Exploration of healthy worker survivor bias due to employment status change among middle-aged populations in Korea

Byungyoon Yun

Department of Medicine

The Graduate School, Yonsei University

Exploration of healthy worker survivor bias due to employment status change among middle-aged populations in Korea

Directed by Professor Jin-Ha Yoon

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Medical Science

Byungyoon Yun

December 2023

This certifies that the Doctoral Dissertation of Byungyoon Yun is approved.

윤진하

Thesis Supervisor: Jin-Ha Yoon

김범경

Thesis Committee Member#1: Beom Kyung Kim

김희진

Thesis Committee Member#2: Heejin Kimm

Thesis Committee Member#3: Laura S. Rozek

Thesis Committee Member#4: Sung-In Jang

The Graduate School
Yonsei University

December 2023

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Exploration of healthy worker survivor bias due to employment status change among middle-aged populations in Korea

Byungyoon Yun

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jin-Ha Yoon)

Background: Healthy worker bias (HWB) refers to the reduction in mortality and morbidity attributable to various employment-related factors when comparing workers to the general population. Healthy worker survivor bias (HWSB), a component of HWB, refers to the ongoing selection process where individuals who continue to be employed generally exhibit better health compared to those who leave employment. It is more challenging to adjust, compared with hiring bias, because HWSB is a time-dependent selection bias and also a confounding bias. Minimizing HWSB is important in that HWSB typically weakens any negative impact of the exposure when examining the exposure-health response patterns within the occupational studies. However, there is a lack of study examining the extent of HWSB or adjusting it. This study aims to 1) explore the health effects of employment status considering socioeconomic factors and 2) estimate the healthy worker survivor bias due to employment status change (HWSB−ES) among the entire Korean population.

Methods: From National Health Insurance Service (NHIS) database, individuals aged 30−59 who continuously maintained the same type of insurance, whether employed, self-employed, or unemployed from 2008 to 2010, were included. The index date was January 2011 and the last follow-up date was December 2022. The primary outcome was defined as all-cause mortality. The risk of all-cause mortality was estimated by adjusted hazard ratio (HR) and 95% confidence interval (CI) from multivariable Cox proportional hazard models. The extent of HWSB–ES was defined as attenuated proportion of the relative risk from the initial point ("Origin Point") to the end point ("Current Point") of landmark period, which is measured by age-standardized mortality rate ratio. Each age-standardized mortality rate ratio, at the Origin Point and the Current Point, respectively, was computed with age-standardized mortality rate between the employee and the general population. The Origin Point was set as the index year of the cohort (2011) and the Current Point was set as 1–10 years from the Origin Point (2012–2021) based on different landmark period. Short- and long-term HWSB–ES were estimated in both fixed (only considering those leaving the employee population) and dynamic (considering both those leaving the employee population and those newly hired) cohort settings, respectively. Short-term HWSB–ES was estimated based on the 1-year mortality rate from the Current Point, while long-term HWSB–ES was estimated based on the mortality rate until the last follow-up, December 2022.

Result: Among 18,192,989 participants with median 12-year follow up (median age 44; male 49.05%), 64,177 (1.07%), 153,843 (2.30%), and 253,736 (4.61%) individuals died among the age group of 30s, 40s, and 50s, respectively. In male, unemployed group was significantly associated with a higher risk of mortality compared to employee group, with the highest observed mortality risk being in

male in their 40s, exhibiting an adjusted HR of 4.03 (95% CI 3.94–4.10). In female, self-employed group had a higher mortality risk compared to employee group, with the highest observed risk in female in their 30s, showing an adjusted HR of 3.34 (95% CI: 3.19–3.49). According to the estimation of HWSB–ES, 5-year estimate of age standardized HWSB–ES among the entire population in fixed and dynamic cohort setting was 0.325194 and 0.280361 for short-term and 0.155346 and 0.082751 for long-term. The estimated HWSB–ES value indicates the extent of attenuated risk, which may be underestimated due to workers leaving employment. HWSB–ES was higher among female, older age, and longer length of landmark period. Moreover, HWSB–ES was higher in the setting of fixed cohort and short-term. These estimates of HWSB–ES can be applied to previous literature, especially studies that have examined the standardized mortality ratio of specific workers compared to the general population. Further, a function was developed to explore the actual expected relative risk by extrapolating the healthy worker survival bias, calculated by reflecting the gender distribution and age distribution of men and women from previous studies, along with landmark periods, cohort types, and follow-up durations, into the relative risks of previous studies.

Conclusion: This study highlighted the significant association between employment status and all-cause mortality among middle-aged populations in Korea. This study also estimated the extent of HWSB across the entire population of Korea, as well as how to utilize these estimates. When performing occupational studies in Korea regarding hazardous exposures and health outcomes, the extent

of HWSB measured in this study could be applied to avoid underestimation of such risk.

Key words : change in employment status, healthy worker survivor bias, socioeconomic factors, landmark analysis, health check-up participation, all-cause mortality

Exploration of healthy worker survivor bias due to employment status change among middle-aged populations in Korea

Byungyoon Yun

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jin-Ha Yoon)

I. INTRODUCTION

1. Study backgrounds

Healthy worker bias (HWB) refers to the tendency of employed individuals to appear healthier than the general population.[1] This bias is typically assessed by comparing the mortality rates between employed workers and the general population, as indicated in previous studies.[2-4] HWB is primarily categorized into two forms: "healthy worker hiring bias" and "healthy worker survivor bias (HWSB)". The former refers to the selection of healthy individuals during the hiring process, whereas the latter pertains to an ongoing selection process where individuals who remain employed generally exhibit better health than those who leave employment.[5,6] For instance, consider a factory where workers are exposed to certain chemicals. If those who feel unwell due to exposure are more likely to quit, the remaining workforce may appear healthier than the general population. This scenario illustrates HWSB, as the 'surviving' workers are the ones who could withstand the exposure without significant health issues.

As the HWHB is an initial selection process whereby healthy individuals are more likely to seek and gain employment than those who are relatively less healthy, this bias may easily be minimized via comparison with appropriate reference groups, either internal or external, as long as they have a health status similar to that of the exposure group.[7-10] However, HWSB is more challenging to adjust, compared with HWHB, because HWSB is a time-dependent continuing selection bias as well as a confounding bias.[11-13] Previous studies have developed several methods to minimize HWSB by restricting the analysis to workers with more than 15 years of experience[4], lagging exposure[14,15], and adjusting for current employment status as a covariate[16]. Recently, new advanced and complex statistical methods including the marginal structural model[17,18], G-null test[19,20], G-estimation[21,22], and parametric G-formula[23] have also been developed. However, most occupational studies in Korea were cross-sectional or panel studies based on secondary materials such as the Korean Working Conditions Survey[24] or Korean National Health and Nutrition Examination Surveys[25], which do not contain sufficient information regarding occupational information and mortality. Thus, these fancy methods have limited applications.

Minimizing HWSB is important because it typically weakens any negative impact of exposure when examining exposure-health response patterns within occupational studies.[5] Given the difficulties in applying advanced statistics, estimating the extent of HWSB and applying this estimation to the effect sizes in the existing literature can be a valuable approach to properly interpret the results of cross-sectional Korean studies. Unlike traditional occupational cohort or cross-sectional studies, in which death records regarding censored populations are missing, a registry-based National Health Insurance Service (NHIS) database in Korea provides yearly employment status and death records for the entire population, allowing for a more direct assessment of the association between employment and health outcomes.[26,27] This database enables the estimation of HWSB due to employment status change (HWSB-ES) across various demographic and socioeconomic factors. In other words, the HWSB-ES can be directly calculated using a data-driven decision-making approach with real-world data from the entire Korean population, rather than relying on simulations based on sampled data.[28]

In addition to the complexities surrounding the HWSB, it is imperative to acknowledge that the magnitude of this bias may not be uniform across all segments of the workforce. According to a previous review article, several experts described the effect modifiers of HWB including time since initial employment, cause of death, type of occupation or social class, and social conditions at the workplace.[29] In particular, socioeconomic status, as well as age and sex, can play a significant role in shaping the dynamics of employment patterns and subsequent health outcomes, as it is a fundamental and significant component of the social determinants of , and it significantly impacts the well-being of individuals and communities [30-32] Thus, addressing the difference on the size of HWSB or the effect of employment status on health outcomes across socioeconomic strata is important, prior to the estimation of HWSB.

Additionally, participation in health check-ups could be an important factor in Korea, as they may have a potential modifying effect on employment status and health outcomes. Aimed at preventing various diseases and promoting a healthy lifestyle, the Korean government recommends that all employee insurers, self-employed insurers, and their dependents aged ≥ 40 undergo the National Health Check-up Service biennially since 1995.[33] As they are not mandatory, participation in health checkups might be influenced by a combination of socioeconomic conditions and healthy lifestyle factors; therefore, they should be considered when examining the relationship between employment status and health outcomes.[34-37]

2. A conceptual framework for the estimation of healthy worker survivor bias due to employment status change

A conceptual framework can be developed to estimate the HWSB−ES which can be defined as a HWSB that occurs owing to changes in employment status, regardless of the reason. This framework aims to systematically evaluate the impact of employment dynamics on health outcomes and to provide a more accurate assessment of occupational health risks. The core idea of estimating HWSB-ES involves calculating the proportion of employment status changes and comparing the associated risks to those who have continuously maintained their employment status. This comparison was facilitated using landmark analysis—a statistical method in which a specific point in time after cohort entry is selected as a 'landmark.' At this point, individuals are categorized based on their exposure status, and their outcomes are analyzed from the landmark onward. This method excludes individuals who experienced events or had their data censored before reaching a landmark.[38]

The estimation of the HWSB−ES is divided into two categories: 1) short-term and 2) long-term. For each category, the HWSB-ES was assessed through two types of cohort settings: fixed (traditional) and dynamic (registry-based). In the fixed cohort, only those who changed from an employee at the beginning of the landmark period to a non-employee at the end of the landmark period were considered. In contrast, the dynamic cohort considers also considered newly hired individuals who changed from non-employee status at the beginning of the landmark period to employee status at the end of the landmark period, in addition to the fixed cohort (as shown in Figure 1).

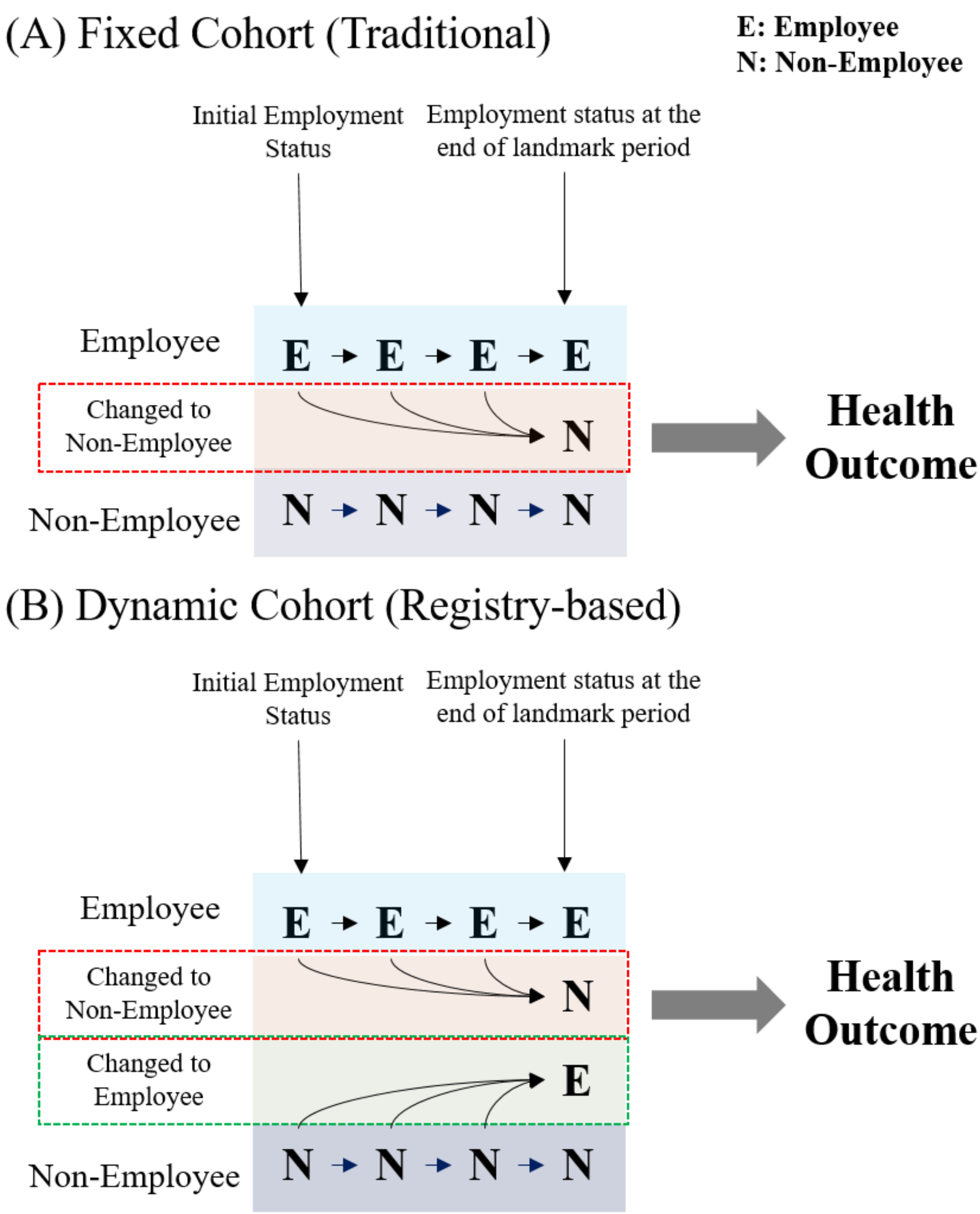


Figure 1. The difference between (A) fixed and (B) dynamic cohorts in estimating healthy worker survivor bias due to employment status change.

In the next step, the term used for HWSB−ES estimation is defined. The Origin Point is defined as the beginning of the landmark period, which refers to the index year of the cohort, whereas the Current Point is defined as the end of the landmark period. The employment status at either the Origin or Current Point was classified into two categories: “employee” or “non-employee.” Non-employees represent all employment statuses, apart from being employees. The general population, which refers to the entire population of the cohort, includes both employee and non-employee groups. Employment status change can be simply defined based on the employment status at the Origin and Current Point. Participants who retained their employee status from Origin Point to Current Point were categorized as “Continuous Employee” group. Those who initially held an employee status but transitioned to non-employee status were labeled as the “Changed to Non-Employee” group, whereas those who were transited from non-employee at Origin Point to employee at Current Point were defined as “Changed to Employee” group. Finally, individuals who maintained a non-employee status both at Origin and Current Point were classified as part of the “Non-Employee” group. The term definitions for estimation of HWSB−ES are summarized in Table 1.

As summarized in Figure 2, to estimate short-term HWSB−ES with a fixed cohort, the mortality rate within one year from the end of the landmark period (Current Point) was calculated among the employee group and the general population at the Origin and Current Points, respectively, considering “Continuous Employee” and “Changed to Non-Employee” groups. For the dynamic cohort, “Changed to Employee” group were also considered when calculating the mortality rate in the employee group at both the Origin and Current Points.

The same methods were applied to estimate long-term HWSB–ES, except for computing the mortality rate per 100,000 PY until the last follow-up.

Table 1. Term definitions for estimating healthy worker survivor bias due to employment status change

Term	Definition
HWSB-ES	healthy worker survivor bias occurring due to employment status change, regardless of any reason
Origin Point	The beginning of the landmark period, referring to the index year of cohort.
Current Point	The end of the landmark period
Landmark Period	Spanning from the Origin Point to the Current point
Employment status	
Continuous Employee	Maintained employee status at both the Origin and Current Points.
Changed to Non-Employee	Transitioned from employee status at the Origin Point to non-employee status at the Current Point.
Changed to Employee	Transitioned from non-employee status at the Origin Point to employee status at the Current Point.
Continuous Non-Employee	Maintained non-employee status at both the Origin and Current Points.

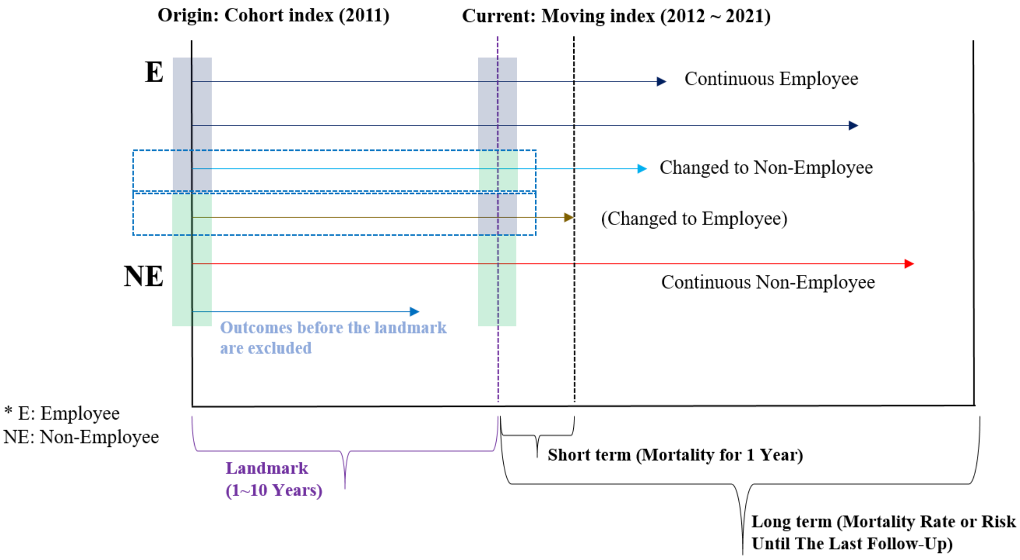


Figure 2. The process of landmark analysis for estimating the extent of healthy worker survivor bias due to employment status change. Individuals who died before the landmark period were excluded. The length of landmark period can be expanded from 1 to 10 years. Short- and long-term healthy worker survivor bias due to employment status change were estimated with comparing 1-year mortality and mortality rate with last follow-up between employee and the general population in both Origin and Current Points.

Table 2 illustrates how HWSB–ES may affect the risk of mortality based on the framework of the HWSB–ES estimation explained above. Suppose that employees at the Current Point presented a 65% lower mortality risk than the general population. However, some individuals may have transitioned from being previously employed at the Origin Point to their non-employee status at the Current Point. After reflecting on the proportion of employment status change and its relative risk of mortality, which may usually have a larger risk than the continuous employee group, the employee group at the Origin Point presented a 44% lower mortality risk than the non-employee group, which was higher than that at the Current Point. The proportion of attenuated risk, or the rate ratio of mortality from the Origin Point to the Current Point, can be defined as the estimated value of the HWSB–ES. This reduction in the mortality risk, which may result from HWSB–ES, can attenuate the extent of the association between exposure and health outcomes.

Similarly, a dynamic cohort (with registry-based NHIS data) can also consider some individuals who were non-employees at the Origin Point but converted into employees at the Current Point (Table 3). This example also considers the proportion and mortality risk of newly hired individuals at the Origin. A reduction in mortality risk is also expected in this case, according to the results of the simulation example.

In conclusion, this framework may elucidate the intricate mechanisms by which changes in employment status obscure the genuine effects of occupational exposures on health outcomes. By differentiating among diverse cohort settings and examining a range of employment transitions, this method facilitates a more detailed exploration of the relationship between employment status and the associated health risks.

Table 2. An illustration of healthy worker survivor bias due to employment status change from employee to non-employee

	Employment status		**Population**	**Rate**	**Death**	**Rate ratio**
Origin Point	General Population	Total	13000	0.071	920	1
	Employee	Total	5000	0.04	200	0.56
		Continuous Employee	4000	0.025	100	
		Changed to Non-Employee	***1000***	***0.1***	***100***	
	Non-Employee	Total	8000	0.09	720	1.27
		Continuous Non-Employee	8000	0.09	720	
Current Point	**Employment status**		**Population**	**Rate**	**Death**	**Rate ratio**
	General Population	Total	13000	0.071	920	1
	Employee	Total	4000	0.025	100	0.35
		Continuous Employee	4000	0.025	100	
	Non-Employee	Total	9000	0.091	820	1.28
		Continuous Non-Employee	8000	0.09	720	
		Changed to Non-Employee	***1000***	***0.1***	***100***	

* The mortality ratio between employee and the general population might be expressed as a rate ratio at Current Point, in most cross-section studies. However, when we trace back to the past (Origin Point), some individuals have been changed from employee to non-employee (changed group). The number of individuals and the mortality rate of changed group will result in workers' healthier outcome. The attenuated proportion of relative risk of mortality between employee and the general population at Origin and Current Point can be an estimate of healthy worker survivor bias due to employment status.

Table 3. An illustration of healthy worker survivor bias due to employment status change from employee to non-employee and vice versa

	Employment status		Population	Rate	Death	Rate ratio
Origin Point	General Population	Total	15000	0.066	1000	1
	Employee	Total	5000	0.04	200	0.61
		Continuous Employee	4000	0.025	100	
		Changed to Non-Employee	***1000***	***0.1***	***100***	
	Non-Employee	Total	10000	0.08	800	1.21
		Continuous Non-Employee	8000	0.09	720	
		Changed to Employee	***2000***	***0.04***	***80***	
	Employment status		**Population**	**Rate**	**Death**	**Rate ratio**
Current Point	General Population	Total	15000	0.066	1000	1
	Employee	Total	6000	0.03	180	0.45
		Continuous Employee	4000	0.025	100	
		Changed to Employee	***2000***	***0.04***	***80***	
	Non-Employee	Total	9000	0.091	820	1.38
		Continuous Non-Employee	8000	0.09	720	
		Changed to Non-Employee	***1000***	***0.1***	***100***	

* In addition to individuals with changed to non-employee, individuals who were changed from non-employee to employee can be considered in registry-based cohort when estimating healthy worker survivor bias due to employment status change.

3. Study objectives

Understanding the relationship between employment status and health outcomes is of paramount importance in occupational health research. Employment status dynamics, influenced by various socioeconomic factors, may play a critical role in shaping an individual's health. This study aims to investigate these complex interactions, offer insights into how employment status, underpinned by socioeconomic variables, affects health, and explains the implications for public health.

With this foundational understanding and using real-world data generated from the conceptual framework, the main objectives of this thesis are as follows: 1) to explore the health effects of employment status by considering socioeconomic factors; and 2) to estimate the HWSB–ES among the entire Korean population using data-driven decision-making approach.

II. MATERIALS AND METHODS

1. Data Source

The National Health Insurance Service (NHIS) Database in Korea contains demographic and SES information, including income level, residential area, employment status, company information, hospital visit records, drug prescription records, and procedure records of all citizens in Korea, with a universal coverage of 97.2% in 2019.[39-41] Moreover, health check-up records with anthropometric measurements, blood tests, radiography, urine tests, and various questionnaires for the history of diseases and lifestyles also exist for the population who participated in national health check-ups, which was 74.8% in 2014.[39] This retrospective cohort study used this database to collect data on study populations.

This study adhered to the ethical principles of the Declaration of Helsinki and was approved by the institutional review board of Yonsei University Hospital (IRB 4-2023-0106). The need for informed consent was waived due to the retrospective nature of this study.

