Received: 4 May 2019 | Revised: 28 August 2019 | Accepted: 19 October 2019

DOI: 10.1002/hec.3986

SPECIAL ISSUE PAPER

Health Economics WILEY

Future projection of the health and functional status of older people in Japan: A multistate transition microsimulation model with repeated cross-sectional data

Megumi Kasajima[1] | **Hideki Hashimoto**[2] | **Sze-Chuan Suen**[3] | **Brian Chen**[4] | **Hawre Jalal**[5] | **Karen Eggleston**[6] | **Jay Bhattacharya**[7]

[1]Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Japan

[2]Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Japan

[3]Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California

[4]Department of Health Services Policy and Management, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina

[5]Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[6]FSI Shorenstein Asia Pacific Research Center, Stanford University, Stanford, California

[7]Center for Primary Care and Outcomes Research, Stanford School of Medicine, Stanford, California

Correspondence
Megumi Kasajima, Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Tokyo, Japan.
Email: kasajimamegumi@gmail.com

Funding information
Cabinet of Office, Japan, Grant/Award Number: 2015-PM16-02-01; Ministry of Health, Labour and Welfare in Japan, Grant/Award Number: H26-Chikyukibo-ippan-001; National Institutes on Health, Grant/Award Number: P30 AG17253

Abstract

Accurate future projections of population health are imperative to plan for the future healthcare needs of a rapidly aging population. Multistate-transition microsimulation models, such as the U.S. Future Elderly Model, address this need but require high-quality panel data for calibration. We develop an alternative method that relaxes this data requirement, using repeated cross-sectional representative surveys to estimate multistate-transition contingency tables applied to Japan's population. We calculate the birth cohort sex-specific prevalence of comorbidities using five waves of the governmental health surveys. Combining estimated comorbidity prevalence with death record information, we determine the transition probabilities of health statuses. We then construct a virtual Japanese population aged 60 and older as of 2013 and perform a microsimulation to project disease distributions to 2046. Our estimates replicate governmental projections of population pyramids and match the actual prevalence trends of comorbidities and the disease incidence rates reported in epidemiological studies in the past decade. Our future projections of cardiovascular diseases indicate lower prevalence than expected from static models, reflecting recent declining trends in disease incidence and fatality.

KEYWORDS

demographic trends, economics of the elderly, forecasting models, health and economic development, simulation methods

1 | INTRODUCTION

Rapid population aging is a risk to social sustainability in countries like Japan that have low fertility rates and high life expectancy. Japan's overall population is decreasing, and the proportion of the population aged ≥65 years has reached 27.3% (Cabinet Office, 2017). Future years will bring increasing demand for old-age pensions, medical services, and long-term care, while revenues to support such services will decrease. Moreover, the distribution of economic, health, and social resources among older people exhibits considerable disparities (Ichimura, Shimizutani, & Hashimoto, 2009). Accurate estimates of future demand for health and social services in heterogeneous older populations are imperative to design sustainable healthcare and social security systems.

Currently available projections of future population and health status assume a static and average status for comorbidity prevalence and mortality by age and sex strata (National Institute of Population and Social Security Research, 2017a); however, such projections fail to incorporate the diverse and dynamic associations between health, economic, and social conditions among older people. To address this gap, a microsimulation model, the U.S. Future Elderly Model, was developed using comprehensive information available in large panel datasets such as the Health and Retirement Study (Goldman et al., 2015; Goldman, Shekelle, Bhattacharya, Hurd, & Joyce, 2004). Chen et al. (2016) have constructed a similar projection model using the existing available panel dataset for Japan, but the fact that the panel dataset they use does not sample the oldest-old (above age 75) limits their findings. The key input that panel data provide for these models is a set of health and mortality transition probabilities representative of the population at large (Chen et al., 2016). Panel datasets are preferred for modeling highly complex processes in human aging (Saksena & Maldonado, 2017) because they provide data on the dynamics of individual transitions. Longitudinal datasets enable us to disentangle age, cohort, and period effects and help control for unobserved heterogeneity (Roßmann & Gummer, 2016).

Although microsimulation models based on panel data have been useful in modeling dynamic population changes in health, the limited availability of panel data for estimating transition probabilities between health states often precludes developing such models. By contrast, repeated cross-sectional survey data with a consistent sampling frame over time (e.g., nationally representative surveys of health conditions) are more readily available in many countries ,including Japan. In this study, we propose an alternative approach to estimate multistate transition contingency tables using repeated cross-sectional survey data for microsimulations.

