 (34.97)	***
LOS (inpatient)	11,429	1.03 (7.47)	26,555	5.79 (37.76)	***
NOV (outpatient)	11,429	8.67 (13.37)	26,555	8.56 (14.97)	

Ages 40-64	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Exp. for inpatient care	3,878	3.78 (26.22)	11,336	14.21 (75.71)	***
Exp. for outpatient care	3,878	5.06 (11.63)	11,336	8.28 (39.34)	***
LOS (inpatient)	3,878	1.08 (9.30)	11,336	7.32 (45.10)	***
NOV (outpatient)	3,878	6.70 (16.04)	11,336	7.20 (17.21)	*

Ages 65-74	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Exp. for inpatient care	7,551	5.99 (34.18)	15,219	14.63 (71.72)	***
Exp. for outpatient care	7,551	7.64 (9.58)	15,219	10.16 (31.29)	***
LOS (inpatient)	7,551	1.00 (6.33)	15,219	4.65 (31.15)	***
NOV (outpatient)	7,551	9.68 (11.64)	15,219	9.57 (12.97)	

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table B1 presents summary statistics for annual healthcare outcome measures, comparing individuals who participated in the SHC with those who did not participate, stratified by age cohort. The upper panel shows all individuals aged 40–74 years eligible for the SHC, the middle panel shows individuals aged 40–64 years, and the lower panel shows individuals aged 65–74 years. Exp. denotes healthcare expenditures (expressed in one thousand units), LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The t-test column reports the results of two-sample t-tests comparing means between participants and non-participants. Statistical significance: ***$p<0.01$, **$p<0.05$, *$p<0.1$.*

Table B2 Summary statistics for covariates by SHC participation and age cohort

Ages 40-74	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Age	11,429	65.36 (7.78)	26,555	63.41 (8.93)	***
Gender (male=1)	11,429	0.45 (0.50)	26,555	0.45 (0.50)	
Household head	11,429	0.59 (0.49)	26,555	0.62 (0.49)	***
Household members	11,429	2.10 (0.87)	26,555	1.99 (0.88)	***
Eq. income (1K JPY)	11,429	2,837 (2,360)	26,555	2,160 (2,141)	***

Ages 40-64	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Age	3,878	56.68 (6.92)	11,336	55.20 (7.68)	***
Gender (male=1)	3,878	0.46 (0.50)	11,336	0.49 (0.50)	***
Household head	3,878	0.57 (0.49)	11,336	0.65 (0.48)	***
Household members	3,878	2.30 (1.14)	11,336	2.00 (1.04)	***
Eq. income (1K JPY)	3,878	2,431 (2,905)	11,336	1,603 (2,267)	***

Ages 65-74	Taking health checkups		Not taking health checkups		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Age	7,551	69.81 (2.91)	15,219	69.52 (2.79)	***
Gender (male=1)	7,551	0.45 (0.50)	15,219	0.42 (0.49)	***
Household head	7,551	0.59 (0.49)	15,219	0.60 (0.49)	
Household members	7,551	2.00 (0.67)	15,219	1.98 (0.74)	***
Eq. income (1K JPY)	7,551	3,045 (1,993)	15,219	2,574 (1,941)	***

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table B2 presents summary statistics for demographic and socioeconomic covariates used in the analysis, comparing individuals who participated in the SHC with those who did not participate, stratified by age cohort. The upper panel shows all individuals aged 40–74 years eligible for the SHC, the middle panel shows individuals aged 40–64 years, and the lower panel shows individuals aged 65–74 years. Gender is coded as 1 for male and 0 for female. Household head is a binary indicator equal to 1 if the individual is the head of household. Household members indicate the number of individuals in the household. Eq. income denotes equivalized annual household income (expressed in one thousand JPY), calculated as total household income divided by the square root of household size. The t-test column reports the results of two-sample t-tests comparing means between participants and non-participants. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table B3 Summary statistics for monthly outcome measures by SHG interventions