2. Study Population

The target population comprised individuals in their 30s–50s, which is the age of the working population. In Korea, Article 19 of the Prohibition of Age Discrimination in Employment and the Promotion of Older Workers Act stipulates that employers should set employees' retirement age at 60 or older[42]. To avoid potential bias, individuals in their 20s were not included in this study because of the absence of data regarding their educational background and military service status (all males in Korea are obligated to complete military service). Therefore, this study included individuals aged 30–59 years who continuously maintained the same type of insurance–either employed, self-employed, or unemployed–between 2008 and 2010. The employment status for each year was inferred from the annual information on insurance type from the following year. Medical aid recipients were classified as unemployed group. The exclusion criteria were as follows: 1) foreigners, 2) individuals with disabilities, 3) missing values for income level or residential area, and 4) Charlson Comorbidity Index (CCI) ≥ 6 (used only for evaluating the risk of all-cause mortality according to employment status within socioeconomic strata). The CCI score was calculated by adding the scores assigned to each disease. A history of each disease was defined based on the criteria of one or more hospitalization or minimum of three outpatient visits. The ICD-10 codes used to define each CCI disease group are summarized in Appendix 1.[43] The index date was January 1, 2011.

3. Outcomes

The primary outcome measure was all-cause mortality. According to statistics on the causes of death in 2021 from Statistics Korea, the most common causes of death among the middle-aged population include cancer, suicide, cardiovascular disease (CVD), and liver disease. The secondary outcomes included a range of diseases commonly linked to death or occupational hazards. These were identified as the occurrence of all types of cancer, with the exception of basal cell carcinoma and thyroid cancer (excluding codes C44 and C73), cardiovascular diseases (codes I21–I22, I25, I63–I64), various mental health conditions (codes F00–F69), liver diseases (codes K70–K77), injuries from trauma (encompassing all S and T codes), and musculoskeletal disorders, specifically the knee (code M23), cervical disc (code M50), lumbar disc (code M51), and shoulder (code M75). For cancer outcomes, the diagnosis date was defined as the earliest hospital visit date, with the code "V193" referring to the registration program to lower co-payments for rare and intractable diseases, which began in 2006.[44] For other diseases, the date of disease occurrence was the earliest hospitalization date. Participants were followed up until December 2022 or until the occurrence of the outcomes, whichever came first.

4. Variables

The covariates used in the baseline cohort analyses were age, sex, income level, residential area, disease status, and health check-up participation. Income level was inferred from the premium insurance paid to the NHIS among individuals with economic activity (employees and self-employed individuals) and divided into four quartile groups: High, High-Middle, Low-Middle, and Low. Due to different payment system between employees and the self-employed, their income levels were defined separately. The residential areas were divided into four groups: Seoul, Gyeonggi, other metropolitan cities, and other provinces. Disease status was stratified into three groups according to the CCI: 0,1, and 2–5. Whether individuals participated in health check-ups or not during 2009–2010 was also used as a covariate in the 40s and 50s age groups.

Baseline employment status was defined as employee, self-employed, or unemployed according to the type of insurance maintained for the past three years (2008–2010) from the index date. Non-employees were defined as either self-employed or unemployed. For employees, detailed industry classifications using the criteria of the Korea Standard Industry Classification (KSIC) and information on company size were also collected. Employees' industry classification was divided into three groups according to the KSIC's industry sections: public/private school, $1^{st}/2^{nd}$ industry, and 3^{rd} industry. Industry section was used to estimate the HWSB–ES. Employed individuals without industry section information, "Professional Soldiers", "Activities of households as employers; undifferentiated goods-and services-producing activities of households for own use", and "Activities of extraterritorial organizations and bodies" were excluded in the analysis regarding industry section or classification. Company size was divided into four groups according to the number of employees: <5, <50, <300, and ≥300.

5. Statistical Analyses

All data manipulation and analyses were performed with SAS version 9.4 (SAS Institute Inc., Cary, NC, USA and R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). All analyses were performed with stratification of sex and age group (30s, 40s, and 50s).

A. Baseline characteristics and the risk of outcomes according to the employment status

The baseline characteristics of the cohorts are expressed as frequency (%) according to sex and employment status (employee, self-employed, or unemployed). The incidence rates of the outcomes and 95% confidence intervals (CIs) were estimated as the number of cases per 100,000 person-years (PY) with a Poisson distribution. Adjusted hazard ratios (HRs) and 95% CI of the primary and secondary outcomes according to baseline employment status were estimated using univariable and multivariable Cox proportional hazard models. Model 1 was adjusted for age. Model 2 was adjusted for age, residential area, disease status. Participation in health check-ups was additionally adjusted using model 3 for the 40s and 50s age groups. For sensitivity analysis, the risk of all-cause mortality with employment status was estimated at the last follow-up of December 2019, before the COVID-19 pandemic.

B. Difference in the risk of all-cause mortality with employment status according to socioeconomic factors

The risk of all-cause mortality with employment status was estimated across different SES groups in several ways. First, the risk of all-cause mortality among employees or self-employed individuals in each income quartile versus the unemployed group was estimated. Second, the employee group was stratified according to company size or industry classification and compared with the self-employed and unemployed groups. Third, interaction and causal mediation analyses were performed to estimate the interactive and mediating roles of non-participation in health check-ups in the association between employment status and all-cause mortality among individuals in their 40s and 50s. The relative excess risk due to the interaction (RERI) and 95% CIs were calculated to measure the extent of the additive interaction.[45] The causal mediation analysis exploring the mediation effect of health check-up participation on the association between 5-year all-cause mortality and employment status was performed using “medflex” R package, which enables flexible estimation of direct and indirect effects using the nested counterfactual concept.[46] The weight-based approach was applied and standard errors were estimated using the robust sandwich method. Fourth, the risk of mental diseases, injury, or muscle diseases according to employment status was estimated by stratifying the industry classification of employees into public/private school, 1st/2nd industry, and 3rd industry. Fifth, female individuals in their 30s were stratified into binary groups according to whether they had experienced pregnancy between January 2009 and December 2010. The risk of all-cause mortality according to employment status was examined by stratifying pregnancy experience. Pregnancy was defined as at least three hospital visits with ICD-10 codes related to normal pregnancy (O, Z321, and Z34–Z36)[47]

C. Estimation of healthy worker survivor bias due to employment status change across various demographic strata

To assess the extent of the HWSB–ES across different demographic groups within the cohort, a landmark analysis was employed based on the conceptual framework mentioned in the introduction. In this study, the term "Origin Point" was defined as an index year (2011) of cohort, whereas the term "Current Point" was defined as the end year of the landmark period, which could be a moving range from 2012 to 2021. The "landmark period" extends from the Origin Point to the Current Point, with the duration of this period varying annually, ranging from 1 year to 10 years.

Short-term HWSB–ES was estimated with the length of the landmark period from 1 to 10 years, whereas long-term HWSB–ES was estimated with that from one to seven years, to fulfill at least a five-year follow-up period (the outcome from the seven-year landmark, 2011–2018, will be followed up for five years (January 2018 to December 2022)). Individuals who died during the landmark period or those with missing employment status information were excluded. The relative risk (RR) of mortality between employees and the general population at the Origin (RRorigin) and Current (RRcurrent) Points was estimated. The size of the HWSB–ES can be defined as the proportion of attenuated relative risk from the Origin Point to the Current Point and is expressed by the following formula:

$$HWSB - ES = \frac{RRorigin - RRcurrent}{RRorigin}$$

The size of the HWSB−ES was estimated from the ratio of the relative risk of mortality across each age group, sex, and length of the landmark period. The main concept of estimation age-standardized HWSB−ES was estimated by computing the age-standardized mortality rate from both points (RRorigin and RRcurrent were calculated based on the ratio of age-standardized rates between employees and the general population at each time point). Age standardization was performed with 1-unit age using the direct method with the 2010 census population in Korea.

For sensitivity analyses, crude HWSB−ES was first estimated using the crude mortality rate ratio between employees and the general population at the Origin and Current Points. Second, a Poisson regression model for the association of mortality with employees (versus non-employee group) of both Origin and Current Points was used in the short-term HWSB−ES estimation. The age-adjusted odds ratio (OR) of mortality among employees (versus the non-employee group) was estimated at both points. The attenuated percentage of the odds ratio (which can be converted into relative risk owing to the rare occurrence of mortality) from the Origin Point to the Current Point was defined as the age-adjusted HWSB. The same method was applied to estimate long-term HWSB−ES and an age-adjusted Cox proportional hazard model was used to compute the hazard ratio (HR) of mortality among the employee and non-employee groups at both Origin and Current Points. Age-standardized and age-adjusted HWSB−ES were estimated across all demographic strata for each landmark year.

Third, subgroup analyses were performed based on the industry section. In the analysis within each industry section, the HWSB−ES was estimated through a fixed cohort setting, as changing to an employee group might excessively influence the effect size owing to the small number of employees. In the same context, long-term HWSB−ES was estimated because individuals in each industry section have a small number of deaths, and the results with short-term follow-up may fluctuate.

Fourth, employees were categorized as either “remaining in the same industry section” or “Transitioning to different industry section,” as determined by annual data on industry section from 2008 to 2010. The age-standardized short- and long-term HWSB−ES in the fixed cohort setting were calculated for both categories, also due to small sample size of individuals who transitioned to different industry section.

Multiple linear regression models were used to fit the best predictive model for both short- and long-term HWSB-ES estimation, based on the estimated values of HWSB−ES according to age group, sex strata, and length of the landmark period. The goodness of fit of the model was assessed using the coefficient of determination, commonly known as R-squared (R^2).

III. RESULTS

1. Baseline characteristics of middle-aged population cohort

After exclusion, 18,192,989 participants were included in the analysis (Figure 1). The median (interquartile range) age was 44 years (range: 39–49 years), and 49.05% of the patients were men. According to the baseline characteristics, the proportion of employees in Gyeonggi in their 30s was higher (68.15%), and the higher CCI group had a higher proportion of unemployed individuals in their 50s (Table 4–5). Moreover, health checkup participation was prominently high among employees in their 40s and 50s, while those in their 50s showed a higher participation rate among the unemployed.

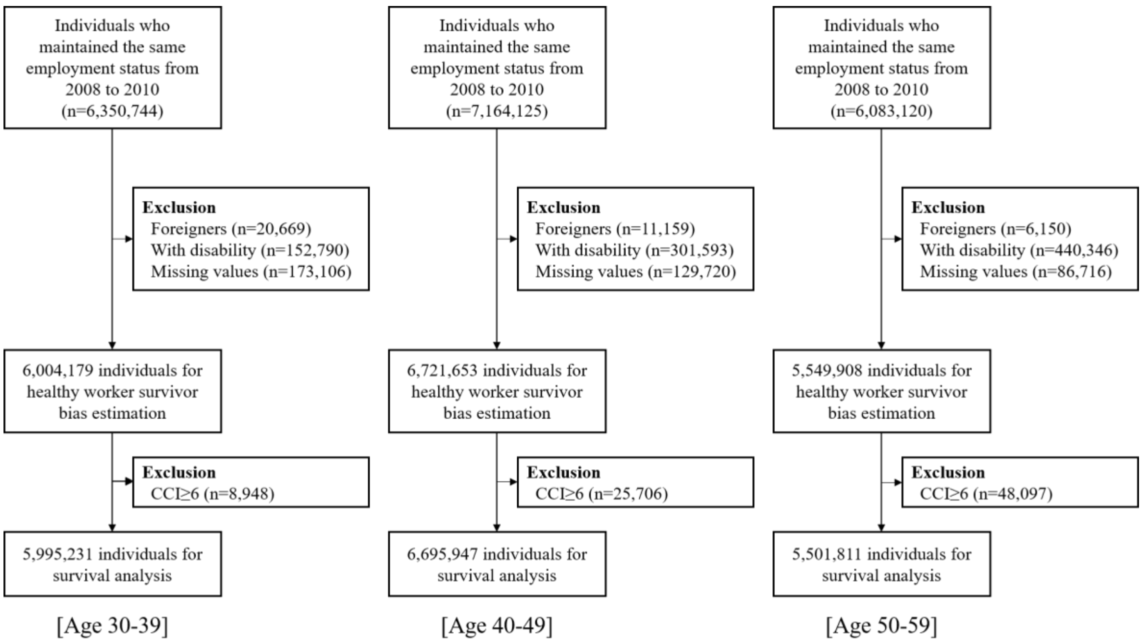


Figure 3. Flow chart of participant selection process across 30s–50s age group

Table 4. Baseline characteristics of male individuals according to age group and employment status

	Variable	Men (n=8,924,179)		
		Employee (n=4,893,792)	Non-Employee (n=4,030,387)	
			Self-Employed (n=2,531,206)	Unemployed (n=1,499,181)
30s	**Age group**			
	30-34	894865 (64.01)	148818 (10.65)	354271 (25.34)
	35-39	1006274 (63.39)	318078 (20.04)	263012 (16.57)
	Residential area			
	Seoul	418722 (61.98)	102478 (15.17)	154405 (22.85)
	Gyeonggi	514421 (68.15)	117117 (15.52)	123283 (16.33)
	Other metropolitan	456766 (61.87)	113433 (15.37)	168032 (22.76)
	Other provinces	511230 (62.60)	133868 (16.39)	171563 (21.01)
	CCI			
	0	1517250 (63.24)	372601 (15.53)	509290 (21.23)
	1	315662 (66.09)	75210 (15.75)	86740 (18.16)
	2~5	68227 (62.84)	19085 (17.58)	21253 (19.58)
40s	**Age group**			
	40-44	1011010 (57.76)	509979 (29.14)	229364 (13.10)
	45-49	823873 (51.72)	578401 (36.31)	190619 (11.97)
	Residential area			
	Seoul	360884 (54.37)	220214 (33.18)	82645 (12.45)
	Gyeonggi	489967 (57.64)	270048 (31.77)	89973 (10.59)
	Other metropolitan	479075 (55.79)	270421 (31.49)	109284 (12.73)
	Other provinces	504957 (52.02)	327697 (33.76)	138081 (14.22)
	CCI			
	0	1239120 (53.87)	766485 (33.32)	294622 (12.81)
	1	429425 (58.15)	225459 (30.53)	83574 (11.32)
	2~5	166338 (54.62)	96436 (31.66)	41787 (13.72)
	Health check-up Participation (Column Percent)			
	No	396793 (21.62)	798785 (73.39)	333181 (79.33)
	Yes	1438090 (78.38)	289595 (26.61)	86802 (20.67)

50s	**Age group**			
	50-54	716174 (46.86)	590219 (38.62)	221938 (14.52)
	55-59	441596 (41.38)	385711 (36.14)	239977 (22.48)
	Residential area			
	Seoul	248558 (46.58)	203292 (38.10)	81754 (15.32)
	Gyeonggi	263939 (45.45)	223653 (38.52)	93069 (16.03)
	Other metropolitan	331105 (47.70)	243802 (35.13)	119189 (17.17)
	Other provinces	314168 (39.91)	305183 (38.77)	167903 (21.33)
	CCI			
	0	619198 (44.33)	544765 (39.00)	232912 (16.67)
	1	324496 (46.44)	251733 (36.02)	122563 (17.54)
	2~5	214076 (42.82)	179432 (35.89)	106440 (21.29)
	Health check-up Participation (Column Percent)			
	No	242824 (20.97)	595541 (61.02)	283975 (61.48)
	Yes	914946 (79.03)	380389 (38.98)	177940 (38.52)

Table 5. Baseline characteristics of female individuals according to age group and employment status

	Variable	Women (n=9,268,810)		
		Employee (n=2,065,649)	Non-Employee (n=7,203,161)	
			Self-Employed	Unemployed
30s	**Age group**			
	30-34	507823 (35.97)	87904 (6.23)	815893 (57.80)
	35-39	395804 (24.76)	144445 (9.04)	1058044 (66.20)
	Residential area			
	Seoul	258748 (37.90)	57290 (8.39)	366594 (53.70)
	Gyeonggi	222244 (29.17)	55640 (7.30)	483927 (63.52)
	Other metropolitan	220782 (28.62)	56460 (7.32)	494091 (64.06)
	Other provinces	201853 (25.42)	62959 (7.93)	529325 (66.65)
	CCI			
	0	709831 (30.83)	171305 (7.44)	1421052 (61.73)
	1	161292 (28.13)	47039 (8.20)	365036 (63.67)
	2~5	32504 (24.19)	14005 (10.42)	87849 (65.38)
40s	**Age group**			
	40-44	397996 (22.59)	216271 (12.28)	1147276 (65.13)
	45-49	336606 (21.15)	227049 (14.27)	1027503 (64.58)
	Residential area			
	Seoul	151101 (21.74)	99821 (14.36)	444195 (63.90)
	Gyeonggi	189116 (22.60)	108963 (13.02)	538596 (64.37)
	Other metropolitan	191794 (21.23)	113023 (12.51)	598734 (66.26)
	Other provinces	202591 (22.08)	121513 (13.25)	593254 (64.67)
	CCI			
	0	511854 (22.53)	298941 (13.16)	1461344 (64.32)
	1	168909 (21.67)	101725 (13.05)	508936 (65.28)
	2~5	53839 (17.89)	42654 (14.17)	204499 (67.94)
	Health check-up Participation (Column Percent)			
	No	129853 (17.68)	287541 (64.86)	1314628 (60.45)
	Yes	604749 (82.32)	155779 (35.14)	860151 (39.55)
50s	**Age group**			
	50-54	274740 (16.57)	228025 (13.76)	1154819 (69.67)
	55-59	152680 (12.23)	153098 (12.26)	942834 (75.51)

Residential area			
Seoul	95858 (14.94)	91497 (14.26)	454193 (70.80)
Gyeonggi	98487 (16.32)	82974 (13.75)	422019 (69.93)
Other metropolitan	113593 (14.38)	97139 (12.29)	579419 (73.33)
Other provinces	119482 (13.72)	109513 (12.57)	642022 (73.71)
CCI			
0	234271 (15.79)	205706 (13.86)	1043948 (70.35)
1	122360 (14.72)	103160 (12.41)	606000 (72.88)
2~5	70789 (11.98)	72257 (12.23)	447705 (75.79)
Health check-up Participation (Column Percent)			
No	60677 (14.20)	204463 (53.65)	909010 (43.33)
Yes	366743 (85.80)	176660 (46.35)	1188643 (56.67)

2. Incidence rates and the risk of the outcomes according to employment status

During median 12-year follow up, 64,177 (1.07%), 153,843 (2.30%), 253,736 (4.61%) individuals died among aged 30s, 40s, 50s group, respectively. Mortality rates per 100,000 person-years are summarized in Table 6. The mortality rate was the highest among male unemployed and female self-employed individuals compared to employees, regardless of age group. The difference in mortality rates between self-employed and unemployed individuals increases with age in men and decreases in women.

According to the multivariable Cox models, an increased risk of all-cause mortality was significantly associated with both self-employed individuals and employees (versus employees), regardless of age or sex (Table 7). The risk of mortality was the highest among unemployed males in their 40s, with an adjusted HR (95% CI) of 4.03 (3.94–4.10), and self-employed females in their 30s, with an adjusted HR (95% CI) of 3.34 (3.19–3.49). The risk of all-cause mortality among self-employed and unemployed individuals prominently decreased with adjustment for health check-up participation in both the 40s and 50s age groups.

Table 6. Mortality rate according to age group, sex, and employment status

Age group	Sex	Employment status	N of death	N of cohort	Rate (95% CI)*
30s	Male	Employee	16158	1901139	71.09 (70.00−72.19)
		Self-Employed	10043	466896	180.83 (177.31−184.40)
		Unemployed	14559	617283	198.73 (195.51−201.98)
	Female	Employee	4082	903627	37.72 (36.57−38.89)
		Self-Employed	3794	232349	137.03 (132.71−141.46)
		Unemployed	15541	1873937	69.36 (68.27−70.46)
40s	Male	Employee	34489	1834883	157.89 (156.23−159.57)
		Self-Employed	43513	1088380	338.55 (335.38−341.75)
		Unemployed	30980	419983	638.61 (631.52−645.76)
	Female	Employee	6178	734602	70.33 (68.58−72.1)
		Self-Employed	8383	443320	158.81 (155.43−162.24)
		Unemployed	30300	2174779	116.82 (115.51−118.14)
50s	Male	Employee	51613	1157770	378.51 (375.25−381.79)
		Self-Employed	67387	975930	591.32 (586.86−595.8)
		Unemployed	62556	461915	1211.94 (1202.46−1221.47)
	Female	Employee	6937	427420	136.12 (132.94−139.37)
		Self-Employed	11195	381123	247.66 (243.09−252.29)
		Unemployed	54048	2097653	217.07 (215.24−218.91)

*Rates are expressed per 100,000 person-year, with 95% confidence interval estimated by Poisson distribution

Table 7. Adjusted HR (95% CI) of all−cause mortality associated with baseline employment status

Age group	Sex	Employment status	Crude model	Model 1	Model 2	Model 3
30s	Male	Employee	1.00(reference)	1.00 (reference)	1.00 (reference)	−
		Self−Employed	2.55 (2.48−2.61)	2.34 (2.28−2.40)	2.34 (2.28−2.40)	−
		Unemployed	2.80 (2.74−2.86)	2.97 (2.91−3.04)	2.98 (2.91−3.05)	−
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	−
		Self−Employed	3.64 (3.48−3.80)	3.48 (3.33−3.64)	3.34 (3.19−3.49)	−
		Unemployed	1.84 (1.78−1.90)	1.79 (1.73−1.85)	1.74 (1.68−1.80)	−
40s	Male	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self−Employed	2.15 (2.12−2.18)	2.05 (2.02−2.08)	2.07 (2.04−2.10)	1.57 (1.55−1.60)
		Unemployed	4.06 (4.00−4.12)	4.06 (4.00−4.12)	4.03 (3.97−4.10)	2.97 (2.92−3.02)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self−Employed	2.26 (2.19−2.34)	2.22 (2.14−2.29)	2.17 (2.10−2.24)	1.67 (1.61−1.73)
		Unemployed	1.66 (1.62−1.71)	1.65 (1.61−1.70)	1.62 (1.57−1.66)	1.27 (1.23−1.31)
50s	Male	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self−Employed	1.57 (1.55−1.58)	1.55 (1.54−1.57)	1.56 (1.54−1.57)	1.25 (1.23−1.26)
		Unemployed	3.22 (3.19−3.26)	3.05 (3.01−3.08)	2.98 (2.95−3.02)	2.36 (2.33−2.39)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self−Employed	1.82 (1.77−1.88)	1.78 (1.73−1.83)	1.76 (1.71−1.81)	1.38 (1.34−1.43)
		Unemployed	1.60 (1.56−1.64)	1.52 (1.48−1.56)	1.48 (1.44−1.52)	1.22 (1.19−1.26)

Model 1 adjusted for age
Model 2 adjusted for age, residential area, and disease status
Model 3 adjusted for age, residential area, disease status, and health check−up participation
Abbreviation: HR, hazard ratio; CI, confidence interval

Regarding secondary outcomes, the self-employed and unemployed groups generally had a significantly higher risk of hospitalization due to secondary outcomes, except for malignancy, than employees in each age group and sex strata (Appendix 3). Notably, the risk of cardiovascular diseases was more pronounced among unemployed males in their 40s and self-employed females in their 30s. The risk of mental and liver diseases also presented a similar pattern, emphasizing that unemployed males in their 40s and self-employed females in their 30s are high-risk populations for secondary outcomes.

In a sensitivity analysis of the last follow-up before the beginning of the COVID-19 pandemic, the similar results regarding the risk of all-cause mortality and morbidities, including all type of malignancies, CVD, mental diseases, and liver diseases were obtained (Appendix 4). The risk of mental diseases was prominently higher at the last follow-up, including COVID-19 pandemic, for unemployed male.

3. Difference in the risk of all-cause mortality with employment status according to socioeconomic factors

The all-cause mortality risk according to employment status was estimated for various socioeconomic strata. When comparing employees and self-employed individuals (i.e., individuals with economic activity) using income quartiles, the risk of all-cause mortality among the self-employed group was significantly higher than the employee group and increased as income levels decreased across all demographic strata (Figure 4). When comparing the unemployed with employees of each income stratum, the risk of all-cause mortality in the unemployed group significantly increased among males, while it was slightly increased in women (Figure 5). The risk of all-cause mortality in the unemployed group was higher than that for high-income employees in their 40s.