We face at least two challenges to inferring multistate transition probabilities from cross-sectional data. First, we must estimate disease incidence rates despite not observing individuals over time; instead, we base our estimate on changes in disease prevalence in the population. Second, we must account for the fact that coexisting comorbid statuses are not independent events. We propose in this study a novel approach using multistate transition contingency tables to overcome these challenges. Based on the model, we present our projection of population health among elderly Japanese forward to 2046 and discuss its implications for health policy in the super-aged society of Japan.

2 | METHODS

2.1 | Data sources

Our model requires repeated cross-sectional data of a closed cohort of the target population (a) collected over at least two waves, which allows us to estimate health state transition probabilities, and (b) a consistent sampling frame across survey waves. To estimate disease prevalence between survey waves, we applied a local polynomial smoothing function in which we assumed evenly spaced knots over time.

For this purpose, we analyzed microdata derived from the Comprehensive Survey of Living Conditions (CSLC), a large cross-sectional nationally representative survey conducted every 3 years by the Japan Ministry of Health, Labour, and Welfare. Approximately 600,000 individuals in 295,000 households were sampled by two-stage cluster sampling in each wave. We calculated time trends in disease prevalence and functional status specific to age-sex-specific cohorts to create a synthetic panel from 2001 to 2013. We also incorporated cause-specific mortality data from vital statistics data for 2000 through 2014 that included death records of approximately 1.2 million people per year (Ministry of Health, Labour and Welfare). We obtained the monthly age-sex-disease-specific probability of death to reflect changing trends in cause-specific mortality, and we standardized the cohort population using data from the 2010 population census. (For more details about data selection and adjustment, see the supporting information.)

2.2 | Health and functional status variables

The 14 health status variables consisted of self-reported morbidity for 11 chronic diseases, subjective health status, and two measures of limitations in activities of daily living. Morbidity diagnoses included diabetes, coronary heart disease, stroke, hypertension, hyperlipidemia, cancer, all respiratory diseases, joint disorders, eye diseases, kidney disorders, and other. The "other" category included circulatory diseases other than coronary heart disease (e.g., heart failure), gastric diseases, and noncancer prostatic conditions (e.g., hyperplasia). We also included poor subjective health (measured as five levels of self-reported health status), which is known to be highly related to mortality prognosis (DeSalvo, Bloser, Reynolds, He, & Muntner, 2006; Idler & Benyamini, 1997; Idler, Russell, & Davis, 2000). We indicated "1" if the respondent reported "poor" or "very poor" health. Finally, dysfunctions in activities of daily living were defined as limitations in at least one of the following basic activities: independently getting out of bed, bathing, dressing, and eating. We regarded this as a condition requiring nursing care. In contrast, "mobility dysfunction" was defined as requiring personal care/assistance when leaving the home.

2.3 | Strategy for determining the transition probability parameters

Based on birth cohort- and-sex-specific health and functional status variables obtained from waves of repeated cross-sectional surveys, we determined transition probabilities over time using contingency tables as described shortly.

Brunet and Struchiner proposed a nonparametric continuous-time method to estimate incidence rate based on information for disease prevalence and differential mortality by the disease when only repeated cross-sectional survey data were available (Brunet & Struchiner, 1999). Goldman et al. (2004) developed a discrete-time version of these methods in the context of a simulation model for projecting population aging. Our application of these methods assumes:

1. Disease conditions are absorbing states. All chronic conditions included in the model were assumed to be absorbing states (i.e., there is no recovery from any chronic condition). This assumption is justified in our case by the fact that for many chronic conditions (such as diabetes), there is no cure.
2. The mortality rate of a disease condition is always at least as large as the base rate of the age-sex-specific mortality without comorbid conditions. With these assumptions, Brunet and Struchiner determined that the disease incidence rate (or transition probability from not having the condition to having the condition) is a function of the change in disease prevalence between periods and the difference between the disease-specific mortality rate and the base rate of mortality.
3. Related to the above assumption, we limited the population at risk (or candidate subpopulation) for a disease-specific death to those who had that disease. In other words, we assumed that one could not die from heart disease if one did not have heart disease in the previous period. To account for deaths from diseases that are not reported in the data or categorized in our model, we included a base rate of age-sex-specific mortality from other causes.
4. In addition, following previous models (Goldman et al., 2015), we further assumed Granger causality (Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003; Michaud & Van Soest, 2008; Stowasser, Heiss, McFadden, & Winter, 2011). That is, we predicted future time series (such as health conditions) using prior values of a time series. We adopted this assumption of Granger causality because our aim was to predict future health states not to identify causal pathways.
5. Finally, because of data limitations, we assumed that the total death probability was additive across concurrent conditions; for example, the total probability of death for an individual with stroke and heart disease was the sum of the mortality probability from stroke and the mortality probability from heart disease. If information on multiple causes of death were available, this fifth assumption would have been unnecessary.