Continuous active support (Ages 40-64)	Eligible for health guidance		Not eligible for health guidance		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Exp. for inpatient care	9,594	0.395 (7.56)	78,542	0.325 (5.44)	
Before the SHC	5,111	0.124 (3.34)	40,441	0.236 (4.51)	**
After the SHC	4,483	0.703 (10.47)	38,101	0.418 (6.29)	***
Exp. for outpatient care	9,594	0.314 (1.08)	78,542	0.453 (1.42)	***
Before the SHC	5,111	0.269 (1.00)	40,441	0.420 (1.29)	***
After the SHC	4,483	0.365 (1.16)	38,101	0.488 (1.54)	***
LOS (inpatient)	9,594	0.103 (1.59)	78,542	0.087 (1.34)	
Before the SHC	5,111	0.036 (0.85)	40,441	0.068 (1.20)	**
After the SHC	4,483	0.179 (2.13)	38,101	0.106 (1.48)	***
NOV (outpatient)	9,594	0.432 (1.39)	78,542	0.594 (1.61)	***
Before the SHC	5,111	0.391 (1.40)	40,441	0.564 (1.58)	***
After the SHC	4,483	0.478 (1.39)	38,101	0.625 (1.65)	***

One-off motivative support (Ages 40-74)	Eligible for health guidance		Not eligible for health guidance		t-test
	N	Mean (Std. Dev.)	N	Mean (Std. Dev.)	
Exp. for inpatient care	26,498	0.498 (7.94)	222,024	0.444 (7.18)	
Before the SHC	13,964	0.373 (6.78)	116,978	0.346 (6.35)	
After the SHC	12,534	0.636 (9.06)	105,046	0.553 (8.00)	
Exp. for outpatient care	26,498	0.458 (1.48)	222,024	0.592 (1.42)	***
Before the SHC	13,964	0.422 (1.52)	116,978	0.570 (1.32)	***
After the SHC	12,534	0.499 (1.43)	105,046	0.618 (1.53)	***
LOS (inpatient)	26,498	0.098 (1.35)	222,024	0.089 (1.28)	
Before the SHC	13,964	0.073 (1.11)	116,978	0.069 (1.12)	
After the SHC	12,534	0.125 (1.58)	105,046	0.111 (1.44)	
NOV (outpatient)	26,498	0.554 (1.29)	222,024	0.759 (1.46)	***
Before the SHC	13,964	0.506 (1.23)	116,978	0.744 (1.45)	***
After the SHC	12,534	0.608 (1.35)	105,046	0.776 (1.48)	***

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table B3 presents summary statistics for monthly healthcare outcome measures among SHC participants, comparing individuals eligible for different types of SHG support with those not eligible for SHG support. The upper panel shows continuous active support (six-month intensive guidance for high-risk individuals aged 40–64 years), and the lower panel shows one-off motivational support (brief counseling for moderate-risk individuals aged 40–74 years). For each intervention type, healthcare outcomes are shown for the overall period, before and after the SHC participation. Exp. denotes healthcare expenditures (expressed in one thousand units), LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. All healthcare outcomes are measured on a monthly basis within the observation window around the SHC participation month, as described in Section 4.2. The t-test column reports the results of two-sample t-tests comparing means between eligible and non-eligible individuals. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

APPENDIX C: Heterogeneous effects across demographic and socioeconomic groups

C.1 Effects of SHC participation

Table C1 OLS and IV estimation results for healthcare expenditures and utilization (age cohort 40-64)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.296* (1.391)		-19.481 (18.217)		0.610*** (0.189)		-2.632 (6.477)	
Health checkup (*t*-1)		-1.516 (1.746)		-17.815 (19.444)		0.511* (0.265)		-5.718 (7.345)
Local participation rate (First-stage)			0.138*** (0.031)	0.137*** (0.032)			0.138*** (0.031)	0.137*** (0.032)
N of obs.	14,673	14,227	14,673	14,227	14,673	14,227	14,673	14,227
Adj. / Centered R2	0.654	0.648	0.738	0.734	0.806	0.780	0.854	0.834
Cragg-Donald Wald F stat.	–	–	89.49	86.46	–	–	89.49	86.46

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.961 (0.690)		-7.172 (7.841)		0.680*** (0.217)		-2.435 (2.729)	
Health checkup (*t*-1)		-0.915 (0.912)		-15.702* (9.244)		0.378 (0.252)		-1.108 (4.180)
Local participation rate (First-stage)			0.138*** (0.031)	0.137*** (0.032)			0.138*** (0.031)	0.137*** (0.032)
N of obs.	14,673	14,227	14,673	14,227	14,673	14,227	14,673	14,227
Adj. / Centered R2	0.794	0.790	0.844	0.836	0.820	0.791	0.864	0.843
Cragg-Donald Wald F stat.	–	–	89.49	86.46	–	–	89.49	86.46