In the context of company size, both the self-employed and unemployed groups presented a higher risk of all-cause mortality than the employee group with a larger company size (Figure 6). For industry classification, the risk of all-cause mortality was higher for the reference employees working public/private school companies than for those working in the 1st/2nd or 3rd industries (Figure 7).

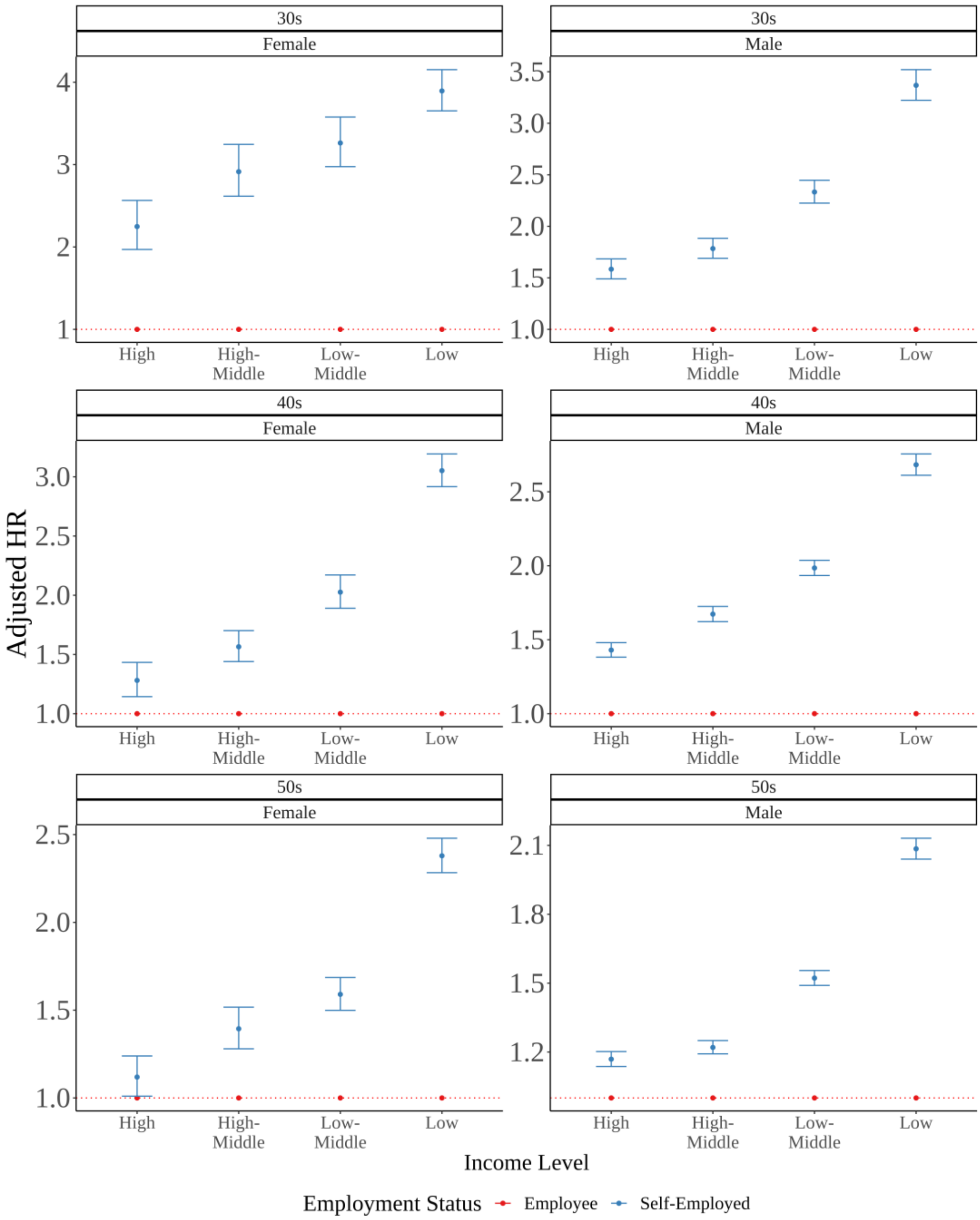


Figure 4. Adjusted HR (95% CI) of all-cause mortality by employment status across income gradient among individuals with economic activity

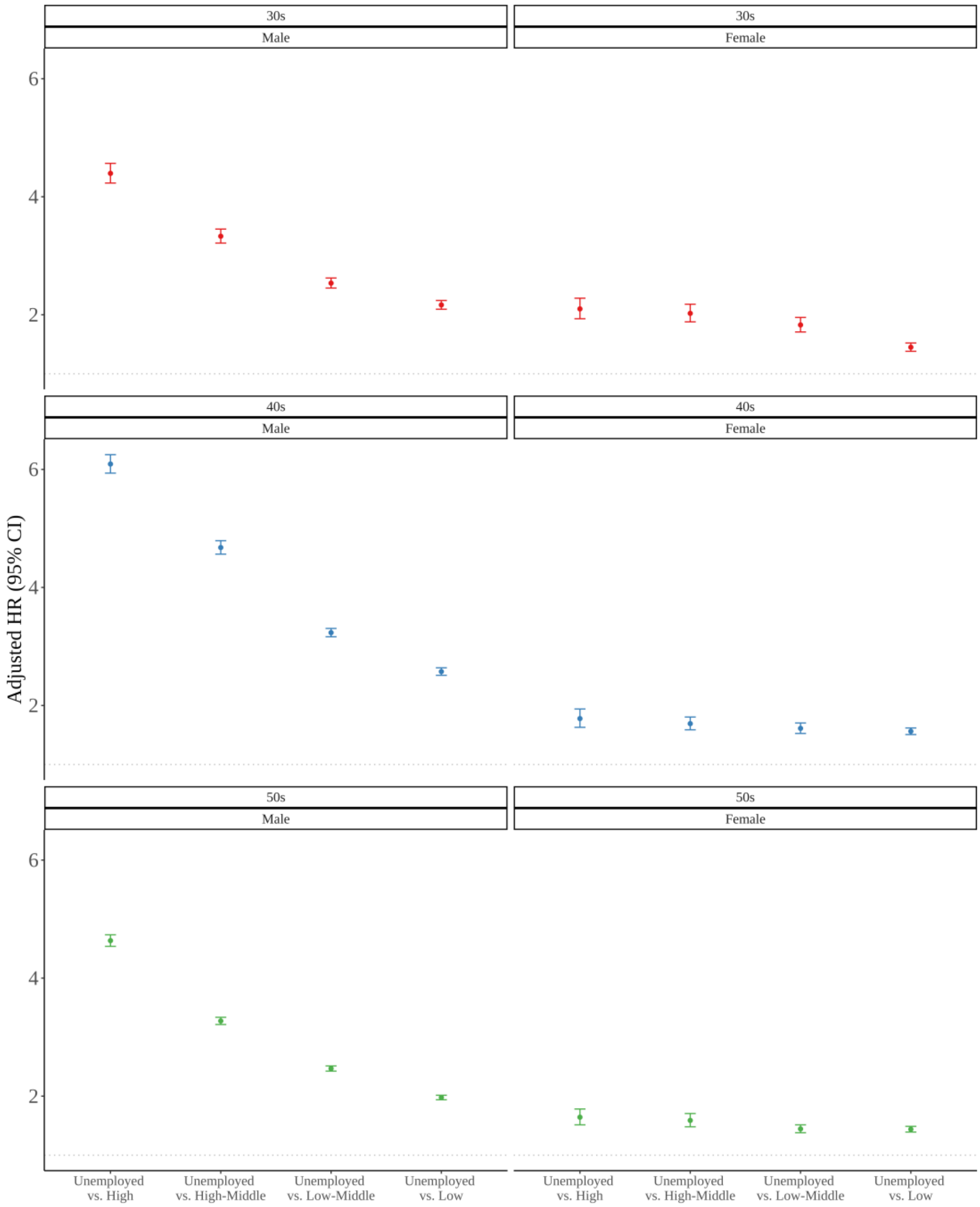


Figure 5. Adjusted HR (95% CI) of all-cause mortality among unemployed, compared with employee of each income quartile

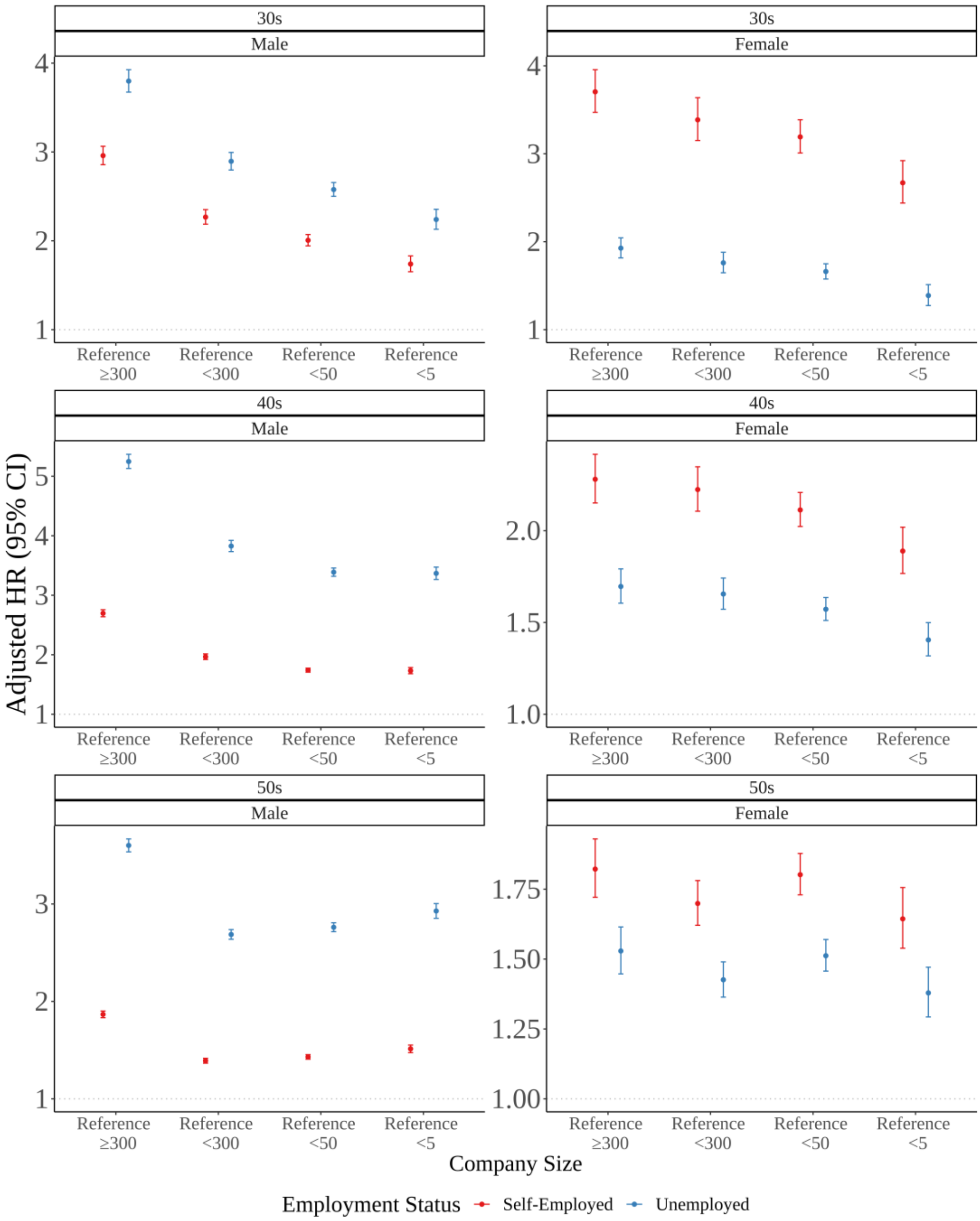


Figure 6. Adjusted HR (95% CI) of all-cause mortality among self-employed and unemployed via comparison with employee of different company size

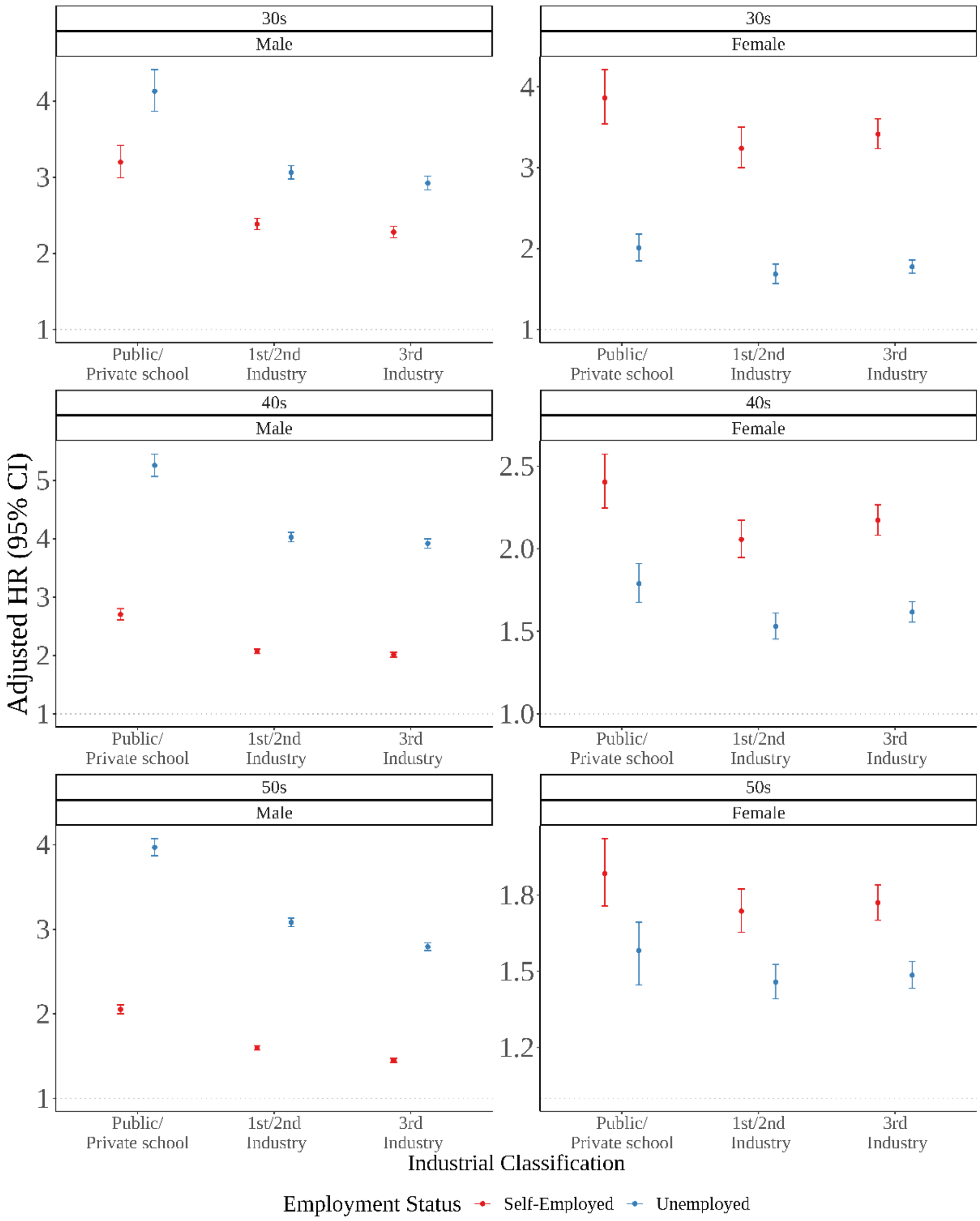


Figure 7. Adjusted HR (95% CI) of all-cause mortality among self-employed and unemployed via comparison with employee in different industry classification

In the analysis of occupational diseases stratified by industry classification among employees (Appendix 5), the risk of mental diseases was significantly higher among the self-employed and unemployed groups than among employees, and the extent of the risk was similar within the specific age and sex groups. Meanwhile, when compared to employees working in the 1st/2nd industries, self-employed and unemployed individuals presented a lower risk of injury or muscle disease than employees in other industries. Specifically, unemployed males in their 30s and 40s showed a lower risk of muscle diseases than employees in any industry.

According to the interaction analysis, non-participation in health checkups and employment status presented significant additive interactions on the risk of all-cause mortality across all age groups and sex strata (Appendix 6–7). The magnitude of the additive interaction was higher in the male population than in the female population (RERI 1.382 [1.334-1.430] for men in their 40s and 1.188 [1.158-1.219] for those in their 50s; 0.659 [0.579-0.739] for women in their 40s and 0.639 [0.559-0.720] for those in their 50s). According to the mediation analysis with a weight-based approach (Appendix 8), health check-up participation has a significant natural indirect effect on the association between employment status and all-cause mortality, with the mediating effect especially prominent among women (proportion of indirect effect: 31.36% for men in their 40s, 39.19% for women in their 40s, 29.77% for men in their 50s, and 43.51% for women in their 50s). Regarding the risk of all-cause mortality with employment status and pregnancy stratification among females in their 30s, the risk of all-cause mortality among unemployed female versus employed female was significantly lower among pregnancy subgroup (Appendix 9).

4. Estimation of healthy worker survivor bias due to employment status change across various demographic strata

   A. Estimation of short-term healthy worker survivor bias due to employment status change

Table 8. Estimation of age-standardized short-term HWSB−ES among the total and sex-stratified populations in both cohorts

Cohort	Length of landmark (Year)	Total	Male	Female
Fixed	1	0.154679	0.129506	0.224945
	2	0.223451	0.195843	0.293336
	3	0.260136	0.219848	0.357205
	4	0.280707	0.243617	0.366526
	5	0.325194	0.296108	0.391802
	6	0.339354	0.305707	0.422749
	7	0.35309	0.32621	0.418381
	8	0.381181	0.35434	0.444221
	9	0.394042	0.363352	0.463866
	10	0.405541	0.386034	0.451208
Dynamic	1	0.127308	0.100229	0.202896
	2	0.196074	0.159586	0.288437
	3	0.23151	0.179942	0.35576
	4	0.259258	0.205905	0.382706
	5	0.280361	0.245134	0.361033
	6	0.305664	0.257212	0.425754
	7	0.313085	0.273413	0.409451
	8	0.332953	0.297912	0.415255
	9	0.350271	0.302597	0.458733
	10	0.375005	0.331693	0.476401

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Table 9. Estimation of age-standardized short-term HWSB–ES from the fixed cohort

Sex	Length of landmark (Year)	30s	40s	50s
Male	1	0.112982	0.115347	0.141972
	2	0.169968	0.137298	0.237262
	3	0.157808	0.178659	0.258485
	4	0.210748	0.207633	0.271841
	5	0.214024	0.237874	0.35076
	6	0.249032	0.268272	0.341162
	7	0.224102	0.271517	0.383107
	8	0.235531	0.317875	0.402625
	9	0.289563	0.327965	0.399922
	10	0.294411	0.349199	0.426013
Female	1	0.146692	0.154943	0.303582
	2	0.186593	0.25877	0.361693
	3	0.27253	0.33198	0.403498
	4	0.316461	0.373534	0.380729
	5	0.292007	0.400404	0.419799
	6	0.321752	0.407083	0.472407
	7	0.325187	0.420222	0.448183
	8	0.383281	0.412712	0.48427
	9	0.367041	0.463069	0.499431
	10	0.35791	0.410301	0.507223

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Table 10. Estimation of age-standardized short-term HWSB−ES from the dynamic cohort

Sex	Length of landmark (Year)	30s	40s	50s
Male	1	0.070126	0.074893	0.122676
	2	0.102757	0.105561	0.206096
	3	0.0906	0.129242	0.230635
	4	0.134166	0.158415	0.250029
	5	0.152691	0.181737	0.305419
	6	0.174197	0.212755	0.303158
	7	0.153798	0.20423	0.343001
	8	0.17078	0.239172	0.360929
	9	0.180339	0.241549	0.36454
	10	0.219971	0.266771	0.391059
Female	1	0.080278	0.161319	0.276518
	2	0.161298	0.263627	0.357051
	3	0.236131	0.35876	0.3966
	4	0.320839	0.364511	0.419248
	5	0.270427	0.338116	0.406174
	6	0.311551	0.410026	0.480433
	7	0.220918	0.425411	0.460404
	8	0.310707	0.396482	0.461632
	9	0.319089	0.450474	0.514911
	10	0.320442	0.428758	0.556962

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Table 8 summarizes the estimated age-standardized short-term HWSB–ES among the total and sex-stratified populations based on the ratio of relative risk of one-year all-cause mortality between the Origin and Current Point with the setting of fixed and dynamic cohorts. Table 9 and 10 summarizes the estimated size of age standardized short-term HWSB–ES according to age group, sex, and the length of landmark in both cohort settings. The size of the short-term HWSB–ES increased as the length of the landmark period increased in both males and females across all age groups in both settings, with the highest increase among those in their 40s. The size of HWSB–ES estimate was lower in the dynamic cohort than in the fixed cohort. The model that best describes the relationship between short-term HWSB–ES, and the length of the landmark period is represented by the equation *y ~ x + exp(-x)*(Figure 8). This suggests that the estimated HWSB–ES experienced a prominent increase during the initial years of the landmark period, followed by a gradual decrease in the rate of increase. According to the model fitting with the linear regression model, the best fitted model for short-term HWSB–ES was as follows, with an R-squared value of 0.937 for the traditional setting and 0.935 for the registry-based setting:

$$\begin{aligned}\text{short} - \text{term}\ \ \text{HWSB} - \text{ES} \\ = \ \beta_0 + \beta_1 Age\ \ Group + \beta_2 Sex + \beta_3 Year + \beta_4 e^{-Year} + \varepsilon\end{aligned}$$

The estimated coefficients for age group (β_1), sex (β_2), and year (β_3), and reverse of exponential year (β_4), were 0.059, 0.103, 0.017, and -0.241, for fixed cohort and 0.08, 0.145, 0.015, and -0.260, for dynamic cohort, respectively. According to the analysis of the crude (Appendix 10–11) and age-adjusted RR (Appendix 12–13) of mortality, the estimated short-term HWSB–ES was similar to the main result in both settings.