2.3.1 | Estimating the prevalence of coexisting health states

Previous studies used regression-based prediction models of health status transition with panel data to model changing trends in correlated health states over waves (Chen et al., 2016; Goldman et al., 2015). In their models, they regressed each comorbidity status on a set of time-lagged comorbid conditions judged to predict future acquisition of health conditions, while controlling for age, sex, and other demographic or socio-behavioral characteristics such as race, education, body mass index, and smoking status. They bring the covariance structure between coexisting morbidity

conditions into the model by including past comorbid condition variables as regressors, which is possible because they rely on panel data. Instead, we explicitly account for the joint distribution of coexisting morbidity prevalence. The existing literature related to our approach treats each health state as mutually exclusive (Brunet & Struchiner, 1999; Goldman et al., 2004; Hallett et al., 2008). To account for the joint distribution of disease statuses, we estimated two-by-two contingency tables for each pair of disease statues by age, sex, and time period.

Because CSLC data had a 3-year interval between waves, we smoothed the required numbers for each cell in the 2 × 2 contingency tables using the local polynomial method to obtain monthly prevalence. We chose 1 month as the unit of time interval for incidence estimation, because the number of incident cases and deaths should be relatively small during such a short period. Using this assumption, we determined incidence based on the prevalence and death values we obtained from the data sources. (For a depiction of our estimation process, see Appendix Figure 2S1).

We chose this simple 2 × 2 table method because the frequency distributions across two health states ensured non-zero positive numbers in each cell for stable estimation. Also, two-dimensional data captured a large portion of the total variance on 14-dimensional joint distributions (i.e., >80%), which justified the use of two-dimensional tables for model simplicity.

2.3.2 | Mortality estimation

Vital statistics data listing multiple causes of death linked with past comorbidity history are the ideal data to estimate cause-specific case fatality when individuals have multiple comorbidities. Unfortunately, the official Japanese vital statistics data contain only a single cause of death. Because of this limitation, we assumed additive probability of mortality (e.g., those who had heart disease and stroke should have a risk of mortality equal to the risk of mortality from heart disease plus that from stroke). We calculated case-fatality rates for the 11 chronic diseases we listed and the additive probability of mortality from corresponding comorbidities. We attributed additional mortality exits from the poor subjective health cell to mental health conditions. Because of data limitations, we assumed that impaired mobility and dysfunctions in activities of daily living do not independently increase the mortality risk. Therefore, regardless of limitations in activities of daily living and mobility, the probability of mortality depends only on subjective health and the 11 diseases. Appendix Table 2 lists the definitions of cause-specific death based upon the International Classification of Cause of Death version 10 (ICD-10) systemS2. We determined the age-sex-specific base mortality rate so that the total mortality exits in the model agreed with the observed natural decreasing trend in the population of a given birth cohort. We adopted local polynomial smoothing with four age-year width bands around the kernel to obtain age-specific mortality curves.

2.3.3 | Incidence estimation

Under the assumption that during a 1-month interval, prevalence is balanced with incidence (entry into the cohort) and death (exit from the cohort), we evaluated new entry and exit from each cell in the sex-birth cohort-specific 2 × 2 contingency tables to derive state-specific incidence.

For two arbitrary diseases i and j, the 2 × 2 table at time t contains the initial prevalence numbers at the beginning of the month for four comorbidity patterns $(d_i, d_j) = (0,0), (1,0), (0,1), (1,1)$ for each birth cohort (c) and sex (s) where d_i and d_j are diagnostic statuses of diseases. We denoted the population size at the beginning of time t in each cell as: $\mathrm{pop}^{(i,j)}_{c,s,t}(d_i, d_j)$.

Then, between time t and $t + 1$, the cohort population decreases by the number of the deceased population. The population with condition $(d_i, d_j) = (0,0)$ decreases by the base mortality rate, $\alpha^{(i,j)}_{\mathrm{base}(c,s,t)}$, depending on individuals' age and sex. The base mortality rate is determined by the all-cause mortality rate and the case fatality rates, $\alpha_{i(c,s,t)}$ and $\alpha_{j(c,s,t)}$, respectively. The remaining three cells have additive mortality risks attributable to diseases describing individuals' health conditions at time t, and the population in the cells decreases by the corresponding mortality rates.