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C1 presents OLS and IV estimation results for annual healthcare expenditures and utilization among individuals aged 40–64 years. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C2 OLS and IV estimation results for healthcare expenditures and utilization (age cohort 65-74)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.871** (1.146)		7.668 (22.592)		-1.166*** (0.373)		2.672 (7.825)	
Health checkup (*t*-1)		4.573*** (1.650)		-43.056 (29.679)		-0.688 (0.534)		18.189* (10.521)
Local participation rate (First-stage)			0.118*** (0.019)	0.115*** (0.019)			0.118*** (0.019)	0.115*** (0.019)
N of obs.	22,470	21,572	22,470	21,572	22,470	21,572	22,470	21,572
Adj. / Centered R2	0.383	0.320	0.530	0.460	0.535	0.529	0.646	0.623
Cragg-Donald Wald F stat.	–	–	87.01	80.35	–	–	87.01	80.35

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.801*** (0.263)		-6.435 (6.141)		-0.291 (0.218)		-0.002 (3.348)	
Health checkup (*t*-1)		0.734** (0.348)		-10.191 (6.759)		-0.197 (0.217)		4.615 (3.940)
Local participation rate (First-stage)			0.118*** (0.019)	0.115*** (0.019)			0.118*** (0.019)	0.115*** (0.019)
N of obs.	22,470	21,572	22,470	21,572	22,470	21,572	22,470	21,572
Adj. / Centered R2	0.735	0.674	0.796	0.745	0.669	0.652	0.749	0.729
Cragg-Donald Wald F stat.	–	–	87.01	80.35	–	–	87.01	80.35

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C2 presents OLS and IV estimation results for annual healthcare expenditures and utilization among individuals aged 65–74 years. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C3 OLS and IV estimation results for healthcare expenditures and utilization (males)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-4.111*** (1.371)		31.455* (18.660)		-0.434 (0.498)		3.578 (7.077)	
Health checkup (*t*-1)		3.014 (2.170)		-6.485 (22.059)		-1.064 (0.751)		10.902 (9.378)
Local participation rate (First-stage)			0.151*** (0.022)	0.149*** (0.022)			0.151*** (0.022)	0.149*** (0.022)
N of obs.	17,128	16,483	17,128	16,483	17,128	16,483	17,128	16,483
Adj. / Centered R2	0.506	0.448	0.603	0.571	0.664	0.664	0.738	0.734
Cragg-Donald Wald F stat.	–	–	117.34	111.49	–	–	117.34	111.49

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-1.311*** (0.447)		7.773 (5.626)		0.227 (0.250)		-1.963 (2.667)	
Health checkup (*t*-1)		-0.137 (0.621)		-0.280 (6.650)		-0.246 (0.251)		1.640 (3.580)
Local participation rate (First-stage)			0.151*** (0.022)	0.149*** (0.022)			0.151*** (0.022)	0.149*** (0.022)
N of obs.	17,128	16,483	17,128	16,483	17,128	16,483	17,128	16,483
Adj. / Centered R2	0.774	0.743	0.820	0.801	0.786	0.768	0.832	0.819
Cragg-Donald Wald F stat.	–	–	117.34	111.49	–	–	117.34	111.49

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C3 presents OLS and IV estimation results for annual healthcare expenditures and utilization among male individuals aged 40–74 years. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C4 OLS and IV estimation results for healthcare expenditures and utilization (females)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.160** (0.915)		-27.759 (21.246)		-0.779** (0.306)		1.684 (6.011)	
Health checkup (*t*-1)		1.440 (1.044)		-32.219 (25.532)		0.322 (0.360)		-1.514 (8.446)
Local participation rate (First-stage)			0.116*** (0.020)	0.115*** (0.020)			0.116*** (0.020)	0.115*** (0.020)
N of obs.	20,856	20,165	20,856	20,165	20,856	20,165	20,856	20,165
Adj. / Centered R2	0.503	0.481	0.601	0.580	0.697	0.618	0.763	0.702
Cragg-Donald Wald F stat.	–	–	80.03	76.20	–	–	80.03	76.20

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.599* (0.307)		-17.215** (8.243)		-0.051 (0.200)		1.257 (3.036)	
Health checkup (*t*-1)		0.185 (0.343)		-18.366** (8.009)		0.294 (0.216)		2.896 (3.896)
Local participation rate (First-stage)			0.116*** (0.020)	0.115*** (0.020)			0.116*** (0.020)	0.115*** (0.020)
N of obs.	20,856	20,165	20,856	20,165	20,856	20,165	20,856	20,165
Adj. / Centered R2	0.764	0.747	0.799	0.783	0.702	0.656	0.767	0.730
Cragg-Donald Wald F stat.	–	–	80.03	76.20	–	–	80.03	76.20