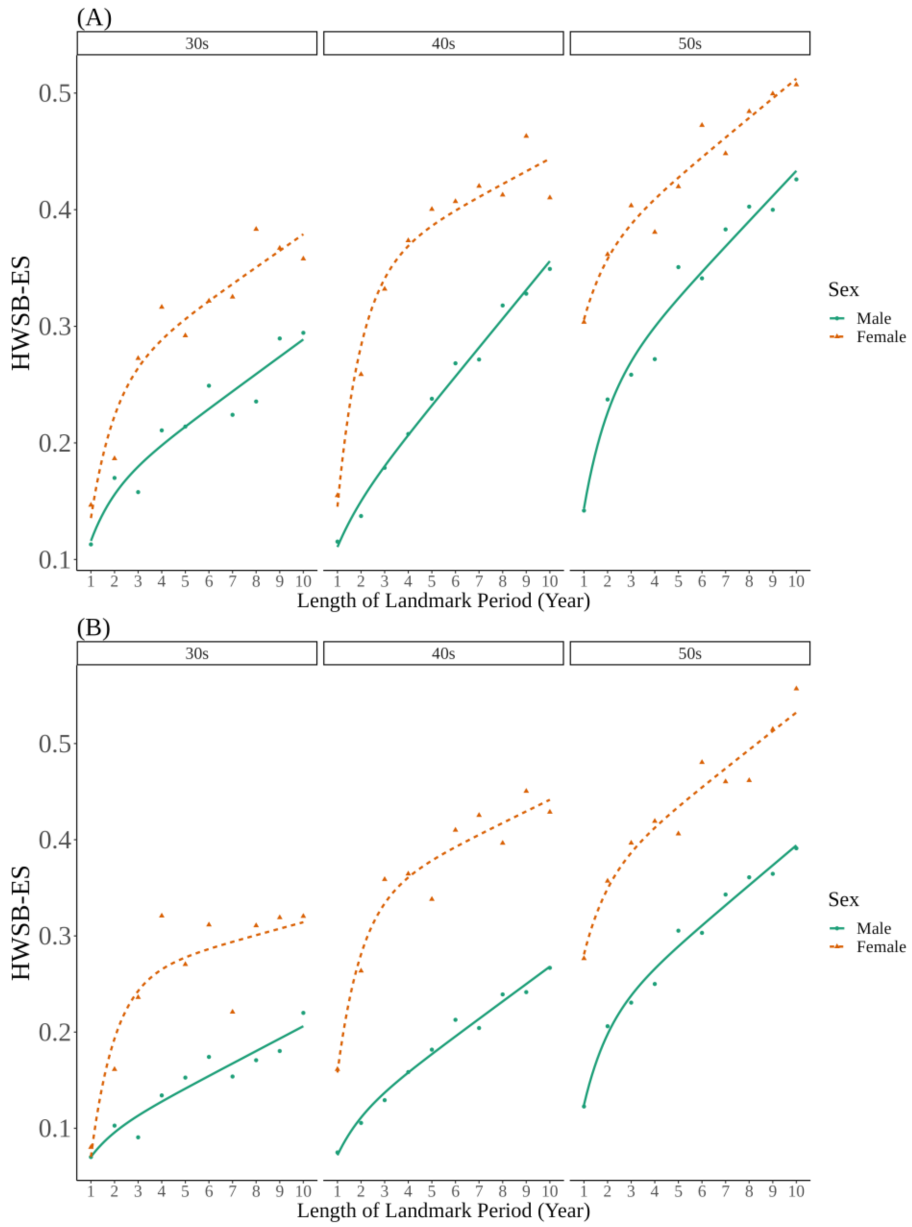


Figure 8. The increasing trend of age standardized short-term HWSB–ES from (A) fixed and (B) dynamic cohorts across age and sex strata

B. Estimation of long-term healthy worker survivor bias due to employment status change

Table 11. Estimation of age-standardized long-term HWSB–ES among the total and sex-stratified populations in both cohorts

Cohort	Length of landmark (Year)	Total	Male	Female
Fixed	1	0.046832	0.046089	0.048616
	2	0.078718	0.075637	0.086047
	3	0.103382	0.098118	0.115848
	4	0.127076	0.120669	0.142232
	5	0.155346	0.146517	0.176278
	6	0.180639	0.171508	0.202384
	7	0.211421	0.198035	0.243104
Dynamic	1	0.006488	0.005295	0.009348
	2	0.025047	0.019873	0.037356
	3	0.043624	0.032755	0.069367
	4	0.062019	0.048492	0.094016
	5	0.082751	0.066886	0.120365
	6	0.111832	0.091696	0.159783
	7	0.139955	0.117792	0.192409

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Table 12. Estimation of age-standardized long-term HWSB−ES from the fixed cohort

Sex	Length of landmark (Year)	30s	40s	50s
Male	1	0.053979	0.048015	0.043139
	2	0.082708	0.072131	0.075948
	3	0.094681	0.093893	0.101301
	4	0.113705	0.115908	0.124973
	5	0.13503	0.136702	0.154697
	6	0.15173	0.161007	0.181928
	7	0.168791	0.180769	0.214285
Female	1	0.047426	0.047419	0.049832
	2	0.085373	0.086386	0.086055
	3	0.1076	0.118015	0.117263
	4	0.135234	0.148839	0.140262
	5	0.142543	0.182621	0.18357
	6	0.160653	0.19991	0.218212
	7	0.193463	0.236309	0.264104

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Table 13. Estimation of age-standardized long-term HWSB−ES from the dynamic cohort

Sex	Length of landmark (Year)	30s	40s	50s
Male	1	-0.00426	0.000024	0.010525
	2	0.00422	0.006502	0.031114
	3	0.001758	0.016862	0.048975
	4	0.019365	0.028441	0.066567
	5	0.028066	0.040183	0.09091
	6	0.048304	0.062543	0.117963
	7	0.06191	0.079097	0.152002
Female	1	-0.00929	0.005964	0.018098
	2	0.019799	0.034765	0.045195
	3	0.04248	0.061903	0.083573
	4	0.05977	0.086546	0.110742
	5	0.060193	0.1126	0.14591
	6	0.079948	0.14672	0.195555
	7	0.098666	0.169572	0.238538

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Table 11 summarizes the estimated age-standardized long-term HWSB–ES among the total and sex-stratified populations based on the ratio of the relative risk of all-cause mortality until the last follow-up between the Origin and Current Points with fixed and dynamic cohorts. Long-term HWSB–ES also increased with the increased length of the landmark period, older age, and female sex and presented lower estimates compared to short-term HWSB–ES. The size of the long-term HWSB–ES in the dynamic cohort setting was also smaller than that in the fixed cohort setting. The relationship between long-term HWSB–ES and length of the landmark period was best described by a linear equation. (Figure 9). According to model fitting with the linear regression model, the best-fitted model for long-term HWSB–ES was as follows, with an R-squared value of 0.947 for the traditional and 0.883 for the registry-based cohort setting:

$$\mathrm{long-term\ \ HWSB-ES} = \ \beta_0 + \beta_1 Age\ \ Group + \beta_2 Sex + \beta_3 Year + \varepsilon$$

The estimated coefficients for age group (β_1), sex (β_2), and year (β_3) were 0.010, 0.021, and 0.026 for fixed cohort and 0.030, 0.043, and 0.021 for dynamic cohort, respectively. The results were similar for the estimation of the crude (Appendix 14–15) and age-adjusted (Appendix 16–17) long-term HWSB–ES.

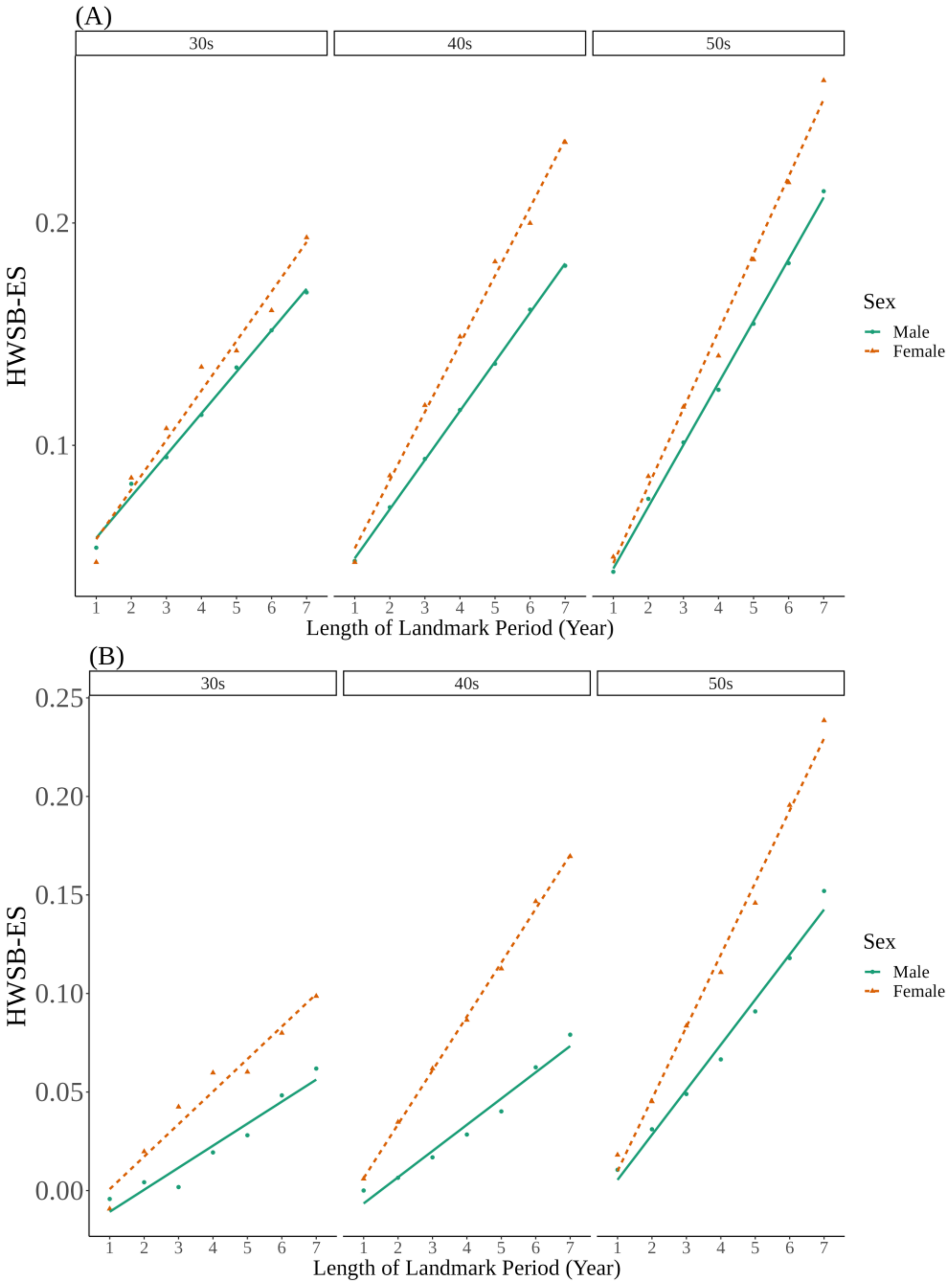


Figure 9. The increasing trend of age standardized long-term HWSB−ES from (A) fixed and (B) dynamic cohorts across age and sex strata

C. Estimation of healthy worker survivor bias due to employment status change of each industry section

Table 14. Estimation of age-standardized long-term HWSB−ES in each industry section from the fixed cohort

Industry section	**Length of landmark (Year)**	**Total**	**Male**	**Female**
Public service	1	0.018138	0.017651	0.019126
	2	0.049382	0.048424	0.051297
	3	0.066252	0.063946	0.070872
	4	0.085697	0.093597	0.069741
	5	0.118441	0.128674	0.097619
	6	0.13852	0.134084	0.147489
	7	0.169994	0.173403	0.163066
Private School	1	0.024213	0.01957	0.032121
	2	0.048898	0.044796	0.055773
	3	0.068778	0.067734	0.070539
	4	0.083185	0.083018	0.083465
	5	0.10387	0.11603	0.08256
	6	0.144451	0.144794	0.14386
	7	0.159949	0.170302	0.142298
Agriculture, forestry, and fishing	1	0.072902	0.058875	0.125082
	2	0.103945	0.069953	0.23022
	3	0.097777	0.081818	0.158095
	4	0.150347	0.159989	0.116432
	5	0.145666	0.147841	0.13786
	6	0.206463	0.195684	0.243958

	7	0.234994	0.194509	0.376475
Mining and quarrying	1	0.05699	0.035099	0.092197
	2	0.097234	0.076112	0.130836
	3	0.100668	0.116158	0.074926
	4	0.138804	0.119722	0.169049
	5	0.25322	0.16435	0.396661
	6	0.322279	0.143594	0.577432
	7	0.303639	0.161217	0.570607
Manufacturing	1	0.041448	0.041412	0.041527
	2	0.078715	0.076734	0.083028
	3	0.109764	0.101052	0.128486
	4	0.137822	0.126113	0.162899
	5	0.172577	0.157812	0.20435
	6	0.200466	0.19187	0.219061
	7	0.233499	0.215838	0.27154
Electricity, gas, steam, and air conditioning supply	1	0.049953	0.031878	0.091316
	2	0.115645	0.059552	0.240313
	3	0.080393	0.05955	0.133682
	4	0.166848	0.086133	0.356227
	5	0.2008	0.105469	0.401947
	6	0.191367	0.117937	0.350107
	7	0.213718	0.116632	0.39383
Water supply; sewage, waste management, materials recovery	1	0.052804	0.045354	0.073064
	2	0.057048	0.059668	0.050166
	3	0.121247	0.123452	0.115434
	4	0.140253	0.154693	0.102401

	5	0.158302	0.200144	0.047938
	6	0.227965	0.264702	0.129542
	7	0.253531	0.288585	0.167162
Construction	1	0.04937	0.047227	0.054604
	2	0.074355	0.076109	0.070052
	3	0.085729	0.085912	0.085277
	4	0.114018	0.105443	0.135081
	5	0.134087	0.119855	0.169435
	6	0.144138	0.136689	0.163447
	7	0.17514	0.157111	0.220444
Wholesale and retail trade	1	0.042753	0.046729	0.033802
	2	0.07605	0.073939	0.08077
	3	0.096526	0.100168	0.088353
	4	0.110672	0.118301	0.093571
	5	0.133707	0.143329	0.112204
	6	0.162024	0.176279	0.130175
	7	0.185999	0.200706	0.153459
Transportation and storage	1	0.053194	0.051359	0.0587
	2	0.087874	0.083141	0.10179
	3	0.110029	0.115751	0.092948
	4	0.134936	0.138049	0.125792
	5	0.159004	0.167393	0.134771
	6	0.199338	0.192789	0.217986
	7	0.230261	0.227787	0.237513
Accommodation and food service activities	1	0.081105	0.074222	0.096356
	2	0.117413	0.115173	0.122228

	3	0.144665	0.137768	0.159591
	4	0.170099	0.161471	0.188953
	5	0.174983	0.164266	0.198698
	6	0.234032	0.21172	0.283396
	7	0.269082	0.228755	0.355587
Information and communication	1	0.043541	0.03122	0.069722
	2	0.054804	0.073692	0.013658
	3	0.060955	0.085286	0.00644
	4	0.095859	0.150846	-0.0254
	5	0.078768	0.150044	-0.0856
	6	0.100458	0.160793	-0.04077
	7	0.139541	0.192793	0.020377
Financial and insurance activities	1	0.010922	0.010458	0.011758
	2	0.030816	0.041974	0.011071
	3	0.062697	0.063334	0.061568
	4	0.116972	0.109666	0.130264
	5	0.13386	0.118395	0.160483
	6	0.194079	0.148155	0.279628
	7	0.190777	0.19011	0.19202
Real estate activities	1	0.045011	0.046318	0.042149
	2	0.069906	0.069101	0.07164
	3	0.094633	0.090904	0.102513
	4	0.122701	0.113936	0.141153
	5	0.156234	0.123946	0.225518
	6	0.167206	0.14104	0.22489
	7	0.224516	0.187453	0.30568

Professional, scientific, and technical activities	1	0.039069	0.039571	0.038081
	2	0.058664	0.060884	0.054426
	3	0.091755	0.083951	0.106573
	4	0.115279	0.090901	0.162346
	5	0.161587	0.106772	0.265408
	6	0.159565	0.10758	0.25543
	7	0.220451	0.137439	0.372549
Business facilities management and business support services; rental and leasing activities	1	0.052136	0.053956	0.047334
	2	0.08568	0.081923	0.095548
	3	0.124847	0.117102	0.144847
	4	0.156806	0.141199	0.196696
	5	0.190462	0.185028	0.204429
	6	0.221645	0.22042	0.224728
	7	0.253587	0.245753	0.272691
Public administration and defense; compulsory social security	1	0.035855	0.040987	0.021135
	2	0.090893	0.081022	0.118812
	3	0.111434	0.088124	0.180521
	4	0.128362	0.115043	0.166935
	5	0.161557	0.133907	0.241505
	6	0.205794	0.168458	0.313324
	7	0.241223	0.207683	0.336717
Education	1	0.028421	0.024378	0.039445
	2	0.067989	0.052109	0.110064
	3	0.100412	0.091304	0.124584
	4	0.112083	0.096983	0.152487
	5	0.139028	0.133488	0.153604

	6	0.171311	0.171654	0.170373
	7	0.214926	0.204629	0.242824
Human health and social work activities	1	0.047252	0.047551	0.046607
	2	0.071367	0.064548	0.086185
	3	0.090628	0.082745	0.107981
	4	0.12317	0.105287	0.162744
	5	0.14154	0.121412	0.187101
	6	0.174958	0.157588	0.215811
	7	0.204065	0.176029	0.270997
Arts, sports and recreation related services	1	0.053598	0.065262	0.024022
	2	0.062336	0.060289	0.067419
	3	0.103714	0.081367	0.158125
	4	0.150706	0.150493	0.151208
	5	0.165199	0.18349	0.121764
	6	0.197499	0.198814	0.194473
	7	0.217514	0.225395	0.198323
Membership organizations, repair, and other personal services	1	0.065113	0.052725	0.093923
	2	0.094187	0.082446	0.12147
	3	0.107663	0.10224	0.120107
	4	0.135606	0.128411	0.152156
	5	0.184596	0.155606	0.250578
	6	0.206582	0.181307	0.264384
	7	0.218549	0.204748	0.249781

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

According to the Table 14, the age-standardized long-term HWSB–ES of each industry section in a fixed cohort setting was estimated for the entire and sex-stratified industrial population. The estimated HWSB–ES differed considerably across industry sections, with female presenting higher estimates of HWSB–ES in most cases. The female group generally presented a higher estimate of the HWSB–ES than the male group, except for the public services and private school industry sections. The number of employees at Origin Point and the proportion of Changed to Non-Employee group in each industry section are summarized in Appendix 18.

D. Estimation of healthy worker survivor bias due to employment status change according to industry section continuity.

Table 15. Estimation of age-standardized short-term HWSB−ES according to industry section change in the fixed cohort

Industry section	Length of landmark (Year)	Total	Male	Female
Remaining in the same industry section	1	0.143075	0.118207	0.211794
	2	0.216184	0.187097	0.289074
	3	0.250306	0.210611	0.345554
	4	0.276233	0.238911	0.362432
	5	0.314654	0.290394	0.369698
	6	0.332566	0.299462	0.413814
	7	0.342225	0.319374	0.397258
	8	0.37181	0.348103	0.427274
	9	0.380973	0.35522	0.438898
	10	0.40242	0.385552	0.442119
Transitioning to different industry section	1	0.213684	0.204222	0.242867
	2	0.268696	0.241011	0.354917
	3	0.313481	0.271067	0.42363
	4	0.289603	0.251091	0.380997
	5	0.396104	0.345751	0.519509
	6	0.378252	0.3348	0.501305
	7	0.426629	0.392218	0.521742
	8	0.426671	0.389572	0.52434
	9	0.481958	0.435463	0.601047
	10	0.427142	0.387319	0.525674

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Table 16. Estimation of age-standardized long-term HWSB−ES according to industry section change in the fixed cohort

Industry section	Length of landmark (Year)	Total	Male	Female
Remaining in the same industry section	1	0.041734	0.041127	0.043179
	2	0.073898	0.071489	0.079578
	3	0.098694	0.094094	0.109493
	4	0.123153	0.117231	0.137031
	5	0.150362	0.142267	0.169368
	6	0.176247	0.16854	0.194424
	7	0.206918	0.19526	0.234246
Transitioning to different industry section	1	0.067264	0.067869	0.06564
	2	0.098606	0.092683	0.114369
	3	0.122819	0.118159	0.135093
	4	0.143145	0.135443	0.163466
	5	0.181622	0.172097	0.207024
	6	0.203105	0.189228	0.240549
	7	0.243372	0.222397	0.299597

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Based on the analysis presented in Tables 15 and 16, both short- and long-term HWSB−ES trends, according to the length of the landmark period, aligned with the primary results, irrespective of variations in the industry section. Moreover, the HWSB-ES estimates were found to be greater among individuals who transitioned to different industry sections than among those who remained in the same industry section (Figure 10). This trend was particularly significant in females with short-term HWSB−ES estimates.

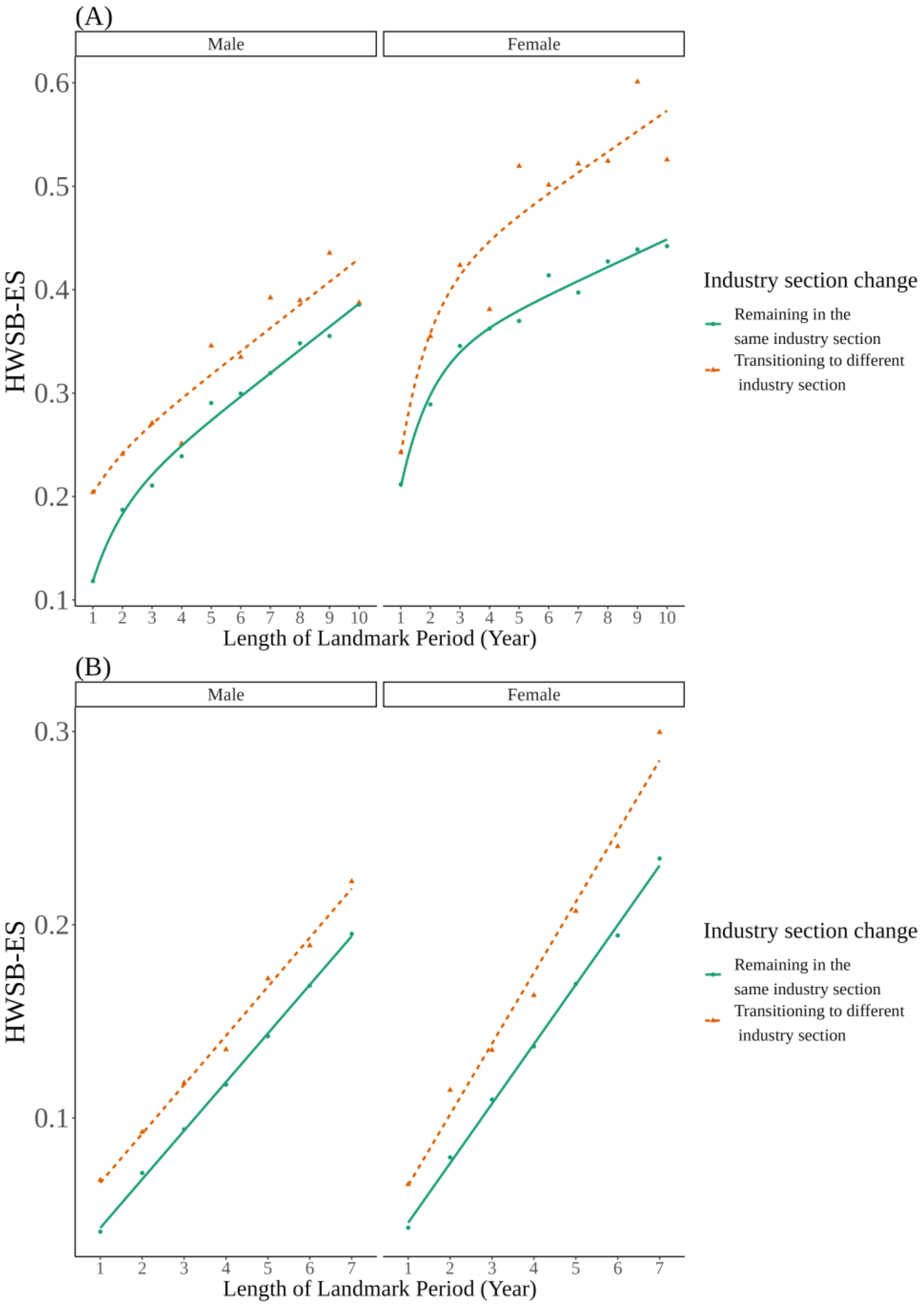


Figure 10. Trend of (A) short- and (B) long-term HWSB−ES estimates in relation to the length of landmark period, associated with industry section change

IV. DISCUSSION

1. Summary of the findings

This study showed that, in middle-aged populations, self-employed and unemployed individuals have a significantly increased risk of all-cause mortality than employees, especially among self-employed women in their 30s and unemployed men in their 40s. The extent of risk varies across SES according to income level, company size, and industry classification. Regarding secondary outcomes, the self-employed and unemployed groups had a significantly increased risk of hospitalization due to CVD, mental disorders, and liver diseases, whereas the risk of malignancy showed a slight or insignificant association. This study also highlighted the interaction/mediating effect of non-participation in health check-ups on the association between employment status and all-cause mortality. The risk of the primary disease and outcomes remained significant in the pre-COVID-19 era. The risk of injury or muscle diseases among the self-employed and employed was the lowest when compared to employees of the $1^{st}/2^{nd}$ industries. Among female in their 30s, the extent of the risk of all-cause mortality among unemployed females (versus employed females) was lower in the pregnancy subgroup.