We write the number of survivors at the end of the month t in each cell for that closed cohort population, $\mathrm{surv}^{(i,j)}_{c,s,t}(d_i, d_j)$ as

$$\mathrm{surv}^{(i,j)}_{c,s,t}(0,0) = \mathrm{pop}^{(i,j)}_{c,s,t}(0,0) \times \left(1 - \alpha^{(i,j)}_{\mathrm{base}(c,s,t)}\right),$$

$$\mathrm{surv}^{(i,j)}_{c,s,t}(1,0) = \mathrm{pop}^{(i,j)}_{c,s,t}(1,0) \times \left(1 - \alpha_{i(c,s,t)} - \alpha^{(i,j)}_{\mathrm{base}(c,s,t)}\right),$$

$$\mathrm{surv}^{(i,j)}_{c,s,t}(0,1) = \mathrm{pop}^{(i,j)}_{c,s,t}(0,1) \times \left(1 - \alpha_{j(c,s,t)} - \alpha^{(i,j)}_{\mathrm{base}(c,s,t)}\right),$$

$$\mathrm{surv}^{(i,j)}_{c,s,t}(1,1) = \mathrm{pop}^{(i,j)}_{c,s,t}(1,1) \times \left(1 - \alpha_{i(c,s,t)} - \alpha_{j(c,s,t)} - \alpha^{(i,j)}_{\mathrm{base}(c,s,t)}\right).$$

Next, we compared the numbers of survivors at the end of the month (t) in the four cells with the estimated prevalence numbers in the corresponding cells at the beginning of the subsequent month $t + 1$. We attributed differences to changes in comorbidity prevalences because of disease incidence during the month. For the population with condition $(d_i, d_j) = (0,0)$, there are two possible status changes: to develop disease i from the (0,0) condition and to develop disease j from the (0,0) condition. Thus, for the incidence rates of diseases i and j from the precondition *(0,0),* the following equation holds

$$\mathrm{incidence}^{(i,j)}_{i(c,s,t)}(0,0) + \mathrm{incidence}^{(i,j)}_{j(c,s,t)}(0,0) = \frac{\mathrm{surv}^{(i,j)}_{c,s,t}(0,0) - \mathrm{pop}^{(i,j)}_{c,s,t+1}(0,0)}{\mathrm{pop}^{(i,j)}_{c,s,t}(0,0)} \quad ((1))$$

However, this equation cannot be uniquely solved for the incidence rates because of a lack of constraint conditions. Therefore, we solved the equilibrium of the solutions using a relevant set of multiple 2 × 2 tables.

$$\mathrm{incidence}^{(i,j)}_{i(c,s,t)}(0,0) - \mathrm{incidence}^{(i,j)}_{j(c,s,t)}(0,0)$$

$$\approx \frac{1}{12}\left\{ \sum_{\substack{k \neq i \\ k = 1,\ldots,14}} \left(\mathrm{incidence}^{(i,k)}_{i(c,s,t)}(0,0) + \mathrm{incidence}^{(i,k)}_{k(c,s,t)}(0,0)\right) - \sum_{\substack{k \neq j \\ k = 1,\ldots,14}} \left(\mathrm{incidence}^{(j,k)}_{j(c,s,t)}(0,0) + \mathrm{incidence}^{(j,k)}_{k(c,s,t)}(0,0)\right) \right\}. \quad ((2))$$

Equations (1) and (2) provide the incidence rates of diseases i and j from the (0,0) condition. Similarly, in the 2 × 2 table, developments of diseases affect population sizes in adjoining cells. The incidence of disease i from the (0,1) condition reduces $\mathrm{pop}^{(i,j)}_{c,s,t}(0,1)$, and the incidence of disease j from the (0,0) condition increases $\mathrm{pop}^{(i,j)}_{c,s,t}(0,1)$. This relationship leads to the incidence rate of disease i from the (0,1) condition.

$$\mathrm{incidence}^{(i,j)}_{i(c,s,t)}(0,1) = \mathrm{incidence}^{(i,j)}_{j(c,s,t)}(0,0) \times \frac{\mathrm{pop}^{(i,j)}_{c,s,t}(0,0)}{\mathrm{pop}^{(i,j)}_{c,s,t}(0,1)} - \frac{\mathrm{surv}^{(i,j)}_{c,s,t}(0,1) - \mathrm{pop}^{(i,j)}_{c,s,t+1}(0,1)}{\mathrm{pop}^{(i,j)}_{c,s,t}(0,1)}.$$