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C4 presents OLS and IV estimation results for annual healthcare expenditures and utilization among female individuals aged 40–74 years. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C5 OLS and IV estimation results for healthcare expenditures and utilization (below 50% of the median income)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-4.212** (2.028)		15.060 (32.871)		0.031 (0.514)		16.220 (17.034)	
Health checkup (*t*-1)		0.978 (3.092)		-20.475 (42.878)		1.396** (0.681)		-13.880 (21.390)
Local participation rate (First-stage)			0.101** (0.040)	0.096** (0.041)			0.101** (0.040)	0.096** (0.041)
N of obs.	10,233	9,880	10,233	9,880	10,233	9,880	10,233	9,880
Adj. / Centered R2	0.711	0.679	0.781	0.758	0.840	0.789	0.872	0.836
Cragg-Donald Wald F stat.	–	–	33.83	29.96	–	–	33.83	29.96

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-1.994* (1.044)		0.621 (14.875)		0.589 (0.381)		-0.199 (6.278)	
Health checkup (*t*-1)		-0.461 (1.653)		-21.354 (17.897)		1.021** (0.445)		1.591 (8.310)
Local participation rate (First-stage)			0.101** (0.040)	0.096** (0.041)			0.101** (0.040)	0.096** (0.041)
N of obs.	10,233	9,880	10,233	9,880	10,233	9,880	10,233	9,880
Adj. / Centered R2	0.850	0.845	0.888	0.878	0.809	0.778	0.857	0.834
Cragg-Donald Wald F stat.	–	–	33.83	29.96	–	–	33.83	29.96

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 201–2016.
*Notes: Table C5 presents OLS and IV estimation results for annual healthcare expenditures and utilization among individuals with equivalized household income below 50% of the median income. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C6 OLS and IV estimation results for healthcare expenditures and utilization (over 50% of the median income)

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.735*** (0.892)		3.721 (14.613)		-0.808*** (0.308)		-1.030 (4.638)	
Health checkup (*t*-1)		2.738** (1.252)		-17.855 (19.321)		-0.637 (0.458)		8.477 (6.421)
Local participation rate (First-stage)			0.144*** (0.019)	0.143*** (0.019)			0.144*** (0.019)	0.143*** (0.019)
N of obs.	26,480	25,537	26,480	25,537	26,480	25,537	26,480	25,537
Adj. / Centered R2	0.312	0.305	0.472	0.462	0.581	0.542	0.679	0.644
Cragg-Donald Wald F stat.	–	–	153.60	148.46	–	–	153.60	148.46

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.632*** (0.193)		-3.888 (4.149)		-0.089 (0.174)		-0.514 (2.127)	
Health checkup (*t*-1)		0.236 (0.279)		-4.521 (4.807)		-0.167 (0.182)		2.730 (2.699)
Local participation rate (First-stage)			0.144*** (0.019)	0.143*** (0.019)			0.144*** (0.019)	0.143*** (0.019)
N of obs.	26,480	25,537	26,480	25,537	26,480	25,537	26,480	25,537
Adj. / Centered R2	0.486	0.447	0.604	0.574	0.695	0.654	0.766	0.732
Cragg-Donald Wald F stat.	–	–	153.60	148.46	–	–	153.60	148.46

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C6 presents OLS and IV estimation results for annual healthcare expenditures and utilization among individuals with equivalized household income over 50% of the median income. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

C.2 Impacts of SHG support among SHC participants

Table C7 DID estimation results for healthcare expenditures and utilization (age cohorts)

	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
One-off motivatioal support	-0.064	-0.115	-0.029	-0.033	0.014	0.010	0.062*	0.068*
(under 65 years of age)	(0.174)	(0.124)	(0.056)	(0.042)	(0.029)	(0.024)	(0.037)	(0.035)
N of obs.	47,385	84,466	47,385	84,466	47,385	84,466	47,385	84,466
Adj. R2	0.013	0.031	0.182	0.182	0.545	0.544	0.660	0.644
One-off motivatioal support	0.222	0.062	0.039	0.017	0.064**	0.043*	0.087***	0.060***
(over 65 years of age)	(0.247)	(0.193)	(0.038)	(0.034)	(0.026)	(0.023)	(0.024)	(0.023)
N of obs.	91,549	164,056	91,549	164,056	91,549	164,056	91,549	164,056
Adj. R2	0.011	0.014	0.090	0.077	0.224	0.245	0.448	0.429