Furthermore, the sizes of the short- and long-term HWSB−ES were estimated across age groups and sex strata through landmark analysis. The estimated HWSB−ES was similar regardless of the statistical methods, including crude rate ratio, age-standardized rate ratio, and age-adjusted RR of all-cause mortality. HWSB−ES tended to be higher among female, older age, and increased length of the landmark period. The estimated size of long-term

HWSB−ES was lower than that of short-term HWSB-ES. In a fixed cohort setting, the estimated HWSB−ES was higher than in a dynamic cohort setting. Specifically, the HWSB−ES for each industry section was estimated and applied to previous studies on specific industries. Female workers typically exhibited higher HWSB−ES values, except in the public service and private school sections. The estimation of HWSB−ES according to industry section change reproduced similar results, with higher estimates among individuals who transitioned to different industry sections.

2. Interpretation of findings

A. The difference among employment status and its health effect across various socioeconomic strata

The increased risk of all-cause mortality observed among self-employed or unemployed individuals compared with that among employees can be attributed to several factors. A study conducted by Cho et al. indicated a notably higher risk of suicide mortality among the self-employed, particularly women, which may be linked to the precarious nature of self-employment and the association between insecure work environments and increased suicide risks[48-50] These results coincide with the increased risk of hospitalization due to mental disorders among the self-employed group in this study, particularly in women. Another study revealed that self-employed women faced work–life balance challenges and reported high time constraints.[51] The prominently higher risk of suicide mortality among self-employed women may account for the elevated all-cause mortality risk observed in this study.

Furthermore, the study revealed that self-employed individuals faced a notably higher risk of ischemic stroke and heart disease than employed women. This can be attributed to factors such as ambiguous work schedules, income instability, and related work stress experienced by the self-employed[52,53] A similar association was reported in a study by Krittanawong et al., where self-employed status was significantly linked to coronary artery disease, stroke, and heart failure. [54] Moreover, non-participation in periodic health checkups may result in a higher cardiovascular disease risk[55], considering the substantial difference in health check-up participation between the two groups.

A key distinction between mortality risk differences is access to healthcare. The self-employed and unemployed groups exhibited a significantly lower rate of health check-ups than employees, likely because of the absence of responsible employers. This finding is consistent with that of a nationwide survey conducted by Kim et al.[56] Furthermore, self-employed individuals may face financial repercussions when taking time off for health checkups, whereas employees typically receive paid leave for this purpose. Therefore, implementing policies that compensate self-employed individuals for their income loss during health checkups is essential. This notion is reinforced by the subgroup analysis in this study, which showed that the health disparity among the self-employed with health checkup participation was notably lower than that among those without such participation.

Moreover, this study observed a notably higher mortality risk in the self-employed group across all income levels than in the employee group. This disparity can be attributed to factors such as income insecurity, unpredictable work environments, and vulnerability to the economic changes experienced by self-employed individuals.[57] Income insecurity among self-employed individuals can lead to stressful conditions and health problems.[58] Interestingly, the difference in mortality risk between self-employed individuals and employees was more pronounced in the lowest income quartile than in the higher income quartiles, possibly because of a greater income stability gap between the two groups in the low-income categories.

We observed sex-related differences in the association between employment status and mental disorders. Women who were unemployed did not show an increased risk of mortality compared with those who remained employed. These findings, although inconsistent with previous studies on the effects of unemployment on mortality and health outcomes[59-62], can be understood in light of social role theory[63,64]. Women often juggle multiple roles, such as mother, wife, friend, or worker, while men typically prioritize their roles as workers. Many women in Korea voluntarily resign because of societal expectations related to marriage and childcare[65], leading to a high proportion of economic loss among women. This could be a possible explanation for the results of pregnancy stratification among women in their 30s, in which unemployed pregnant women showed a lower risk of mortality than employed pregnant women. Consequently, the impact of the loss of economic activity on mortality may be less significant for women who are less affected by unemployment. In contrast, men who experienced loss of economic activity showed an increased risk of mortality, possibly because of the mental burden of being a breadwinner[66], particularly for mental disorders related to substance abuse. This association was also closely related to the loss of income resulting from unemployment. Men in the high-income subgroup showed the highest risk of all-cause mortality, which is likely attributable to substantial income disparity. This could also be a possible explanation for men in their 40s having the highest risk of all-cause mortality, as they are the most economically active.

B. Estimation of healthy worker survivor bias due to employment status change

This study emphasizes that the estimated HWSB−ES represents a diminished portion of mortality risk, which may differ based on age, sex, and industry section. Furthermore, the total HWSB−ES value estimation revealed that previous studies have underestimated the risk by approximately 30% in the long term and 15% in the short-term mortality. The estimated size of the HWSB−ES was higher in female, older age, fixed cohort setting, and longer landmark periods. The size of the HWSB−ES in a fixed cohort setting can be determined by the proportion of employees who changed to non-employees at the Origin Point and the mortality risk ratio between the continuously employed group and the non-employee group. In the dynamic cohort setting, newly employed group (which refers to "Changed to Employee group") may also impact on the size of HWSB−ES.

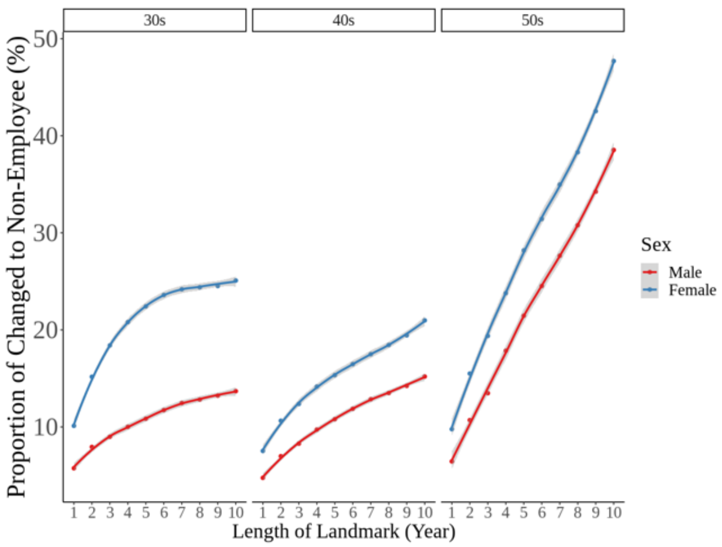


Figure 11. The proportion of changed to non-employee group among employee at the origin point across age and sex strata

According to Figure 11, the proportion of Changed to Non-Employee group among employee at Origin Point was higher among female and prominently higher among individuals in their 50s. Although females in their 30s presented a higher proportion of Changed to Non-Employees than those in their 40s, maternal factors, including marriage, pregnancy, and childcare, can explain the lower estimate of HWSB–ES among those in their 30s. This figure may also explain the positive association between the length of the landmark period and the HWSB–ES estimate, as the proportion of the changed to the non-employee group increased with the length of the landmark period.

Furthermore, the HWSB–ES estimates were higher in the fixed cohort setting. The difference in the HWSB–ES estimates between the two cohort settings stems from the proportion and risk of mortality among the changed employee group. Since the risk of mortality among the changed-to-employee group versus the continuous-employee group is higher, the relative risk of mortality at the Current Point in the dynamic cohort may be higher than that in the fixed cohort. This may finally result in decrease the size of the HWSB–ES estimates.

According to the HWSB–ES estimates across various industry sections, women had higher estimates than men, with the exception of the public service and private school sections. This trend can be attributed to the comparable and lower rates of transition from employee to nonemployee status in both the public and private school sections observed in both male and female groups, as shown in Appendix 18. Given the significant variation in the proportion of individuals leaving their companies in different industry sections, the application of HWSB–ES may vary depending on the specific industry. Our study also highlighted that individuals who transitioned to other industry sections at baseline period presented higher HWSB-ES estimates than those who remained in the same industry. This can be attributed to the unique risks and adaptation challenges associated with entering a new environment. Transitioning to a different industry often involves adapting to new work cultures, job responsibilities, and health and safety risks, which can significantly impact an individual's health and well-being.[67] Hence, to gain deeper insight into the extent of HWSB–ES and its practical implications, it is crucial to carefully consider an individual's employment trajectory, encompassing their industry section and potential transitions to different industry sections.

The HWSB-ES trend differed between short- and long-term HWSB–ES. Short-term HWSB–ES presented a prominent increase in the first few years and gradually decreased as time progressed, whereas the immediate effect of long-term HWSB–ES in the first few years tended to wane but significantly increased thereafter. Short-term HWSB–ES refers to the relatively immediate consequences of changes in employment status on health outcomes within a short timeframe. These effects include stress-related responses, lifestyle changes, and conditions that manifest shortly after a change. However, long-term HWSB–ES might reflect the sustained and cumulative effects of HWSB, which develops or worsens over time owing to prolonged exposure to certain employment conditions or transitions. Diseases with longer latency periods may affect the long-term HWSB–ES. These trends provide valuable insights into how changes in employment status affect health outcomes in both the short- and long-term, offering a comprehensive view of the dynamic relationship between employment transition and well-being.

3. Application of healthy worker survivor bias due to employment status change to the existing literature

As this study emphasized the negative influence of HWSB−ES on the risk of exposure and health outcomes in occupational studies, the extent of short-term and long-term HWSB−ES was estimated in various scenarios (age group, sex, and length of landmark period), using real-world evidence. Since the calculated HWSB−ES was based on the entire Korean population, previous literature regarding the risk of mortality with occupational exposure can be interpreted using the method of extrapolation of HWSB−ES. As an illustration, if the relative risk of exposure A on all-cause mortality is 1.2 among men in their 40s, the estimated five-year HWSB−ES will be 0.237874 for the short term and 0.136702 for the long term in a fixed cohort setting. This implies that a risk attenuation of 23.79% (short) or 13.67 (long) from RRorigin leads to a value of 1.2. The exposure risk of Origin Point, which may be five years past from the Current Point can be estimated with 1.2 multiplied by 100/ (100-23.79) for short-term and 100/ (100-13.67) for long-term, and the relative risk considering short-term or long-term HWSB−ES of 5 years will be 1.57 or 1.39, respectively.

This method can be applied to real-world literature. Lee et al. conducted a study on mortality among agricultural workers compared with the general Korean population.[68] They used death registration data from 2004 to 2008 and aggregated the number of deaths each year and occupation information for that year. The results showed a higher mortality rate among agricultural workers, with a standardized mortality ratio (SMR) of 1.08 (1.07−1.09). However, this result was based on aggregated mortality rates for each year. If we track the mortality rates among individuals who engaged in the agricultural industry in

2004, HWSB–ES may exist owing to employment status changes between agricultural workers in 2004 and workers in 2005–2008 (Figure 12). In other words, the SMR of agricultural workers in 2008 (Current Point) should be adjusted for the four-year length of landmark short-term HWSB if they were followed up from 2004 (Origin Point). If we apply two-year length of landmark short-term HWSB as an average of the overall length of the landmark period between the Origin (2004) and Current (2005–2008) Points, the estimated HWSB–ES will be 0.223451 for the fixed cohort setting and 0.196074 for the dynamic cohort setting. After adjusting the possible attenuation due to HWSB–ES, SMR of agricultural workers may be approximately 1.39 (1.38–1.40) or 1.34 (1.33–1.35), for fixed cohort or dynamic setting, respectively.

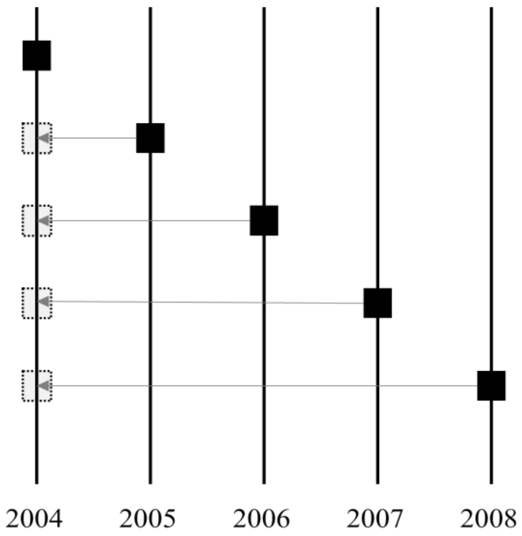


Figure 12. An illustration of application of short-term HWSB–ES in the study of agricultural workers

Moreover, if the estimated HWSB–ES of each industry section is used, two-year HWSB–ES estimate of the agriculture, forestry, and fishing" industry section is 0.103945. Thus, the adjusted SMR of agricultural workers with extrapolation of the HWSB–ES may be 1.21 (1.20–1.22).

In another study, the malignant and non-malignant SMR of male foundry workers were calculated using data from 1992 to 2008 by Yoon et al.[69] which included workers employed in foundry companies for at least one day between 1992 and 2000, considering them as part of the cement industry workforce. These workers were monitored until 2007 or 2008, as shown in Figure 13. The SMR for nonmalignant diseases among production workers was 1.06 (0.97–1.15). Given the long-term follow-up (at least from 2000 to 2007–2008) and similar methodological approaches, a four-year landmark period for long-term HWSB–ES is applicable. When adjusting for potential attenuation owing to HWSB–ES, the estimated SMRs for foundry workers might be approximately 1.21 (1.12–1.30) and 1.13 (1.04–1.22) in fixed and dynamic cohort settings, respectively. If the estimated HWSB–ES of manufacturing industry section is applied, the extrapolated SMR is 1.21 (1.12–1.30).

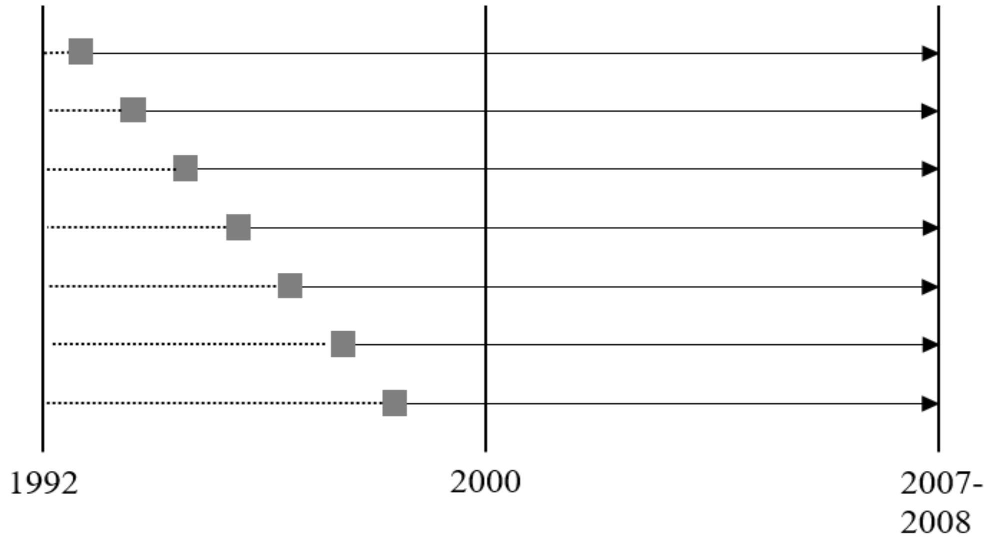


Figure 13. An illustration of application of long-term HWSB−ES in the study of foundry male workers

As seen above, various interpretations can be made from the previous literature using several types of HWSB-ES estimates. The HWSB−ES can be described as the percentage of underestimation, as workers who remain continuously employed tend to show better health outcomes than those who leave employment. Considering the effect of HWSB−ES on the attenuation of the association between risk exposure and mortality might assist in exploring the health effects of hazardous substances concealed by HWSB. The results of estimated HWSB−ES can be briefly summarized with the following heat map for user convenience (Figure 14).

Nonetheless, because of the potential for significant differences in the values according to various scenarios, the results should be cautiously interpreted. For instance, for agricultural workers, it may be deemed most suitable to consider using industry-specific HWSB−ES, given that the traits of agricultural workers can exert a significant impact on this bias. However, it is also crucial to acknowledge that, while the previous study computed SMR values in a short-term manner, the extent of industry-specific HWSB−ES was estimated over a long-term period, which could lead to an underestimation of attenuation owing to HWSB−ES.

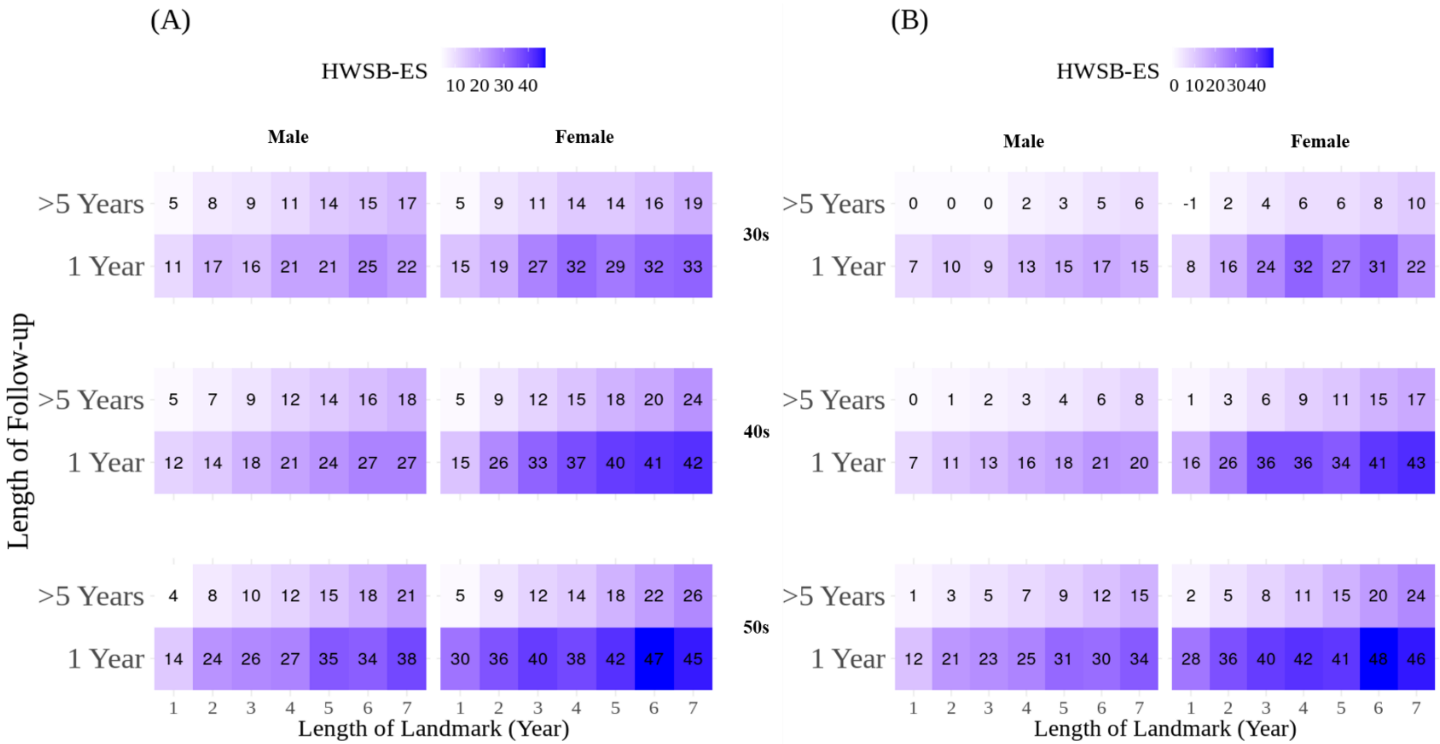


Figure 14. A heatmap of estimated HWSB–ES according to age group, sex, length of landmark, and length of follow-up in (A) fixed and (B) dynamic cohorts for the application of extrapolation

Finally, to enhance understanding and application of the HWSB-ES concept in previous studies, a specific function, 'HWSB_ES_Extrapolation,' was developed. This function calculates expected RRs and their 95% confidence intervals through the extrapolation of age-standardized HWSB-ES. The development process followed these steps.

1) Data import and preparation

The function initiates by importing pre-calculated age-standardized HWSB-ES values along with their standard errors, segmented demographically by age groups (30s, 40s, 50s) and sex (male and female). These values are gathered based on varying cohort types (dynamic or fixed), follow-up durations (short- or long-term), and landmark period lengths (ranging from 1 year to 7–10 years). The 95% CI of these age-standardized HWSB-ES values, utilized in the function, is derived using the Poisson distribution.

2) Demographic proportion calculation

Recognizing that the age and sex distribution in previous studies might differ from the general population, this function estimates the proportion of each demographic segment. It uses the mean age and standard deviation of the previous study population, assuming a normal distribution. Ages above 59 or below 30 are grouped into the 50s and 30s categories, respectively. For studies lacking specific demographic data, general population metrics are applied.

3) Weighted HWSB–ES calculation

The core of the function involves the computation of a weighted HWSB-ES value and its 95% CI. This is done by multiplying the HWSB-ES values for each demographic segment by their respective proportions in the study population. The sum of these weighted values yields an overall HWSB-ES value and 95% CI for each landmark period, cohort type, and follow-up duration, reflecting the demographics of the previous study.

4) Estimation of expected RR (95% CI)

The function calculates the expected RR (referred to as RRorigin in this study) using the overall weighted HWSB-ES and the RR from the previous study with the formula:

$$expected\ \ RR = RR\ \ of\ \ previous\ \ study \times \frac{1}{1 - weighted\ \ HWSB - ES}$$

Additionally, it estimates the 95% confidence interval of the expected RR using the standard error of HWSB-ES.

5) Application of the function

To use this function, several arguments are required: the mean age and standard deviation (SD) for each sex, the proportion of males, and the RR from the previous study. Additional inputs include the length of the landmark period (year), cohort setting (fixed or dynamic), and follow-up duration (short- or long-term). Landmark period can be selected from 1–10 years for short-term follow-up and 1–7 years for long-term follow-up. For instance, if the SMR of employee workers was 1.07 with a 55% male proportion and mean ages (SD) of 46.3±5.0 for males and 43.2±5.5 for females, setting a 5-year landmark period in a fixed cohort with long-term follow-up would result in an expected RR (95% CI) of 1.27 (1.25−1.28) using the 'HWSB_ES_Extrapolation' function. The R codes for “HWSB_ES_Extrapolation” function presented in Appendix 19. Note that this function serves as an example applicable in scenarios where the previous study provides data on the proportion of males and age distribution for each sex. So the function for the situation of studies with non-stratification of sex or lack of information about age distribution should be further developed. Therefore, for studies that do not stratify data by sex or lack detailed information on age distribution, further development of this function is required.