In the same way, the incidence rate of disease j from the (1,0) condition is obtained by

$$\mathrm{incidence}^{(i,j)}_{j(c,s,t)}(1,0) = \mathrm{incidence}^{(i,j)}_{i(c,s,t)}(0,0) \times \frac{\mathrm{pop}^{(i,j)}_{c,s,t}(0,0)}{\mathrm{pop}^{(i,j)}_{c,s,t}(1,0)} - \frac{\mathrm{surv}^{(i,j)}_{c,s,t}(1,0) - \mathrm{pop}^{(i,j)}_{c,s,t+1}(1,0)}{\mathrm{pop}^{(i,j)}_{c,s,t}(1,0)}.$$

From the estimated monthly incidence rates using 91 2 × 2 tables above, we calculated the conditional incidence probabilities under the condition (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, and d_{14}). We converted the incidence

rates in the 2 × 2 table form to conditional incidence probabilities of disease k in the 14-dimensional health status forms using the weighted average:

$$\frac{\sum_{\substack{l=1,\ldots,14 \\ l\neq k}} \text{incidence}^{(k,l)}_{k(c,s,t)}(d_k,d_l)\times \text{pop}^{(k,l)}_{c,s,t}(d_k,d_l)}{\sum_{\substack{l=1,\ldots,14 \\ l\neq k}} \text{pop}^{(k,l)}_{c,s,t}(d_k,d_l).}$$

For example, to calculate the conditional incidence probability of diabetes ($k = 1$) under the condition (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14}) = (0,0,0,1,0,0,0,1,0,0,0,0,0,0), we used the weighted average of the incidence rates of diabetes in 13 2 × 2 tables as follows: $\text{incidence}^{(1,2)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,3)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,4)}_{1(c,s,t)}(0,1)$, $\text{incidence}^{(1,5)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,6)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,7)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,8)}_{1(c,s,t)}(0,1)$, $\text{incidence}^{(1,9)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,10)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,11)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,12)}_{1(c,s,t)}(0,0)$, $\text{incidence}^{(1,13)}_{1(c,s,t)}(0,0)$, and $\text{incidence}^{(1,14)}_{1(c,s,t)}(0,0)$.

Consequently, we obtained a total of 114,688 (= 14 health variables × (2^{14}/2) comorbidity patterns) monthly conditional incidence probabilities as solutions for each sex, birth cohort, and month. We translated the sex-birth cohort-specific conditional incidence probabilities into age-sex-specific conditional incidence probabilities with the following two steps. First, after translating the monthly rates into annual rates for the consecutive 13 years, we regressed the pooled incidence estimates for each of the 114,688 combinations of comorbidities on age, age squared, and birth-cohort dummy variables to incorporate cohort-specific fixed effects, for men and women separately. Second, we used the estimated values for age-sex-specific conditional incidence probabilities for the years 2001–2013 to compare existing data sources for validation purposes. We used the numbers for the years 2010, 2011, and 2012 for projection to account for the recent trend change.

2.4 | Simulation

With the estimated sex-age-comorbidity-specific incidence and case-fatality rates, we performed a microsimulation projecting future distributions of health states. Transitions between health states, including mortality exit, followed a first-order Markov process; specifically, we assumed that disease incidence and mortality at time $t + 1$ depended only on comorbidities and age at time t.

For the microsimulation, we simulated an older Japanese population (aged 60+) with health conditions probabilistically distributed according to 2013 CSLC comorbidity prevalence data (approximately 42 million observations). For each individual, we calculated the conditional incidence and mortality probability, using data for 2010–2012, and assumed that these probabilities were constant in the future. We used a 6-month cycle length in our Markov process and prepared transition probability parameters accordingly.

Every 3 years, we supplemented the simulation population with an incoming cohort of 60- to 62-year-olds, using the birth-cohort population and mortality rates as of 2011–2013 up to age 60 for the standardized years of age. We assumed that the population disease distributions were the same as those of 60-, 61-, and 62-year-olds in the middle of 2013.