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C7 presents difference-in-differences estimation results for monthly healthcare expenditures and utilization among SHC participants, stratified by age cohort. The upper panel shows results for one-off motivational support among individuals aged 40–64 years (who are not eligible for continuous active support), and the lower panel shows results for one-off motivational support among individuals aged 65–74 years. The coefficients represent the treatment effect of SHG eligibility on monthly healthcare outcomes, measured over 6-month and 12-month horizons after SHC participation. Expenditures are measured in one thousand units per month. LOS denotes length of hospital stay (in days) for inpatient care per month, and NOV denotes number of visits for outpatient care per month. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and control variables described in Section 4.2. Robust standard errors reported in parentheses are clustered at the individual level. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C8 DID estimation results for healthcare expenditures and utilization (males and females)

Males	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
Continuous active support	0.568	0.341	0.084	0.099	0.050	0.036	0.063*	0.024
(under 65 years of age)	(0.401)	(0.387)	(0.071)	(0.105)	(0.031)	(0.022)	(0.037)	(0.029)
N of obs.	21,811	38,939	21,811	38,939	21,811	38,939	21,811	38,939
Adj. R2	0.038	0.050	0.150	0.165	0.655	0.661	0.740	0.729
One-off motivatioal support	-0.124	-0.186	-0.034	-0.035	0.025	0.010	0.060**	0.034
	(0.205)	(0.193)	(0.030)	(0.032)	(0.027)	(0.022)	(0.026)	(0.025)
N of obs.	60,363	107,962	60,363	107,962	60,363	107,962	60,363	107,962
Adj. R2	0.020	0.022	0.131	0.114	0.423	0.440	0.616	0.603

Females	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
Continuous active support	0.237	0.212	0.038	0.026	0.074	0.102**	0.089	0.118**
(under 65 years of age)	(0.413)	(0.248)	(0.070)	(0.040)	(0.068)	(0.050)	(0.064)	(0.053)
N of obs.	27,590	49,197	27,590	49,197	27,590	49,197	27,590	49,197
Adj. R2	-0.003	0.016	0.206	0.220	0.363	0.350	0.511	0.472
One-off motivatioal support	0.658	0.391	0.115	0.078	0.107***	0.094***	0.125***	0.117***
	(0.418)	(0.266)	(0.077)	(0.060)	(0.034)	(0.032)	(0.034)	(0.030)
N of obs.	78,571	140,560	78,571	140,560	78,571	140,560	78,571	140,560
Adj. R2	0.009	0.014	0.124	0.123	0.237	0.246	0.434	0.407

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C8 presents difference-in-differences estimation results for monthly healthcare expenditures and utilization among SHC participants, stratified by gender. The upper panel shows results for males, including both continuous active support (for those aged 40–64 years) and one-off motivational support. The lower panel shows results for females, including both types of SHG interventions. The coefficients represent the treatment effect of SHG eligibility on monthly healthcare outcomes, measured over 6-month and 12-month horizons after SHC participation. Expenditures are measured in one thousand units per month. LOS denotes length of hospital stay (in days) for inpatient care per month, and NOV denotes number of visits for outpatient care per month. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and control variables described in Section 4.2. Robust standard errors reported in parentheses are clustered at the individual level. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

Table C9 DID estimation results for healthcare expenditures and utilization (below and over 50% of the median income)

Below 50% of the median income	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
Continuous active support	0.731	0.990	0.146	0.304	0.069	0.037	0.083	0.053
(under 65 years of age)	(0.815)	(1.008)	(0.162)	(0.286)	(0.053)	(0.039)	(0.057)	(0.043)
N of obs.	15,604	27,958	15,604	27,958	15,604	27,958	15,604	27,958
Adj. R2	0.035	0.058	0.169	0.209	0.696	0.701	0.735	0.725
One-off motivatioal support	0.233	0.091	0.049	0.015	0.039	0.033	0.116**	0.141**
	(0.468)	(0.239)	(0.105)	(0.069)	(0.044)	(0.033)	(0.054)	(0.063)
N of obs.	25,754	46,063	25,754	46,063	25,754	46,063	25,754	46,063
Adj. R2	0.061	0.049	0.194	0.184	0.590	0.540	0.689	0.674