In addition, the expected RR extrapolated by HWSB−ES should be understood as an estimated rather than an exact figure. It helps us understand the potential range of the original study's RR when considering the HWSB. As such, the confidence interval of the original study may not play a major role in this context. This approach mainly serves as a tool for finding the need for further well-designed investigation of harmful occupational exposures that potentially overlooked owing to HWSB. Hence, the estimated expected RR and its 95% CI should be used for exploratory purposes only and not be over interpreted, even if statistically significant.

To enhance its accessibility for researchers, an interactive application was developed using Shinyapp in R. The 'HWSB−ES' data tab displays HWSB−ES values and their 95% confidence intervals (CI) as researchers select the cohort type, follow-up duration, and length of the landmark period (see Figure 15). The subsequent tab, 'expected RR calculation', computes the expected RR using the demographic distribution and RR from the existing study, along with weighted HWSB-ES estimates. Researchers can input values from previous studies (indicated by a navy box), choose the cohort setting, follow-up duration, and length of landmark (indicated by a red box), and press the calculation button to generate the expected RR and its error bar plot on the right side of the page (refer to Figure 16). This application enables researchers to swiftly examine previous studies, allowing for various interpretations that consider HWSB.

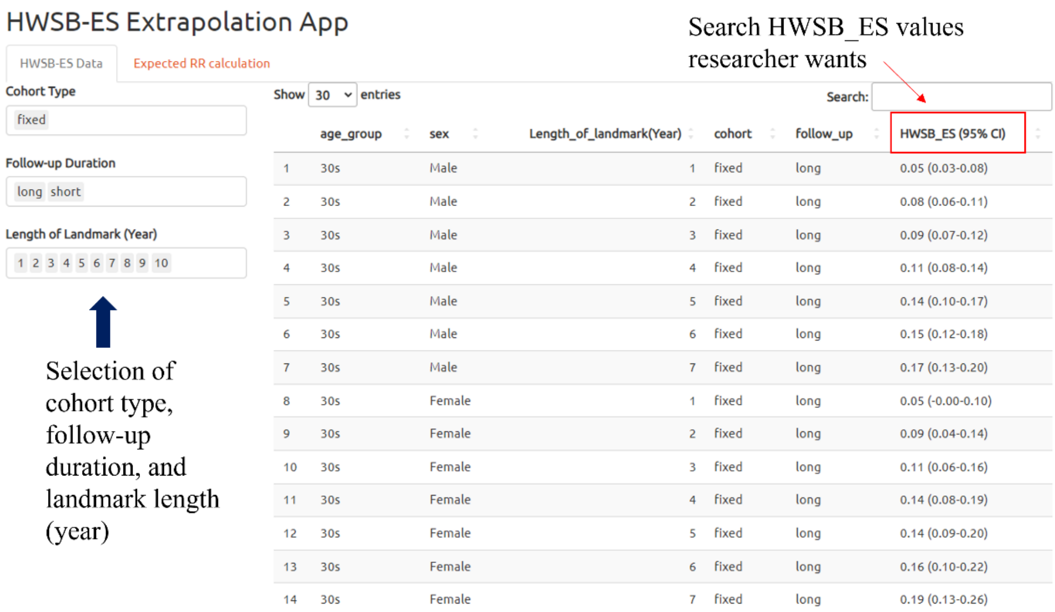


Figure 15. Interactive application for searching HWSB–ES values and 95% CIs based on the length of landmark, cohort type, and follow-up duration

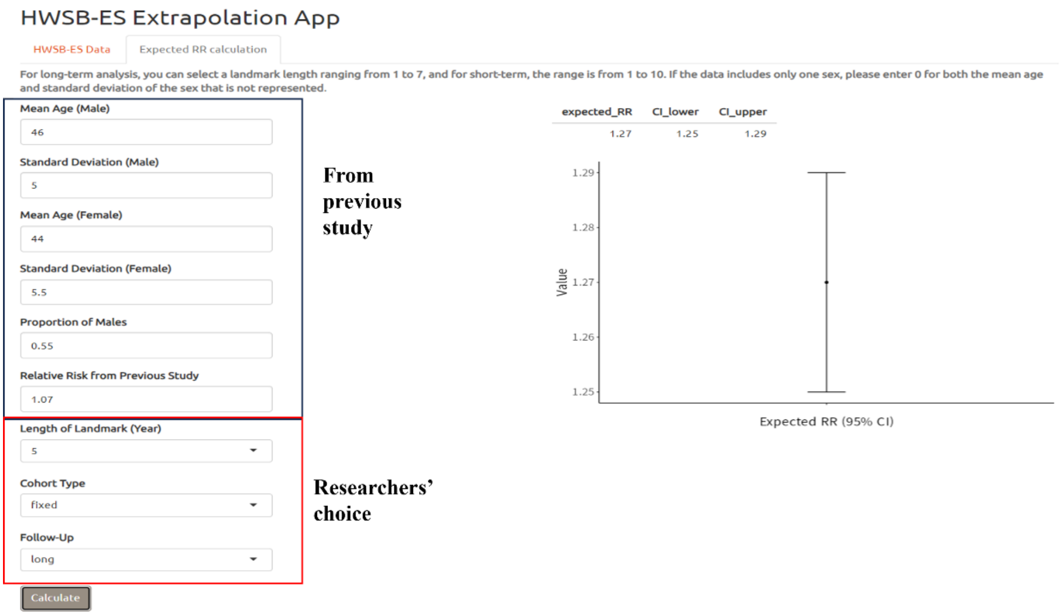


Figure 16. Application for calculating expected RR: utilizing RR from previous studies and extrapolating weighted HWSB−ES estimates based on the length of landmark, cohort type, and follow-up duration

4. Strengths and limitations

The main strength of this study was the comparison of mortality risk among different employment statuses when considering SES, including income levels and industrial traits. Moreover, based on a data-driven decision-making approach with real-world evidence, the different extents of HWSB–ES in Korea were estimated from a nationwide cohort that covered 97.2% of the Republic of Korea's population[70]. A more advanced interpretation of the existing literature is possible by using the method of extrapolating the HWSB estimate from previous literature. The advantage of this approach is its simplicity, which makes it readily adaptable to previous research. Furthermore, stratifying the estimation of the HWSB–ES by sex also offers advantages in interpreting previous studies by considering differences in social roles. This study also examined high-risk populations for health checkup non-participation; the significant additive interaction and mediation effect of health check-up participation underscores the significance of promoting participation in them.

However, this study also has several limitations. First, due to the nature of the NHIS, a lack of information, including education, marital status, cause of death, and occupational classification, could be attributed to bias from unmeasured confounders. The extent of harmful exposure among individuals could not be examined without information on occupational classification. Further well-designed cohort studies with unmeasured confounders should be conducted to overcome this limitation. If possible, the relationship between work environment measurements or special health check-up data and the NHIS database should be examined.

Second, the size of the HWSB was estimated solely based on changes in employment status. Since the reason for the employment status change is unknown, the HWSB size was estimated with a broader spectrum, regardless of the contribution of specific reasons for unemployment, such as health reasons, childcare, and pregnancy. Third, it is essential to exercise caution when applying the HWSB estimate to the existing literature because the patterns of employment and job stability can vary significantly depending on factors such as country, historical events, and the time period under consideration. For instance, in the case of South Korea, job stability has declined since the International Monetary Fund bailout of 1997–1998[71], and significant decrease in the unemployment rate during the 2008 financial crisis and COVID-19 pandemic.[72] Moreover, the proportion of temporary, part-time, and nonstandard employment has been rising globally, and the extent of this trend differs from country to country.[73] Therefore, using the HWSB estimate in the interpretation of the existing literature should be restricted to specific periods in Korea.

To expand the concept of HWSB–ES worldwide, a methodology similar to the Global Burden of Disease (GBD) approach should be developed and implemented. For example, the International Agency for Research on Cancer regularly updates information on the worldwide burden of cancer, as demonstrated in its most recent GLOBOCAN 2020 report, detailing estimates of cancer incidence and mortality rates.[74] This approach, which involves the systematic collection and analysis of mortality data in relation to employment status and its changes by country, may contribute to a better understanding of the occupational research field.

V. CONCLUSION

This study highlighted the significant association between employment status and all-cause mortality among the middle-aged population in Korea, with the extent of risk differing according to socioeconomic factors. This study also estimated the extent of HWSB–ES in the entire Korean population using a data-driven decision-making approach and exemplified the application of the estimated HWSB–ES to the existing literature. When performing occupational studies in Korea regarding hazardous exposures and health outcomes, the extent of the HWSB–ES measured in this study could be applied to avoid risk underestimation. Furthermore, there is a need to broaden our comprehension of the health effects associated with diverse socioeconomic traits. This can be achieved by estimating the survival effect of healthy workers in individual countries, considering the specific characteristics unique to each country.

REFERENCES

1. Li CY, Sung FC. A review of the healthy worker effect in occupational epidemiology. Occup Med (Lond) 1999;49:225-9.

2. McMichael AJ. Standardized mortality ratios and the "healthy worker effect": Scratching beneath the surface. J Occup Med 1976;18:165-8.

3. Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. Br J Prev Soc Med 1976;30:225-30.

4. Fox AJ, Collier PF. Mortality experience of workers exposed to vinyl chloride monomer in the manufacture of polyvinyl chloride in Great Britain. Br J Ind Med 1977;34:1-10.

5. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. Epidemiology 1994;5:189-96.

6. Buckley JP, Keil AP, McGrath LJ, Edwards JK. Evolving methods for inference in the presence of healthy worker survivor bias. Epidemiology 2015;26:204-12.

7. Chowdhury R, Shah D, Payal AR. Healthy Worker Effect Phenomenon: Revisited with Emphasis on Statistical Methods - A Review. Indian J Occup Environ Med 2017;21:2-8.

8. Chowdhury R, Darrow L, McClellan W, Sarnat S, Steenland K. Incident ESRD among participants in a lead surveillance program. Am J Kidney Dis 2014;64:25-31.

9. Chowdhury R, Sarnat SE, Darrow L, McClellan W, Steenland K. Mortality among participants in a lead surveillance program. Environ Res 2014;132:100-4.

10. Roberts MH, Sim MR, Black O, Smith P. Occupational injury risk among ambulance officers and paramedics compared with other healthcare workers in Victoria, Australia: analysis of workers' compensation claims from 2003 to 2012. Occup Environ Med 2015;72:489-95.

11. Buckley JP, Keil AP, McGrath LJ, Edwards JKJE. Evolving methods for inference in the presence of healthy worker survivor bias. 2015;26:204-12.

12. Hertz-Picciotto I, Arrighi HM, Hu S-WJAjoe. Does arsenic exposure increase the risk for circulatory disease? 2000;151:174-81.

13. Keil AP, Richardson DB. Reassessing the Link between Airborne Arsenic Exposure among Anaconda Copper Smelter Workers and Multiple Causes of Death Using the Parametric g-Formula. Environ Health Perspect 2017;125:608-14.

14. Richardson DB, Cole SR, Chu H, Langholz B. Lagging exposure information in cumulative exposure-response analyses. Am J Epidemiol 2011;174:1416-22.

15. Garshick E, Laden F, Hart JE, Davis ME, Eisen EA, Smith TJ. Lung cancer and elemental carbon exposure in trucking industry workers. Environ Health Perspect 2012;120:1301-6.

16. Gilbert ES. Some confounding factors in the study of mortality and occupational exposures. Am J Epidemiol 1982;116:177-88.

17. Dumas O, Le Moual N, Siroux V, Heederik D, Garcia-Aymerich J, Varraso R, et al. Work related asthma. A causal analysis controlling the healthy worker effect. Occup Environ Med 2013;70:603-10.

18. Thygesen LC, Hvidtfeldt UA, Mikkelsen S, Brønnum-Hansen H. Quantification of the healthy worker effect: a nationwide cohort study among electricians in Denmark. BMC Public Health 2011;11:571.

19. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chronic Dis 1987;40 Suppl 2:139s-61s.

20. Hertz-Picciotto I, Arrighi HM, Hu SW. Does arsenic exposure increase the risk for

circulatory disease? Am J Epidemiol 2000;151:174-81.

21. Neophytou AM, Picciotto S, Hart JE, Garshick E, Eisen EA, Laden F. A structural approach to address the healthy-worker survivor effect in occupational cohorts: an application in the trucking industry cohort. Occup Environ Med 2014;71:442-7.

22. Naimi AI, Cole SR, Hudgens MG, Richardson DB. Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. Epidemiology 2014;25:246-54.

23. Cole SR, Richardson DB, Chu H, Naimi AI. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. Am J Epidemiol 2013;177:989-96.

24. Cho SS, Kang MY. Association between occupational exposure to chemical or physical factors and sleep disturbance: An analysis of the fifth Korean Working Conditions Survey. Sleep Health 2022;8:521-7.

25. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014;43:69-77.

26. Kang YJ, Lee DW, Park MY, Kang MY. Employment Status Transition Predicts Adult Obesity Trajectory: Evidence From a Cohort Study in South Korea. J Occup Environ Med 2021;63:e861-e7.

27. Cho HW, Song IA, Oh TK. Quality of Life and Long-Term Mortality Among Survivors of Extracorporeal Membrane Oxygenation: A Nationwide Cohort Study in South Korea. Crit Care Med 2021;49:e771-e80.

28. Pappaioanou M, Malison M, Wilkins K, Otto B, Goodman RA, Churchill RE, et al. Strengthening capacity in developing countries for evidence-based public health:: the data

for decision-making project. 2003;57:1925-37.

29. Choi BC. Definition, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect. J Occup Med 1992;34:979-88.

30. Saegert SC, Adler NE, Bullock HE, Cauce AM, Liu WM, Wyche KFJW, DC: American Psychological Association. Report of the APA task force on socioeconomic status. 2007.

31. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, et al. Socioeconomic status and health. The challenge of the gradient. Am Psychol 1994;49:15-24.

32. Hill-Briggs F, Adler NE, Berkowitz SA, Chin MH, Gary-Webb TL, Navas-Acien A, et al. Social Determinants of Health and Diabetes: A Scientific Review. Diabetes Care 2020;44:258-79.

33. Shin DW, Cho J, Park JH, Cho BJP, Medicine F. National General Health Screening Program in Korea: history, current status, and future direction. 2022;6:9-31.

34. Amoah PA, Boateng MO, Koduah AO, Acheampong PR. Interplay of health literacy, healthcare access and health behaviors with oral health status among older persons. Front Public Health 2022;10:997987.

35. Shin HY, Kang HT, Lee JW, Lim HJ. The Association between Socioeconomic Status and Adherence to Health Check-up in Korean Adults, Based on the 2010-2012 Korean National Health and Nutrition Examination Survey. Korean J Fam Med 2018;39:114-21.

36. Park M, Lee JS, Kim YH. Association of physical activity, smoking, and socioeconomic factors on health checkup participation in community-dwelling stroke survivors aged 50 years or older. BMC Public Health 2023;23:502.

37. Lee J. Health Inequality in Health Checkups. Korean J Fam Med 2018;39:65-6.

38. Gleiss A, Oberbauer R, Heinze G. An unjustified benefit: immortal time bias in the

analysis of time-dependent events. Transpl Int 2018;31:125-30.
39. Cheol Seong S, Kim Y-Y, Khang Y-H, Heon Park J, Kang H-J, Lee H, et al. Data resource profile: the national health information database of the National Health Insurance Service in South Korea. 2017;46:799-800.
40. Kwon S. Thirty years of national health insurance in South Korea: lessons for achieving universal health care coverage. Health Policy Plan 2009;24:63-71.
41. Kim HK, Song SO, Noh J, Jeong I-K, Lee B-WJD, Journal M. Data configuration and publication trends for the Korean national health insurance and health insurance review & assessment database. 2020;44:671-8.
42. labor moea. Article 19, Act on prohibition of age discrimination in employment and elderly employment promotion. In: labor moea, editor.; 2016.
43. Lee C, Sung NJ, Lim HS, Lee JH. Emergency Department Visits Can Be Reduced by Having a Regular Doctor for Adults with Diabetes Mellitus: Secondary Analysis of 2013 Korea Health Panel Data. J Korean Med Sci 2017;32:1921-30.
44. Park JH, Hong JY, Han K, Shen JJ. Association Between Glycemic Status and the Risk of Kidney Cancer in Men and Women: A Nationwide Cohort Study. Diabetes Care 2023;46:38-45.
45. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012;41:514-20.
46. Steen J, Loeys T, Moerkerke B, Vansteelandt SJJoSS. Medflex: an R package for flexible mediation analysis using natural effect models. 2017;76:1-46.
47. Lim GY, Jung NY, Jun KY, Kang JY, Kim MK, Lee HE, et al. Pregnancy loss and Income in the Republic of Korea using National Health Insurance Service Data, 2008-2014. BMC Public Health 2022;22:188.

48. Kim C, Cho Y. Does Unstable Employment Have an Association with Suicide Rates among the Young? Int J Environ Res Public Health 2017;14.

49. Min KB, Park SG, Hwang SH, Min JY. Precarious employment and the risk of suicidal ideation and suicide attempts. Prev Med 2015;71:72-6.

50. Cho Y, Choi S, Kim K, Park SM. Sociodemographic Factors, Health Behaviors, and Biological Indicators Associated With Suicide Mortality Among Young Adults in South Korea: A Nationwide Cohort Study Among 15 Million Men and Women. J Clin Psychiatry 2020;82.

51. Hagqvist E, Toivanen S, Vinberg S. Time strain among employed and self-employed women and men in Sweden. Society, Health & Vulnerability 2015;6:29183.

52. Matilla-Santander N, Muntaner C, Kreshpaj B, Gunn V, Jonsson J, Kokkinen L, et al. Trajectories of precarious employment and the risk of myocardial infarction and stroke among middle-aged workers in Sweden: A register-based cohort study. Lancet Reg Health Eur 2022;15:100314.

53. Schonfeld IS, Mazzola JJ. A qualitative study of stress in individuals self-employed in solo businesses. J Occup Health Psychol 2015;20:501-13.

54. Krittanawong C, Kumar A, Wang Z, Baber U, Bhatt DL. Self-employment and cardiovascular risk in the US general population. Int J Cardiol Hypertens 2020;6:100035.

55. Park BH, Lee BK, Ahn J, Kim NS, Park J, Kim Y. Association of Participation in Health Check-ups with Risk Factors for Cardiovascular Diseases. J Korean Med Sci 2021;36:e19.

56. Kim S, Song JH, Oh YM, Park SM. Disparities in the utilisation of preventive health services by the employment status: An analysis of 2007-2012 South Korean national survey. PLoS One 2018;13:e0207737.

57. Baron RAJAomR. The role of affect in the entrepreneurial process. 2008;33:328-40.

58. Wiklund J, Nikolaev B, Shir N, Foo M-D, Bradley SJJobv. Entrepreneurship and well-being: Past, present, and future. 2019;34:579-88.

59. Garcy AM, Vågerö D. The length of unemployment predicts mortality, differently in men and women, and by cause of death: a six year mortality follow-up of the Swedish 1992-1996 recession. Soc Sci Med 2012;74:1911-20.

60. Hammarström A, Gustafsson PE, Strandh M, Virtanen P, Janlert U. It's no surprise! Men are not hit more than women by the health consequences of unemployment in the Northern Swedish Cohort. Scand J Public Health 2011;39:187-93.

61. Norström F, Janlert U, Hammarström A. Is unemployment in young adulthood related to self-rated health later in life? Results from the Northern Swedish cohort. BMC Public Health 2017;17:529.

62. Norström F, Waenerlund AK, Lindholm L, Nygren R, Sahlén KG, Brydsten A. Does unemployment contribute to poorer health-related quality of life among Swedish adults? BMC Public Health 2019;19:457.

63. Heggebø K. Gendered health consequences of unemployment in Norway 2000-2017: a register-based study of hospital admissions, health-related benefit utilisation, and mortality. BMC Public Health 2022;22:2447.

64. Forret ML, Sullivan SE, Mainiero LAJJoOB. Gender role differences in reactions to unemployment: Exploring psychological mobility and boundaryless careers. 2010;31:647-66.

65. Um K, Yang S-EJJoF, Life B. A qualitative study on the career interrupted and the child care of married women. 2011;29:21-40.

66. Tattarini G, Grotti R. Gender roles and selection mechanisms across contexts: a comparative analysis of the relationship between unemployment, self-perceived health and gender. Sociol Health Illn 2022;44:641-62.

67. Raub S, Borzillo S, Perretten G, Schmitt AJIjohm. New employee orientation, role-related stressors and conflict at work: Consequences for work attitudes and performance of hospitality employees. 2021;94:102857.

68. Lee WJ, Cha ES, Moon EK. Disease prevalence and mortality among agricultural workers in Korea. J Korean Med Sci 2010;25:S112-8.

69. Yoon JH, Ahn YS. Cause-specific mortality due to malignant and non-malignant disease in Korean foundry workers. PLoS One 2014;9:e88264.

70. Song SO, Jung CH, Song YD, Park C-Y, Kwon H-S, Cha BS, et al. Background and data configuration process of a nationwide population-based study using the korean national health insurance system. 2014;38:395-403.

71. Kim Do, Kim SJIRAJoE, Society. Globalization, financial crisis, and industrial relations: The case of South Korea. 2003;42:341-67.

72. Nam M, Lee SJSJoE. COVID-19 and employment in South Korea: trends and comparison with the 2008 financial crisis. 2021;34.

73. organisation Il. Non-standard employment around the world: Understanding challenges, shaping prospects: International Labour Office Geneva; 2016.

74. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.

APPENDICES

Appendix 1. Diagnostic ICD-10 codes for calculating CCI score.

Appendix 2. Definitions of industry classification based on industry section from Korea standard industry classification.

Appendix 3. Adjusted HR (95% CI) of secondary outcomes associated with baseline employment status.

Appendix 4. Sensitivity analyses on the risk of primary and secondary outcomes associated with employment status before the COVID-19 pandemic.

Appendix 5. Adjusted HR (95% CI) of occupational diseases associated with employment status among individuals stratifying industry classification.

Appendix 6. The interaction of health check-up non-participation and employment status on the risk of all-cause mortality among ages 40–49.

Appendix 7. The interaction of health check-up non-participation and employment status on the risk of all-cause mortality among ages 50–59.

Appendix 8. Mediating effect of health check-up non-participation on the association between employment status and 5-year mortality.

Appendix 9. The risk of all-cause mortality associated with employment status among female individuals in their 30s stratified by pregnancy status.

Appendix 10. Estimation of crude short-term HWSB–ES from the fixed cohort.

Appendix 11. Estimation of crude short-term HWSB–ES from the dynamic cohort.

Appendix 12. Estimation of age-adjusted HWSB–ES from the fixed cohort.