2.5 | Corroboration of the simulation parameters

We tested the validity of our simulation parameters by forward corroboration and, external, and backward validation. In the forward corroboration, we projected the future total Japanese population to 2046 using the 2013 population as a baseline, then compared our projected population forecast with the governmental official projection to 2046 (National Institute of Population and Social Security Research, 2017a). For external validation, we compared our values for disease incidence with those reported in existing epidemiological cohort studies conducted in Japan for cardiovascular diseases (Kubo et al., 2003) and cancer (National Cancer Center, 2015). To determine age-sex-specific annual incidence rates for cardiovascular diseases and cancer, we pooled birth cohort-sex-disease-specific conditional incidence

probabilities for 2001 through 2013, and then calculated pooled means weighted by the frequency of the comorbid combinations. Finally, as backward validation, we performed a simulation with cohorts born between 1924 and 1953 based on CSLC 2001 prevalence data as a baseline population to predict the disease prevalence for 2013 and then compared the projected results with the actual prevalence trends of corresponding age-sex strata in CSLC 2013 survey data.

After validating our parameters, we projected the age-sex-specific prevalence of comorbid conditions to 2046, using the initial conditions as of 2013. As a reference for comparison, to demonstrate the importance of dynamic microsimulation, we calculated disease prevalence based on the static model by multiplying sex cohort-specific prevalence rates by the governmental official population estimates for 2022, 2034, and 2046.

3 | RESULTS

3.1 | Forward corroboration

Figure 1 shows the future projection of the age-sex-specific population for men on the left-hand side and for women on the right-hand side, by 3-year-interval birth cohorts for 2022, 2034, and 2046, using the 2013 population structure as a baseline. Population Pyramids in black are the governmental population projections published by the National Institute of Population and Social Security Research. Our projected population for 2022 was statistically equivalent to the existing governmental projection of the Japanese population structure (Dinno, 2017; Schuirmann, 1987; Table S1). However, we observed a small but significant difference between our projections and official projections of those aged ≥75 years in