Over 50% of the median income	Inpatient care				Outpatient care			
	Expenditures		Utilization		Expenditures		Utilization	
	6 months	12 months	6 months	12 months	6 months	12 months	6 months	12 months
Continuous active support	0.498	0.137	0.075*	0.019	0.038	0.040	0.045	0.022
(under 65 years of age)	(0.335)	(0.198)	(0.043)	(0.028)	(0.034)	(0.025)	(0.036)	(0.033)
N of obs.	32,958	58,803	32,958	58,803	32,958	58,803	32,958	58,803
Adj. R2	-0.019	0.003	0.145	0.106	0.246	0.244	0.447	0.400
One-off motivatioal support	0.083	-0.051	0.005	-0.008	0.054**	0.036*	0.078***	0.052**
	(0.169)	(0.155)	(0.022)	(0.026)	(0.024)	(0.021)	(0.022)	(0.021)
N of obs.	112,178	200,891	112,178	200,891	112,178	200,891	112,178	200,891
Adj. R2	0.003	0.012	0.082	0.081	0.233	0.250	0.438	0.422

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table C9 presents difference-in-differences estimation results for monthly healthcare expenditures and utilization among SHC participants, stratified by income level. The upper panel shows results for individuals with equivalized household income below 50% of the median income, and the lower panel shows results for those over 50% of the median income. Each panel includes both continuous active support (for those aged 40–64 years) and one-off motivational support. The coefficients represent the treatment effect of SHG eligibility on monthly healthcare outcomes, measured over 6-month and 12-month horizons after SHC participation. Expenditures are measured in one thousand units per month. LOS denotes length of hospital stay (in days) for inpatient care per month, and NOV denotes number of visits for outpatient care per month. All specifications include individual fixed effects, fiscal year fixed effects, calendar month fixed effects, SHC participation event fixed effects, and control variables described in Section 4.2. Robust standard errors reported in parentheses are clustered at the individual level. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*

APPENDIX D: Robustness check for effects of SHC participation

Table D1 OLS and IV estimation results for healthcare expenditures and utilization: Individuals with positive healthcare outcomes

Expenditures (1K units)	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-2.526*** (0.823)		2.261 (14.294)		-0.782*** (0.266)		2.923 (5.132)	
Health checkup (*t*-1)		2.719** (1.137)		-17.618 (18.178)		-0.335 (0.388)		4.637 (6.227)
Local participation rate (First-stage)			0.132*** (0.018)	0.131*** (0.018)			0.132*** (0.018)	0.131*** (0.018)
N of obs.	34,242	32,983	34,242	32,983	34,242	32,983	34,242	32,983
Adj. / Centered R2	0.364	0.328	0.507	0.475	0.644	0.611	0.723	0.698
Cragg-Donald Wald F stat.	–	–	168.31	162.14	–	–	168.31	162.14

Utilization	Inpatient care				Outpatient care			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Health checkup (*t*)	-0.595*** (0.222)		-5.148 (4.289)		-0.075 (0.145)		-0.206 (2.088)	
Health checkup (*t*-1)		0.266 (0.268)		-5.792 (4.848)		-0.033 (0.158)		1.786 (2.721)
Local participation rate (First-stage)			0.132*** (0.018)	0.131*** (0.018)			0.132*** (0.018)	0.131*** (0.018)
N of obs.	34,242	32,983	34,242	32,983	34,242	32,983	34,242	32,983
Adj. / Centered R2	0.613	0.579	0.698	0.671	0.736	0.708	0.796	0.773
Cragg-Donald Wald F stat.	–	–	168.31	162.14	–	–	168.31	162.14

Source: Authors' calculations using administrative data from the NHI program, City X, Japan, FY 2011–2016.
*Notes: Table D1 presents OLS and IV estimation results for annual healthcare expenditures and utilization among individuals with positive healthcare outcomes aged 40–74 years. The IV estimates use the local participation rate (leave-one-out mean SHC participation rate in the individual's residential block) as an instrument for individual SHC participation. Health checkup (t) indicates participation in the SHC during fiscal year t (contemporaneous effects), while Health checkup (t-1) indicates participation during fiscal year t-1 (lagged effects). Expenditures are measured in one thousand units. LOS denotes length of hospital stay (in days) for inpatient care, and NOV denotes number of visits for outpatient care. The first-stage coefficient on the local participation rate and the Cragg-Donald Wald F-statistic are reported to assess instrument strength. All specifications include individual fixed effects, fiscal year fixed effects, and control variables described in Section 4.1. Standard errors reported in parentheses are clustered at both the individual and block levels. Statistical significance: ***p<0.01, **p<0.05, *p<0.1.*