Appendix 1. Diagnostic ICD-10 codes for calculating CCI score

Diseases	ICD-10 code	CCI scores
Diabetes mellitus	E10-E14	1
Myocardial infarct	I21, I22, I25	1
Congestive heart failure	I50	1
Peripheral vascular disease	I70-I79	1
Cerebrovascular disease	I60-I69	1
Dementia	F03, G30	1
Chronic pulmonary disease	J41-45, J47, J64	1
Rheumatic or connective tissue disease	M30-M36, M06	1
Gastric or peptic ulcer	K25, K26	1
Mild liver disease	B18, B19, K70-K77	1
Hemiplegia or paraplegia	G80-G82	2
Moderate or severe renal disease	N17-N19	2
Any malignancy, including lymphoma and leukemia, except basal cell cancer of skin and thyroid cancer	C00-C41, C43, C45-C72, C74, C75, C81-C96	2
Metastatic solid tumor	C76-C80	6
Acquired immune deficiency syndrome	B20-B24	6

Abbreviation: ICD, international classification of disease; CCI, charlson comorbidity index

Appendix 2. Definitions of industry classification based on industry section from Korea standard industry classification

Industry Classification	**Industry Section**
Public/Private school	Public Service
	Private School
1st/2nd industry	Agriculture, forestry and fishing
	Mining and quarrying
	Manufacturing
	Electricity, gas, steam and air conditioning supply
	Water supply; sewage, waste management, materials recovery
	Construction
3rd industry	Wholesale and retail trade
	Transportation and storage
	Accommodation and food service activities
	Information and communication
	Financial and insurance activities
	Real estate activities
	Professional, scientific and technical activities
	Business facilities management and business support services; rental and leasing activities
	Public administration and defense; compulsory social security
	Education
	Human health and social work activities
	Arts, sports and recreation related services
	Membership organizations, repair and other personal services

Appendix 3. Adjusted HR (95% CI) of secondary outcomes associated with baseline employment status

Age group	Sex	Employment Status	All types of Malignancies*	CVD	Mental diseases	Liver diseases
30s	Male	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.07 (1.04-1.10)	1.52 (1.48-1.56)	2.34 (2.29-2.39)	1.54 (1.52-1.56)
		Unemployed	1.03 (1.04-1.06)	1.61 (1.57-1.65)	3.33 (3.27-3.39)	1.46 (1.44-1.48)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	0.97 (0.94-0.99)	1.97 (1.87-2.08)	2.81 (2.74-2.88)	1.89 (1.84-1.94)
		Unemployed	0.92 (0.91-0.94)	1.30 (1.25-1.35)	1.80 (1.76-1.83)	1.20 (1.18-1.22)
40s	Male	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.09 (1.08-1.11)	1.49 (1.47-1.51)	1.94 (1.91-1.97)	1.60 (1.58-1.62)
		Unemployed	19 (1.17-1.21)	1.72 (1.69-1.74)	3.68 (3.62-3.74)	2.19 (2.16-2.22)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.01 (0.99-1.02)	1.76 (1.71-1.82)	1.92 (1.88-1.96)	1.61 (1.58-1.64)
		Unemployed	0.99 (0.98-1.00)	1.37 (1.33-1.40)	1.48 (1.46-1.51)	1.12 (1.18-1.22)
50s	Male	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.05 (1.04-1.06)	1.38 (1.37-1.40)	1.58 (1.56-1.61)	1.41 (1.40-1.43)
		Unemployed	1.20 (1.19-1.21)	1.60 (1.58-1.62)	2.75 (2.71-2.79)	2.13 (2.11-2.16)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.09 (1.07-1.11)	1.63 (1.59-1.67)	1.65 (1.61-1.68)	1.38 (1.35-1.41)
		Unemployed	1.05 (1.04-1.07)	1.39 (1.36-1.42)	1.46 (1.44-1.49)	1.19 (1.17-1.21)

All models were adjusted for age, residential area, and disease status
* thyroid cancer and basal skin cancer were excluded. 4-year time lag was applied due to the chronic nature of malignancy
Abbreviation: HR, hazard ratio; CI, confidence interval; CVD, cardiovascular diseases

Appendix 4. Sensitivity analyses on the risk of primary and secondary outcomes associated with employment status before the COVID-19 pandemic

			adjusted HR (95% CI)				
Age group	**Sex**	**Employment Status**	**All-cause Mortality**	**Cancer**	**CVD**	**mental diseases**	**liver diseases**
30s	**Male**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	2.21 (2.14-2.28)	0.98 (0.95-1.01)	1.50 (1.45-1.55)	2.26 (2.20-2.32)	1.58 (1.55-1.61)
		Unemployed	3.10 (3.01-3.18)	1.00 (0.96-1.03)	1.57 (1.52-1.62)	2.58 (2.51-2.64)	1.39 (1.37-1.42)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	3.42 (3.24-3.61)	0.97 (0.94-1.01)	2.15 (2.00-2.31)	2.75 (2.65-2.84)	2.13 (2.06-2.21)
		Unemployed	1.80 (1.72-1.88)	0.93 (0.92-0.95)	1.33 (1.27-1.41)	1.64 (1.59-1.68)	1.25 (1.21-1.28)
40s	**Male**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.98 (1.94-2.01)	1.05 (1.03-1.06)	1.48 (1.46-1.51)	1.90 (1.86-1.94)	1.63 (1.61-1.66)
		Unemployed	4.56 (4.48-4.64)	1.31 (1.28-1.33)	1.71 (1.67-1.75)	2.81 (2.74-2.88)	2.00 (1.96-2.03)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

		Self-Employed	2.14 (2.05-2.23)	0.99 (0.97-1.01)	1.83 (1.76-1.91)	1.82 (1.77-1.87)	1.65 (1.60-1.69)
		Unemployed	1.67 (1.62-1.73)	0.96 (0.95-0.97)	1.35 (1.31-1.40)	1.37 (1.34-1.40)	1.21 (1.18-1.24)
50s	**Male**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.48 (1.45-1.50)	1.01 (1.00-1.03)	1.41 (1.39-1.43)	1.52 (1.49-1.55)	1.40 (1.38-1.42)
		Unemployed	3.43 (3.39-3.48)	1.32 (1.31-1.35)	1.63 (1.61-1.66)	2.29 (2.24-2.34)	1.91 (1.87-1.94)
	Female	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.76 (1.69-1.82)	1.04 (1.02-1.07)	1.66 (1.60-1.71)	1.58 (1.54-1.63)	1.33 (1.30-1.37)
		Unemployed	1.54 (1.50-1.59)	1.01 (1.00-1.03)	1.35 (1.31-1.39)	1.36 (1.33-1.39)	1.15 (1.13-1.18)

All models are adjusted for age, residential area, and disease status.

Abbreviation: HR, hazard ratio; CI, confidence interval; COVID, coronavirus disease

Appendix 5. Adjusted HR (95% CI) of occupational diseases associated with employment status among individuals stratifying industry classification

				Public/Private school	1st/2nd industry	3rd industry
30s	Male	Mental Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	2.23 (2.11-2.36)	2.32 (2.25-2.39)	2.18 (2.11-2.25)
			Unemployed	2.49 (2.33-2.61)	2.56 (2.49-2.64)	2.41 (2.33-2.48)
		Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.28 (1.26-1.30)	1.32 (1.31-1.34)	1.35 (1.34-1.37)
			Unemployed	1.08 (1.06-1.10)	1.12 (1.11-1.13)	1.14 (1.13-1.15)
		Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.04 (1.01-1.06)	1.02 (1.01-1.03)	1.10 (1.09-1.11)
			Unemployed	0.83 (0.81-0.84)	0.82 (0.81-0.83)	0.87 (0.86-0.88)
	Female	Mental Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	2.73 (2.58-2.88)	2.90 (2.75-3.01)	2.71 (2.62-2.81)
			Unemployed	1.63 (1.55-1.72)	1.74 (1.65-1.82)	1.63 (1.58-1.67)
		Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.87 (1.83-1.91)	1.64 (1.60-1.67)	1.62 (1.60-1.65)
			Unemployed	1.26 (1.23-1.28)	1.10 (1.08-1.12)	1.09 (1.08-1.10)
		Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.57 (1.53-1.61)	1.24 (1.22-1.27)	1.34 (1.32-1.37)
			Unemployed	1.20 (1.17-1.23)	0.95 (0.94-0.97)	1.03 (1.02-1.04)
40s	Male	Mental Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.77 (1.71-1.84)	1.93 (1.89-1.98)	1.86 (1.82-1.91)

			Unemployed	2.60 (2.50-2.70)	2.83 (2.76-2.91)	2.74 (2.66-2.81)
		Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.34 (1.33-1.36)	1.24 (1.24-1.25)	1.39 (1.37-1.40)
			Unemployed	1.29 (1.27-1.31)	1.20 (1.19-1.21)	1.33 (1.32-1.34)
		Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	0.98 (0.97-1.00)	0.95 (0.94-0.95)	1.11 (1.10-1.12)
			Unemployed	0.86 (0.85-0.87)	0.83 (0.82-0.84)	0.97 (0.96-0.99)
	Female	**Mental Diseases**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.70 (1.62-1.77)	1.90 (1.83-1.98)	1.79 (1.74-1.85)
			Unemployed	1.27 (1.22-1.33)	1.43 (1.38-1.48)	1.35 (1.31-1.38)
		Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.53 (1.50-1.56)	1.18 (1.16-1.20)	1.26 (1.25-1.28)
			Unemployed	1.23 (1.21-1.25)	0.95 (0.94-0.96)	1.02 (1.01-1.03)
		Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.42 (1.39-1.44)	0.95 (0.94-0.97)	1.11 (1.10-1.12)
			Unemployed	1.23 (1.21-1.25)	0.83 (0.82-0.84)	0.96 (0.96-0.97)
50s	**Male**	**Mental Diseases**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.62 (1.57-1.68)	1.53 (1.50-1.57)	1.49 (1.46-1.53)
			Unemployed	2.39 (2.31-2.47)	2.25 (2.19-2.31)	2.19 (2.14-2.25)
		Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
			Self-Employed	1.40 (1.38-1.42)	1.15 (1.14-1.16)	1.30 (1.29-1.31)
			Unemployed	1.50 (1.48-1.52)	1.23 (1.22-1.24)	1.39 (1.38-1.41)
		Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)

		Self-Employed	1.08 (1.07-1.10)	0.93 (0.92-0.94)	1.10 (1.09-1.11)
		Unemployed	1.10 (1.08-1.11)	0.94 (0.93-0.95)	1.11 (1.10-1.13)
Female	**Mental Diseases**	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.52 (1.44-1.61)	1.62 (1.55-1.68)	1.52 (1.47-1.58)
		Unemployed	1.32 (1.25-1.39)	1.40 (1.35-1.46)	1.32 (1.28-1.36)
	Injury	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.30 (1.28-1.33)	1.06 (1.05-1.08)	1.10 (1.09-1.12)
		Unemployed	1.20 (1.18-1.23)	0.98 (0.97-0.99)	1.02 (1.01-1.03)
	Muscle Diseases	Employee	1.00 (reference)	1.00 (reference)	1.00 (reference)
		Self-Employed	1.34 (1.30-1.37)	0.92 (0.91-0.94)	1.02 (1.00-1.03)
		Unemployed	1.30 (1.27-1.33)	0.90 (0.89-0.91)	0.99 (0.98-1.00)

All models are adjusted for age, residential area, and disease status.

Appendix 6. The interaction of health check-up non-participation and employment status on the risk of all-cause mortality among ages 40−49

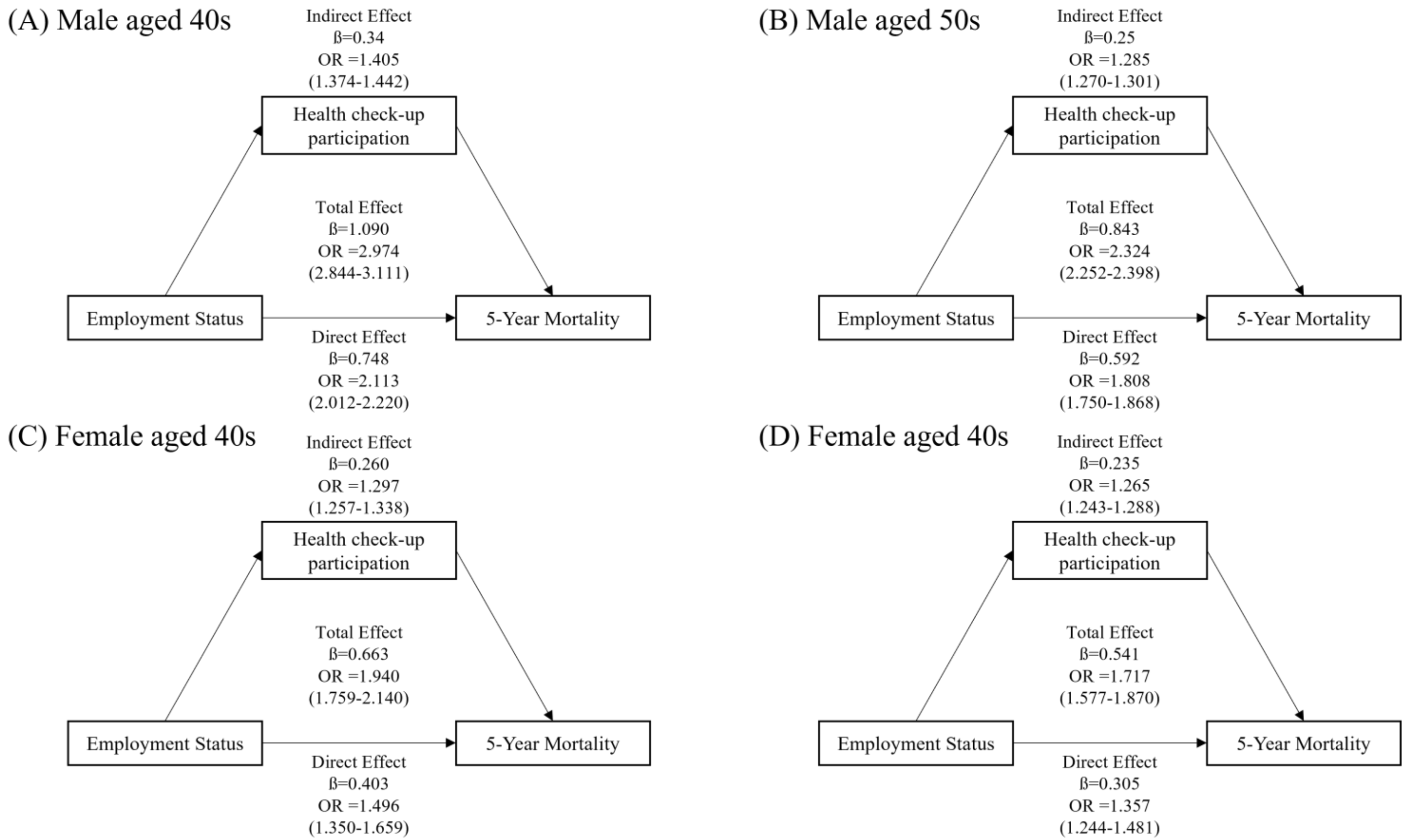
(A) Male	**Employee**			**Non-Employee**			
Health Check-ups	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI) for non-employee within strata of health check-up participation**
Participation	17,129,787	146.80	1.00 (reference)	4,457,560	251.82	1.53 (1.50−1.57)	1.53 (1.50−1.57)
Non-participation	4,713,573	198.21	1.38 (1.35−1.42)	13,246,325	477.63	3.30 (3.25−3.35)	2.39 (2.32−2.44)
Adjusted HR (95% CI) for non-participation in health check-ups within strata of employment status			1.38 (1.35−1.42)			2.15 (2.11−2.19)	Additive scale (RERI): 1.38 (1.33−1.43) Multiplicative Scale: 1.56 (1.51−1.61)
(B) Female	**Employee**			**Non-Employee**			
Health Check-ups	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI) for non-employee within strata of health check-up participation**
Participation	7,233,456	67.06	1.00 (reference)	12,136,620	88.18	1.81 (1.77−1.86)	1.81 (1.77−1.86)
Non-participation	1,551,377	85.54	1.33 (1.25−1.42)	19,079,155	146.66	2.21 (2.14−2.28)	1.66 (1.58−1.72)
Adjusted HR (95% CI) for non-participation in health check-ups within strata of employment status			1.33 (1.25−1.42)			1.22 (1.18−1.26)	Additive scale (RERI): 0.66 (0.58−0.74) Multiplicative Scale: 1.36 (1.28−1.45)

All models were adjusted for age, residential area, and disease status

Appendix 7. The interaction of health check-up non-participation and employment status on the risk of all-cause mortality among ages 50−59

(A) Male	**Employee**			**Non-Employee**			
Health Check-ups	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI) for non-employee within strata of health check-up participation**
Participation	10,783,543	365.73	1.00 (reference)	6,521,351	528.13	1.29 (1.27−1.30)	1.29 (1.27−1.30)
Non-participation	2,852,226	426.82	1.21 (1.18−1.23)	10,036,382	951.56	2.68 (2.65−2.71)	2.22 (2.19−2.25)
Adjusted HR (95% CI) for non-participation in health check-ups within strata of employment status			1.21 (1.18−1.23)			2.09 (2.06−2.11)	Additive scale (RERI): 1.19 (1.16−1.22) Multiplicative Scale: 1.73 (1.69−1.77)
(B) Female	**Employee**			**Non-Employee**			
Health Check-ups	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Person-Year**	**Rate**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI) for non-employee within strata of health check-up participation**
Participation	4,373,829	130.92	1.00 (reference)	16,248,343	171.52	1.16 (1.13−1.20)	1.16 (1.13−1.20)
Non-participation	722,223	167.68	1.34 (1.26−1.43)	13,170,783	283.76	2.14 (2.09−2.20)	1.60 (1.52−1.66)
Adjusted HR (95% CI) for non-participation in health check-ups within strata of employment status			1.34 (1.26−1.43)			1.85 (1.82−1.87)	Additive scale (RERI): 0.64 (0.56−0.72) Multiplicative Scale: 1.37 (1.29−1.50)

All models were adjusted for age, residential area, and disease status



Appendix 8. Mediating effect of health check-up non-participation on the association between employment status and 5-year mortality.

Appendix 9. The risk of all-cause mortality associated with employment status among female individuals in their 30s stratified by pregnancy status

Pregnancy	**Employment Status**	**Total**	**Rate**	**adjusted HR (95% CI)**
Yes	Employee	495816	26.02	reference (1.00)
	Self-Employed	80187	74.45	2.52 (2.30−2.76)
	Unemployed	1044358	50.53	1.72 (1.63−1.82)
No	Employee	409368	54.28	reference (1.00)
	Self-Employed	152626	173.03	2.51 (2.45−2.57)
	Unemployed	833866	99.87	2.01 (1.97−2.05)

Abbreviation HR, hazard ratio; CI, confidence interval

Appendix 10. Estimation of crude short-term HWSB–ES from the fixed cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.112442	0.114396	0.145024
	2	0.169151	0.134813	0.23936
	3	0.157073	0.178675	0.26657
	4	0.209921	0.207101	0.283468
	5	0.213295	0.23625	0.361496
	6	0.248327	0.267108	0.351878
	7	0.224451	0.270203	0.397068
	8	0.235062	0.317844	0.399509
	9	0.290264	0.326769	0.414044
	10	0.293833	0.349601	0.439478
Female	1	0.148929	0.159216	0.306513
	2	0.185208	0.257385	0.372331
	3	0.267123	0.329493	0.408663
	4	0.304227	0.370949	0.38884
	5	0.277863	0.405962	0.431346
	6	0.302453	0.407846	0.484556
	7	0.304998	0.41872	0.479967
	8	0.370228	0.416772	0.496468
	9	0.35508	0.457599	0.509928
	10	0.347207	0.414986	0.514632

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Appendix 11. Estimation of crude short-term HWSB−ES from the dynamic cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.070229	0.073865	0.124995
	2	0.104924	0.102359	0.200849
	3	0.096277	0.126744	0.236712
	4	0.138001	0.160144	0.258751
	5	0.157283	0.18011	0.314199
	6	0.179639	0.209285	0.313154
	7	0.158925	0.201623	0.352475
	8	0.173762	0.235992	0.359152
	9	0.185366	0.239099	0.375717
	10	0.223957	0.265567	0.404717
Female	1	0.077947	0.163796	0.277428
	2	0.144156	0.266174	0.363234
	3	0.214184	0.355198	0.401428
	4	0.300373	0.36645	0.422514
	5	0.237368	0.343931	0.415991
	6	0.273751	0.405383	0.477983
	7	0.197783	0.424487	0.461619
	8	0.282499	0.403056	0.465441
	9	0.284553	0.449815	0.527634
	10	0.282713	0.437633	0.562115

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Appendix 12. Estimation of age-adjusted HWSB–ES from the fixed cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.116343	0.121037	0.152745
	2	0.17341	0.137363	0.251142
	3	0.135802	0.179412	0.272311
	4	0.204204	0.206997	0.273128
	5	0.20178	0.233728	0.352144
	6	0.242604	0.270893	0.335484
	7	0.19697	0.266476	0.374732
	8	0.189189	0.322767	0.384615
	9	0.263514	0.322289	0.37395
	10	0.252669	0.345133	0.392931
Female	1	0.147208	0.163972	0.313725
	2	0.184685	0.269618	0.377432
	3	0.269737	0.340757	0.412478
	4	0.318359	0.388889	0.388985
	5	0.275785	0.422	0.427365
	6	0.307527	0.423664	0.480287
	7	0.296471	0.435993	0.467257
	8	0.369615	0.428571	0.486179
	9	0.364606	0.473684	0.495066
	10	0.343458	0.420082	0.499182

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Appendix 13. Estimation of age-adjusted HWSB−ES from the dynamic cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.111406	0.103933	0.151869
	2	0.171582	0.151042	0.248899
	3	0.169863	0.188525	0.28355
	4	0.237598	0.238095	0.30123
	5	0.273632	0.265789	0.35625
	6	0.31068	0.306329	0.352362
	7	0.285012	0.2975	0.391473
	8	0.307087	0.345679	0.405354
	9	0.326316	0.353535	0.408922
	10	0.378307	0.38835	0.432727
Female	1	0.101235	0.195991	0.30064
	2	0.197872	0.316288	0.395131
	3	0.280808	0.422222	0.440066
	4	0.391608	0.445783	0.463768
	5	0.340594	0.430851	0.458529
	6	0.387755	0.500824	0.521311
	7	0.315895	0.525157	0.504026
	8	0.417603	0.511036	0.51173
	9	0.442341	0.561905	0.569546
	10	0.441989	0.552459	0.599709

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Appendix 14. Estimation of crude long-term HWSB–ES from the fixed cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.052769	0.048051	0.049604
	2	0.08139	0.072225	0.088545
	3	0.093795	0.094122	0.116927
	4	0.113088	0.116275	0.146139
	5	0.134692	0.136723	0.178393
	6	0.151439	0.161502	0.207568
	7	0.16871	0.181585	0.241215
Female	1	0.043735	0.047364	0.055798
	2	0.078047	0.086953	0.098431
	3	0.097867	0.11914	0.131508
	4	0.119952	0.150186	0.159124
	5	0.125033	0.185214	0.203086
	6	0.142768	0.202369	0.242515
	7	0.173936	0.238738	0.292651

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Appendix 15. Estimation of crude long-term HWSB−ES from the dynamic cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	-0.00234	-0.00043	0.016099
	2	0.007335	0.005713	0.042001
	3	0.006138	0.016108	0.062363
	4	0.024095	0.027826	0.083887
	5	0.033257	0.038793	0.109779
	6	0.053277	0.061017	0.138051
	7	0.066805	0.07793	0.172872
Female	1	-0.02168	0.008234	0.020465
	2	-0.00174	0.038351	0.052577
	3	0.013009	0.065944	0.0901
	4	0.02528	0.090182	0.120223
	5	0.01883	0.115957	0.154996
	6	0.037384	0.149418	0.206387
	7	0.056628	0.174443	0.252341

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Appendix 16. Estimation of age-adjusted long-term HWSB−ES from the fixed cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	0.0271	0.036269	0.036072
	2	0.044199	0.052219	0.064128
	3	0.047753	0.071618	0.09018
	4	0.063037	0.093085	0.106
	5	0.084548	0.10992	0.130261
	6	0.096096	0.134771	0.154
	7	0.108025	0.154891	0.184369
Female	1	0.035294	0.046099	0.049839
	2	0.064833	0.085106	0.08903
	3	0.083168	0.119857	0.118859
	4	0.106	0.151786	0.141494
	5	0.106557	0.186594	0.183413
	6	0.124481	0.204753	0.21865
	7	0.158562	0.243494	0.267628

Abbreviation: HWSB−ES, healthy worker survivor bias due to employment status change

Appendix 17. Estimation of age-adjusted long-term HWSB−ES from the dynamic cohort