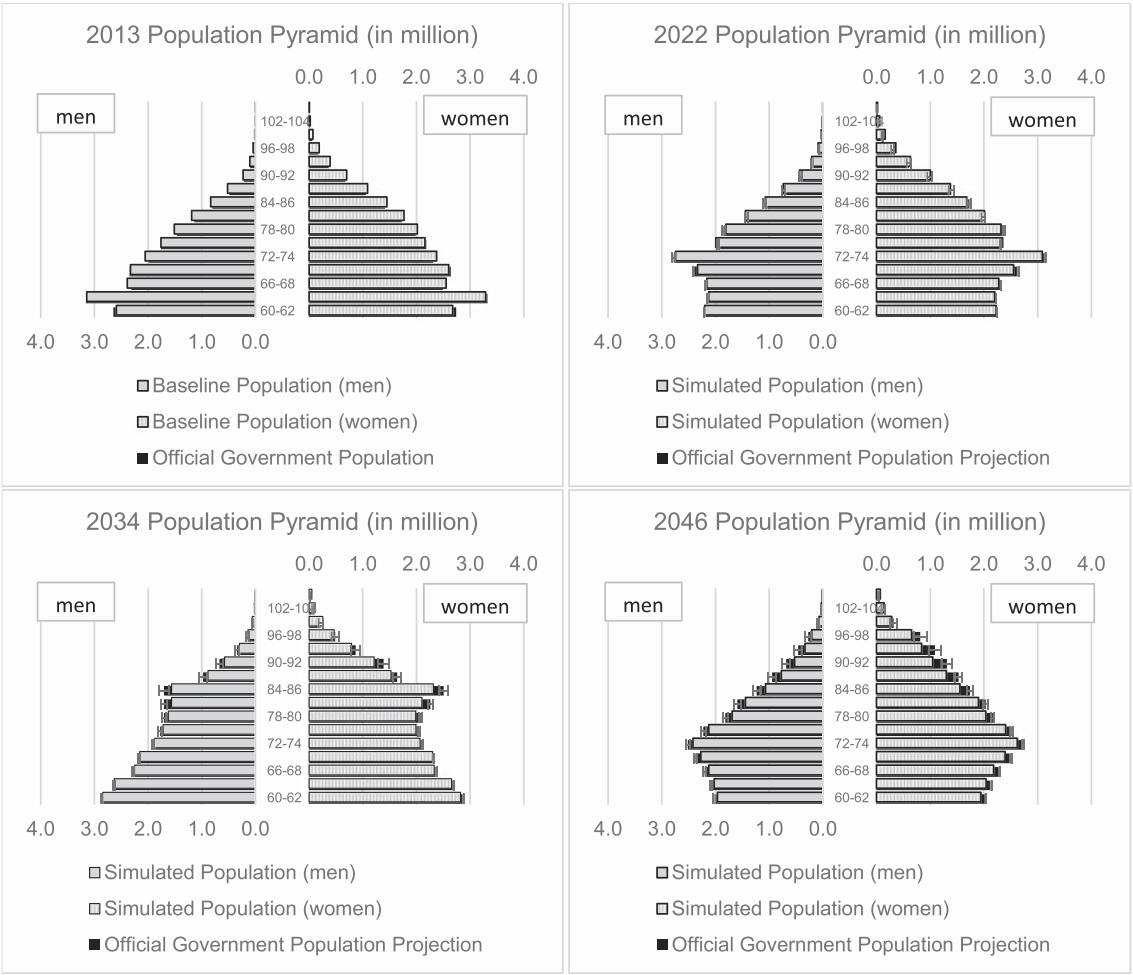


FIGURE 1 Simulated population pyramid (in gray); Japanese governmental official projections (in black); Japanese governmental projections with low-mortality and high-mortality assumptions (error bars)

2034 and 2046. Our long-term simulations predict slightly smaller cohorts of those aged ≥75 years, compared with the official projection.

3.2 | External validation

Figure 2 shows the validated annual incidence rate by age for coronary heart disease (Figure 2a), stroke (Figure 2b), and cancer (Figure 2c) in 2013, for men (circle) and women (diamond). The solid lines depict the weighted averages of the conditional incidence rates, while the shadowed area depicts the range between the 5th and 95th percentiles. The gray dots in Figures 2a and 2b plot the heart and stroke incidence rates from the Hisayama study of the 1990s (Kubo et al., 2003), which were higher in all age groups compared with our estimations. In Figure 2c, we compared our cancer incidence results with the numbers published in the Japanese national cancer registry. The cancer registry showed consistently higher incidence rates for both men and women, compared with our incidence rates. Japan's cancer registry included approximately 8.8% "Death Certificate Only" (DCO) cases (National Cancer Center, 2017). When we accounted for this number and allowed 8.8% death exits from the "no cancer comorbidity" cell, we obtained an incidence rate similar to the registry.

3.3 | Backward validation

As shown in Figure 3 and Table S2, the results of two-sample paired mean-equivalence *t* testing supported that our simulation model produced health status prevalence rates statistically equivalent to observed trends, except for a slight overestimation of the prevalence of hyperlipidemia and joint disorders in those aged ≥80 years. We also confirmed that our model replicated up to third-order joint distribution of comorbidities.

>Overall, women's base mortality was lower than for men. When we plotted the estimated case fatality for each year, we observed a gradual decline in case fatality rates for all diseases in our observation period, and the base mortality rates increased slightly at ≥80 years of age. Incidence trends varied by disease. The incidence of circulatory diseases and cancer increased with age, and men had higher risks compared with women. We observed declining incidence of these diseases over time.(for detailed estimation results, refer to Supplemental Figures S1 and S2)

3.4 | Future projection of disease prevalence

Figure 4 presents our estimates of future disease prevalence for selected key chronic diseases by age and sex. Our projections indicate a lower prevalence of coronary heart disease and stroke compared with what would be expected in a static model based on multiplying 2013 disease prevalence by the projected population published by the National Institute of Population Research (Figures 4a and 4b). In 2034, when the second wave of baby boomers (born 1971–1974) reach 60 years of age, and the proportion of the population over 65 years reaches 33% (Cabinet Office, 2017), our projection estimated stroke prevalence at 1.8 million compared with 2.2 million in a static model. At the same time, the prevalence of difficulties in activities of daily living and impaired mobility was lower in our projection compared with a static model (see supplemental figures).

Our projections of cancer and respiratory disease prevalence among Japan's future older population were similar to those based on a static model. In 2034, cancer prevalence will increase by 109,000, and respiratory prevalence will increase by 213,000 compared with 2013. However, our projection suggests that both incidence and case fatality will decline in the future, with offsetting effects, resulting in similar prevalence but longer survival for the affected population (Figures 4c and 4d).