Sex	Landmark (Year)	30s	40s	50s
Male	1	-0.00526	0.002551	0.011928
	2	0.020833	0.017677	0.039216
	3	0.033592	0.042821	0.064078
	4	0.069231	0.062344	0.078695
	5	0.094388	0.084158	0.106464
	6	0.127877	0.117647	0.136961
	7	0.149485	0.141463	0.173184
Female	1	-0.00963	0.022648	0.028617
	2	0.028409	0.068493	0.063963
	3	0.061798	0.113559	0.110599
	4	0.091078	0.148333	0.143511
	5	0.099251	0.187708	0.182094
	6	0.133581	0.233553	0.238671
	7	0.169145	0.269737	0.286567

Abbreviation: HWSB–ES, healthy worker survivor bias due to employment status change

Appendix 18. The proportion of the changed to non-employee among the employee population at the origin point, stratified by industry section and sex, over the length of the landmark period

Industry section	Sex	Employment status change	Length of landmark period (year)						
			1	**2**	**3**	**4**	**5**	**6**	**7**
Public Service	Male	Employee at origin point	427,588	426,967	426,341	425,693	424,978	424,275	423,453
		Changed to non-employee	3,742 (0.88)	10,849 (2.54)	15,183 (3.56)	29,208 (6.86)	40,349 (9.49)	49,312 (11.62)	59,598 (14.07)
	Female	Employee at origin point	261,266	261,147	260,999	260,850	260,666	260,510	260,257
		Changed to non-employee	2,116 (0.81)	4,875 (1.87)	7,611 (2.92)	12,189 (4.67)	18,309 (7.02)	22,617 (8.68)	27,199 (10.45)
Private School	Male	Employee at origin point	114,386	114,014	113,680	113,421	113,155	112,857	112,562
		Changed to non-employee	1,689 (1.48)	3,453 (3.03)	4,929 (4.34)	7,274 (6.41)	10,264 (9.07)	12,542 (11.11)	15,350 (13.64)
	Female	Employee at origin point	72,501	72,373	72,260	72,182	72,068	71,971	71,887
		Changed to non-employee	2,887 (3.98)	4,749 (6.56)	6,127 (8.48)	7,347 (10.18)	8,848 (12.28)	9,877 (13.72)	11,127 (15.48)
Agriculture, forestry and fishing	Male	Employee at origin point	12,657	12,375	12,244	12,186	12,142	12,092	11,980
		Changed to non-employee	939 (7.42)	1,269 (10.25)	1,476 (12.05)	1,827 (14.99)	2,037 (16.78)	2,223 (18.38)	2,423 (20.23)
	Female	Employee at origin point	4,808	4,770	4,760	4,753	4,752	4,750	4,735
		Changed to non-employee	431 (8.96)	693 (14.53)	839 (17.63)	978 (20.58)	1,030 (21.68)	1,108 (23.33)	1,179 (24.9)
Mining and quarrying	Male	Employee at origin point	11,522	11,434	11,392	11,356	11,321	11,293	11,251
		Changed to non-employee	780 (6.77)	1,228 (10.74)	1,506 (13.22)	1,885 (16.6)	1,989 (17.57)	2,200 (19.48)	2,331 (20.72)
	Female	Employee at origin point	1,754	1,748	1,744	1,740	1,733	1,735	1,727
		Changed to non-employee	160 (9.12)	271 (15.5)	316 (18.12)	395 (22.7)	429 (24.75)	478 (27.55)	472 (27.33)

Manufacturing	Male	Employee at origin point	1,661,263	1,640,734	1,633,090	1,629,725	1,626,564	1,622,644	1,616,918
		Changed to non-employee	78,171 (4.71)	118,966 (7.25)	141,837 (8.69)	170,272 (10.45)	198,057 (12.18)	225,023 (13.87)	246,825 (15.27)
	Female	Employee at origin point	413,878	411,166	410,323	409,789	409,331	408,819	408,138
		Changed to non-employee	41,098 (9.93)	65,013 (15.81)	79,207 (19.3)	93,233 (22.75)	104,185 (25.45)	112,451 (27.51)	118,556 (29.05)
Electricity, gas, steam and air conditioning supply	Male	Employee at origin point	59,436	59,342	59,265	59,174	59,087	59,002	58,895
		Changed to non-employee	1,243 (2.09)	2,201 (3.71)	2,749 (4.64)	3,358 (5.67)	3,938 (6.66)	4,233 (7.17)	4,967 (8.43)
	Female	Employee at origin point	6,598	6,594	6,589	6,588	6,587	6,583	6,582
		Changed to non-employee	234 (3.55)	413 (6.26)	516 (7.83)	603 (9.15)	689 (10.46)	730 (11.09)	771 (11.71)
Water supply; sewage, waste management, materials recovery	Male	Employee at origin point	15,397	15,245	15,203	15,156	15,113	15,066	15,012
		Changed to non-employee	879 (5.71)	1,421 (9.32)	1,697 (11.16)	1,974 (13.02)	2,225 (14.72)	2,401 (15.94)	2,554 (17.01)
	Female	Employee at origin point	3,208	3,196	3,197	3,193	3,188	3,185	3,182
		Changed to non-employee	274 (8.54)	385 (12.05)	504 (15.76)	584 (18.29)	627 (19.67)	691 (21.7)	742 (23.32)
Construction	Male	Employee at origin point	395,947	393,617	392,405	391,622	390,800	389,906	388,902
		Changed to non-employee	29,479 (7.45)	42,505 (10.8)	49,674 (12.66)	55,624 (14.2)	58,023 (14.85)	60,908 (15.62)	63,717 (16.38)
	Female	Employee at origin point	62,165	62,104	62,055	62,003	61,944	61,879	61,822
		Changed to non-employee	7,187 (11.56)	10,835 (17.45)	12,918 (20.82)	14,318 (23.09)	15,158 (24.47)	15,665 (25.32)	16,002 (25.88)
Wholesale and retail trade	Male	Employee at origin point	490,040	488,204	487,076	486,188	485,222	484,200	483,059
		Changed to non-employee	35,479 (7.24)	55,850 (11.44)	66,746 (13.7)	75,771 (15.58)	83,202 (17.15)	89,736 (18.53)	95,141 (19.7)
	Female	Employee at origin point	213,561	213,197	212,955	212,740	212,544	212,331	212,060

		Changed to non-employee	22,854 (10.7)	36,457 (17.1)	44,271 (20.79)	50,302 (23.64)	54,517 (25.65)	57,642 (27.15)	60,355 (28.46)
Transportation and storage	Male	Employee at origin point	308,589	307,762	306,925	305,996	304,974	303,988	302,871
		Changed to non-employee	17,894 (5.8)	29,046 (9.44)	36,117 (11.77)	43,661 (14.27)	50,099 (16.43)	56,460 (18.57)	62,390 (20.6)
	Female	Employee at origin point	38,066	38,037	37,995	37,968	37,938	37,915	37,867
		Changed to non-employee	2,759 (7.25)	4,694 (12.34)	6,002 (15.8)	7,022 (18.49)	7,951 (20.96)	8,690 (22.92)	9,231 (24.38)
Accommodation and food service activities	Male	Employee at origin point	49,474	49,003	48,855	48,736	48,593	48,443	48,296
		Changed to non-employee	5,123 (10.35)	7,679 (15.67)	8,967 (18.35)	10,137 (20.8)	10,840 (22.31)	11,631 (24.01)	12,140 (25.14)
	Female	Employee at origin point	53,047	52,306	52,322	52,455	52,444	52,361	52,338
		Changed to non-employee	7,457 (14.06)	11,390 (21.78)	13,945 (26.65)	15,727 (29.98)	16,854 (32.14)	17,942 (34.27)	18,648 (35.63)
Information and communication	Male	Employee at origin point	171,916	171,706	171,481	171,297	171,095	170,870	170,595
		Changed to non-employee	8,634 (5.02)	12,587 (7.33)	15,080 (8.79)	22,587 (13.19)	22,434 (13.11)	23,148 (13.55)	23,670 (13.87)
	Female	Employee at origin point	41,840	41,812	41,776	41,747	41,714	41,683	41,637
		Changed to non-employee	4,074 (9.74)	6,538 (15.64)	8,345 (19.98)	10,685 (25.59)	11,595 (27.8)	12,263 (29.42)	12,509 (30.04)
Financial and insurance activities	Male	Employee at origin point	207,740	207,480	207,227	206,953	206,678	206,409	206,042
		Changed to non-employee	6,162 (2.97)	10,686 (5.15)	14,275 (6.89)	19,865 (9.6)	25,373 (12.28)	28,597 (13.85)	32,944 (15.99)
	Female	Employee at origin point	103,280	103,230	103,190	103,138	103,098	103,029	102,958
		Changed to non-employee	4,585 (4.44)	7,379 (7.15)	9,686 (9.39)	13,127 (12.73)	15,325 (14.86)	17,248 (16.74)	19,008 (18.46)
Real estate activities	Male	Employee at origin point	164,395	163,994	163,619	163,287	162,934	162,544	162,095
		Changed to non-employee	8,419 (5.12)	13,147 (8.02)	15,319 (9.36)	17,618 (10.79)	19,489 (11.96)	21,275 (13.09)	23,319 (14.39)

	Female	Employee at origin point	61,725	61,656	61,584	61,513	61,434	61,340	61,247
		Changed to non-employee	5,524 (8.95)	8,818 (14.3)	10,643 (17.28)	12,125 (19.71)	13,343 (21.72)	14,450 (23.56)	15,353 (25.07)
Professional, scientific, and technical activities	Male	Employee at origin point	146,935	146,664	146,419	146,202	145,927	145,657	145,371
		Changed to non-employee	8,141 (5.54)	11,515 (7.85)	13,501 (9.22)	15,058 (10.3)	15,895 (10.89)	17,043 (11.7)	18,035 (12.41)
	Female	Employee at origin point	60,375	60,326	60,283	60,228	60,181	60,135	60,075
		Changed to non-employee	6,467 (10.71)	9,743 (16.15)	11,808 (19.59)	12,995 (21.58)	14,053 (23.35)	14,596 (24.27)	15,120 (25.17)
Business facilities management and business support services; rental and leasing activities	Male	Employee at origin point	163,342	162,778	162,396	162,034	161,621	161,218	160,775
		Changed to non-employee	11,670 (7.14)	16,834 (10.34)	19,563 (12.05)	22,220 (13.71)	24,695 (15.28)	27,058 (16.78)	28,904 (17.98)
	Female	Employee at origin point	102,102	101,962	101,868	101,771	101,657	101,547	101,438
		Changed to non-employee	11,967 (11.72)	17,760 (17.42)	21,154 (20.77)	24,014 (23.6)	26,093 (25.67)	27,891 (27.47)	29,614 (29.19)
Public administration and defense; compulsory social security	Male	Employee at origin point	16,515	16,470	16,416	16,370	16,319	16,275	16,226
		Changed to non-employee	712 (4.31)	1,224 (7.43)	1,502 (9.15)	1,810 (11.06)	2,102 (12.88)	2,360 (14.5)	2,769 (17.07)
	Female	Employee at origin point	12,169	12,155	12,138	12,129	12,117	12,105	12,091
		Changed to non-employee	1,005 (8.26)	1,536 (12.64)	1,749 (14.41)	2,012 (16.59)	2,211 (18.25)	2,429 (20.07)	2,570 (21.26)
Education	Male	Employee at origin point	33,966	33,597	33,346	33,102	32,927	32,770	32,638
		Changed to non-employee	3,144 (9.26)	4,783 (14.24)	5,681 (17.04)	6,254 (18.89)	6,667 (20.25)	7,069 (21.57)	7,495 (22.96)
	Female	Employee at origin point	159,247	158,997	158,816	158,640	158,500	158,346	158,177
		Changed to non-employee	15,624 (9.81)	21,626 (13.6)	26,170 (16.48)	29,671 (18.7)	31,932 (20.15)	34,156 (21.57)	36,120 (22.84)
	Male	Employee at origin point	129,387	129,207	129,043	128,854	128,626	128,373	128,101

Human health and social work activities		Changed to non-employee	5,289 (4.09)	7,488 (5.8)	8,799 (6.82)	9,774 (7.59)	10,772 (8.37)	11,597 (9.03)	12,454 (9.72)
	Female	Employee at origin point	247,963	247,772	247,598	247,423	247,214	246,998	246,759
		Changed to non-employee	24,535 (9.89)	35,343 (14.26)	41,656 (16.82)	46,579 (18.83)	50,196 (20.3)	52,687 (21.33)	55,125 (22.34)
Arts, sports and recreation related services	Male	Employee at origin point	41,711	41,619	41,536	41,444	41,360	41,271	41,183
		Changed to non-employee	2,688 (6.44)	4,130 (9.92)	4,831 (11.63)	5,542 (13.37)	6,135 (14.83)	6,674 (16.17)	7,211 (17.51)
	Female	Employee at origin point	21,746	21,724	21,712	21,695	21,672	21,657	21,629
		Changed to non-employee	2,355 (10.83)	3,639 (16.75)	4,338 (19.98)	4,973 (22.92)	5,527 (25.5)	5,825 (26.9)	6,081 (28.12)
Membership organizations, repair and other personal services	Male	Employee at origin point	182,261	181,679	181,259	180,941	180,560	180,171	179,744
		Changed to non-employee	11,196 (6.14)	16,839 (9.27)	19,798 (10.92)	22,896 (12.65)	25,150 (13.93)	27,764 (15.41)	29,921 (16.65)
	Female	Employee at origin point	91,624	91,506	91,436	91,355	91,281	91,190	91,071
		Changed to non-employee	9,696 (10.58)	14,714 (16.08)	17,866 (19.54)	20,274 (22.19)	22,111 (24.22)	23,785 (26.08)	24,678 (27.1)

Appendix 19. Sample R codes for the “HWSB_ES_Extrapolation” function and an example of application

```
HWSB_ES_Extrapolation = function(mean_age_male, std_dev_male, mean_age_female,
std_dev_female, prop_male, RR_prev_study, Year, Cohort, Follow_up) {
  # Load necessary libraries
  library(tidyverse)
  library(data.table)
  library(readxl)
  library(writexl)
  library(htmlTable)
  library(stats)

  # Import HWSB-ES data
  hwsb_es = fread('result/pre-calculated_HWSB_95CI.csv')

  # Define age group boundaries
  age_groups <- c(40, 49)

  # Calculate proportions for males
  cdf_male <- pnorm(age_groups, mean = mean_age_male, sd = std_dev_male)
  proportion_male_30_39 <- cdf_male[1] * prop_male
  proportion_male_40_49 <- (cdf_male[2] - cdf_male[1]) * prop_male
  proportion_male_50_59 <- (1 - cdf_male[2])* prop_male

  # Calculate proportions for females
  cdf_female <- pnorm(age_groups, mean = mean_age_female, sd = std_dev_female)
  proportion_female_30_39 <- cdf_female[1]* (1 - prop_male)
  proportion_female_40_49 <- (cdf_female[2] - cdf_female[1])* (1 - prop_male)
  proportion_female_50_59 <- (1 - cdf_female[2])* (1 - prop_male)
```

```
# Select length of landmark, cohort setting, and follow-up duration
hwsb_es_selected = hwsb_es %>%
  filter(length_of_landmark %in% Year, cohort %in% Cohort, follow_up %in% Follow_up)

# Demographic proportions from a previous study
prev_study_distribution <- data.frame(
  age_group = c('30s', '40s', '50s','30s', '40s', '50s'),
  sex = c("Male", "Male", "Male","Female", "Female", "Female"),
  proportion = c(proportion_male_30_39, proportion_male_40_49, proportion_male_50_59, proportion_female_30_39, proportion_female_40_49, proportion_female_50_59))

# Merge datasets to align the proportions with the HWSB-ES data
merged_data <- merge(hwsb_es_selected, prev_study_distribution, by = c("age_group",'sex'))

# Calculate weighted HWSB-ES
merged_data <- merged_data %>%
  mutate(
    weighted_HWSB_ES = HWSB_ES * proportion,
    weighted_Var_HWSB_ES = (SE_HWSB_ES^2) * (proportion^2))

# Sum the weighted HWSB-ES and Variance across all groups
total_weighted_HWSB_ES <- sum(merged_data$weighted_HWSB_ES)
total_weighted_Var_HWSB_ES <- sum(merged_data$weighted_Var_HWSB_ES)

# Calculate SE for the weighted HWSB-ES
```

```
  SE_total_weighted_HWSB_ES <- sqrt(total_weighted_Var_HWSB_ES)

  # Adjust RR or SMR using the weighted HWSB-ES
  adjusted_RR <- RR_prev_study * 1 / (1 - total_weighted_HWSB_ES)

  # Calculate 95% CI for the adjusted RR
  CI_lower <- adjusted_RR - 1.96 * SE_total_weighted_HWSB_ES
  CI_upper <- adjusted_RR + 1.96 * SE_total_weighted_HWSB_ES

  # Results
  cat("Adjusted RR: ", adjusted_RR, "\n")
  cat("95% CI: [", CI_lower, ", ", CI_upper, "]\n")
}

# Example usage of the function
HWSB_ES_Extrapolation(
  mean_age_male = 46.3,
  std_dev_male = 5.0,
  mean_age_female = 43.2,
  std_dev_female = 5.5,
  prop_male = 0.55,
  RR_prev_study = 1.07,
  Year = 5,
  Cohort = 'fixed',
  Follow_up = 'long'
)
```

ABSTRACT (IN KOREAN)

국내 중장년층의 근로 형태 변화에 따른 건강 근로자 생존 편향 탐색

<지도교수 윤진하>

연세대학교 대학원 의학과

윤 병 윤

배경: 건강 근로자 편향은 근로자와 일반 인구를 비교할 때 근로자와 관련된 다양한 고용 관련 요소로 인한 사망률 및 질병 발생률의 감소로 정의한다. 건강 근로자 생존자 편향은 건강 근로자 편향의 구성 요소 중 하나로, 지속적으로 일하는 근로자가 퇴사하는 근로자에 비해 일반적으로 더 좋은 건강 상태를 유지하기 때문에 생긴다. 건강 근로자 생존자 편향은 건강 근로자 고용 편향과 비교하였을 때 조절하기가 더 어려운데, 이는 생존자 편향이 시간에 따른 선택적 편향이자 교란 편향이기 때문이다. 건강 근로자 생존자 편향은 직업 연구 내에서 노출-건강 영향을 일반적으로 약화시키기 때문에 이를 최소화하는 것은 매우 중요하다. 그러나 건강 근로자 생존자 편향의 정도를 조사하거나 이를 보정하는 연구가 부족한 실정이다. 이에 본 연구는 1) 사회경제적 요인을 고려한 근로 형태에 따른 건강 영향을 탐색하고 2) 근로 형태의 변화로 인한 건강한 근로자 생존자 편향을 대한민국 전체 인구에서 추정하는 것을 목표로 한다.

방법: 국민건강보험공단 데이터베이스에서 2008년부터 2010년까지 직장가입자, 지역가입자, 피부양자 중 동일한 유형의 보험 자격을 계속 유지한 30세에서 59세 사이의 인구 집단을 대상으로 하였으며, 시작점은 2011년 1월으로 2022년 12월까지 추적 조사되었다. 주요 종속변수는 전체 사망으로 정의되었다. 근로 형태에 따른 전체 사망 위험도(HR) 및 95% 신뢰 구간(CI)은 다변량 콕스 비례 위험 모형을 사용하여 추정하였다. 근로 형태의 변화에 따른 건강한 근로자 생존자 편향(HWSB−ES)의 크기는 랜드마크의 시작점인 초기 시점과 종점인 현재 시점의 감쇠된 상대 위험도의 정도로 추정하였다. 상대 위험도는 연령 표준화 사망률 비로 측정하였으며 연령 표준화 사망률 비는 각 시점에서 근로자 집단과 전체 인구 집단의 연령 표준화된 사망률을

통해 산출하였다. 초기 시점은 코호트의 시작 연도(2011)로 설정되었으며 현재 시점은 랜드마크의 길이에 따라 원점으로부터 1~10년까지 변화되도록 설정하였다. 직장에서 퇴사한 근로자만 고려하는 고정 코호트 방식과 근로자 중 퇴사한 경우와 비근로자가 근로자로 입사한 경우를 모두 고려하는 다이나믹 코호트 설정에서 단기 및 장기 HWSB–ES가 각각 추정되었다. 단기 HWSB–ES는 현재 시점으로부터 1년 동안의 사망률을 기반으로 추정되었으며, 장기 HWSB–ES는 마지막 추적일인 2022년 12월까지의 사망률을 기반으로 추정되었다.

결과: 중위 12년 추적 기간 동안 18,192,989명의 참가자(중위 연령 44세; 남성 49.05%) 중 30대, 40대, 50대 연령 그룹에서 각각 64,177명(1.07%), 153,843명(2.30%), 253,736명(4.61%)이 사망하였다. 남성의 경우, 무직이 근로자에 비해 사망 위험도가 높았으며 이는 40대 남성에서 보정된 HR (95% CI)가 4.03 (3.94–4.10)로 가장 높은 사망 위험도로 관찰되었다. 여성의 경우, 자영업자가 근로자에 비해 사망 위험도가 높았으며, 30대 여성에서 보정된 HR (95% CI)가 3.34 (3.19–3.49)로 가장 높은 사망 위험도가 관찰되었다. HWSB–ES의 크기 추정 분석에 따르면 고정 코호트 및 다이나믹 코호트 방식에서 대한민국 전체 인구 중 5년 랜드마크 기준 연령 표준화된 단기 HWSB–ES 추정치는 각각 0.325194와 0.280361이었으며 장기 HWSB–ES 경우에는 각각 0.155346과 0.082751이었다. HWSB–ES는 여성, 더 높은 연령 및 긴 랜드마크 기간인 경우에서 더 높은 값을 가지는 경향을 보였다. 또한, HWSB–ES값은 고정 코호트 내에서 단기 추정치일 때 높은 값을 가졌다. 상기 추정된 HWSB–ES 값을 통해 기존 문헌, 특히 특정 근로자들의 전체 인구 집단 대비 연령표준화 사망률비를 추정한 문헌들에 대해 과소 추정된 값을 보정하여 새로운 의미를 도출해 낼 수 있다. 또한, 본 연구에서는 기존 연구의 남녀의 비율 및 성별에 따른 나이의 분포와 랜드마크 기간, 코호트 유형, 추적관찰기간을 반영하여 계산된 건강 근로자 생존 효과를 기존 연구의 상대 위험도에 외삽하였을 때 실제 기대되는 위험도가 어느 정도인지를 탐색할 수 있는 함수를 개발하였다.

결론: 이 연구는 대한민국 중년 인구에서 근로 형태와 전체 사망률 간의 중요한 연관성을 강조하였다. 또한, 이 연구는 대한민국 전체 인구를 대상으로 HWSB–ES의 크기를 추정하였으며 이를 기존 연구의 결과에 활용하는 법을

개발하였다. 대한민국에서 직업적인 노출과 건강 영향에 관한 연구를 수행할 때, 이 연구에서 측정된 HWSB-ES의 크기를 적용하여 위험을 과소평가하지 않도록 할 수 있다.

핵심되는 말: 근로 형태의 변화, 건강 근로자 생존 편향, 사회경제적요인, 랜드마크 분석, 건강검진 수검, 전체 사망

PUBLICATION LIST

1. Yun B, Oh J, Choi J, Rozek LS, Park H, Sim J, Kim Y, Lee J, Yoon JH. Socioeconomic Disparities in the Association Between All-Cause Mortality and Health Check-Up Participation Among Healthy Middle-Aged Workers: A Nationwide Study. J Korean Med Sci. 2023 Nov;38:e384. https://doi.org/10.3346/jkms.2023.38.e384