4 | DISCUSSION

In this study, we proposed a multistate transition contingency table method for future projection of health conditions in older populations based on a microsimulation using repeated cross-sectional representative surveys in Japan. The statistical equivalence between existing epidemiological cohort data and our estimates of disease incidence and case-

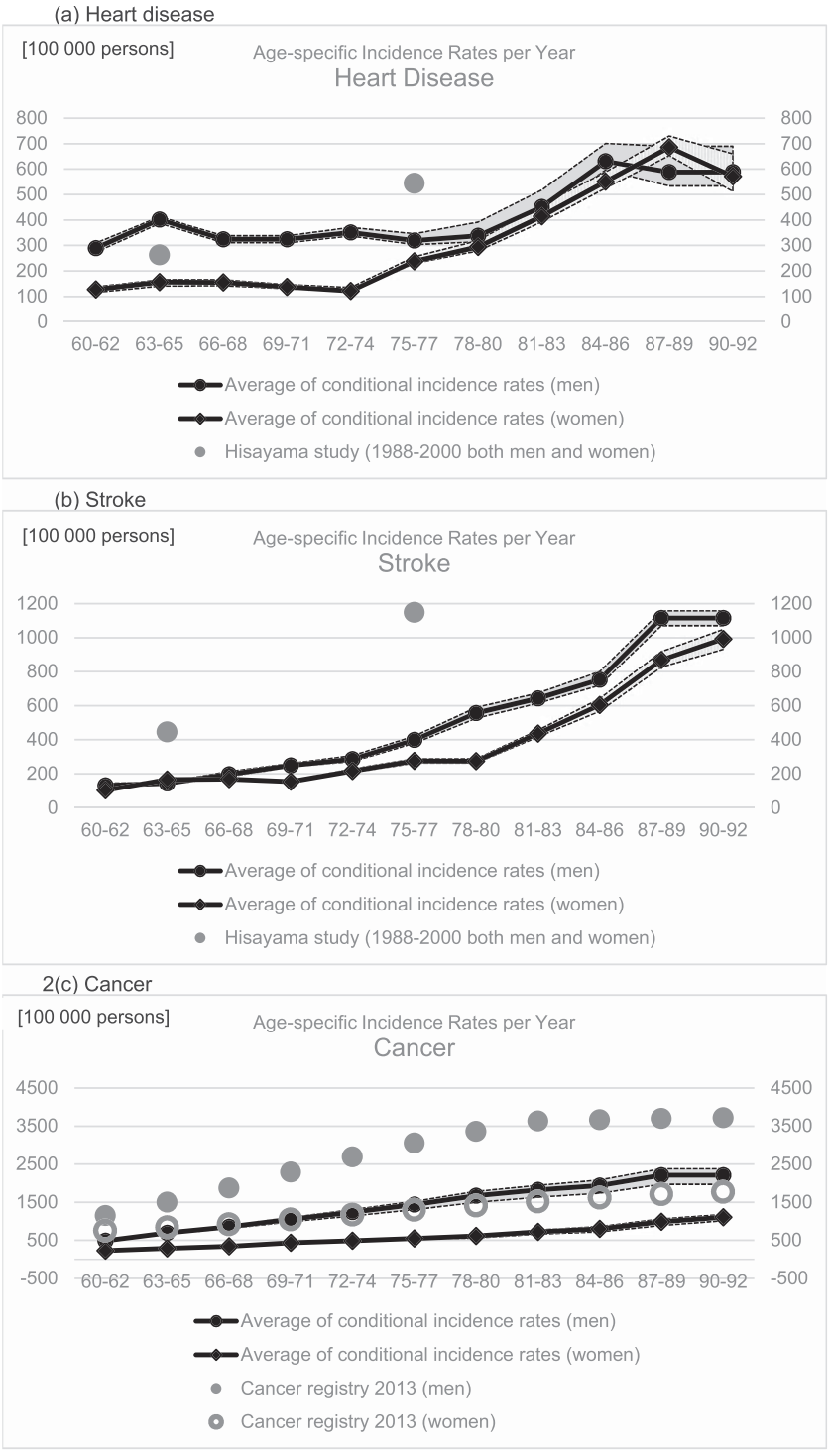


FIGURE 2 The range between the 5th and 95th percentiles is shadowed. The gray plots indicate epidemiological observations derived from the references (Kubo et al., 2003; National Cancer Center, 2017)

fatality rates support the validity of our model. Our projection results imply that traditional static models do not accurately forecast the prevalence of some comorbid conditions.

Our projections of future population pyramids were statistically equivalent to those published by governmental institutes until 2022, supporting the validity of our estimation of transition parameters. However, we found a small but significant gap between the governmental projection and our simulation results for those aged ≥75 years in the long-term simulation, which requires discussion. Governmental projections by the National Institute of Population and Social Security Research assume hypothetical elongation of potential life years in addition to common adjustment with Lee and Carter modeling to account for improved longevity of the Japanese older subpopulation (National Institute of Population and Social Security Research, 2017b). We did not rely on this hypothetical adjustment. Instead, our

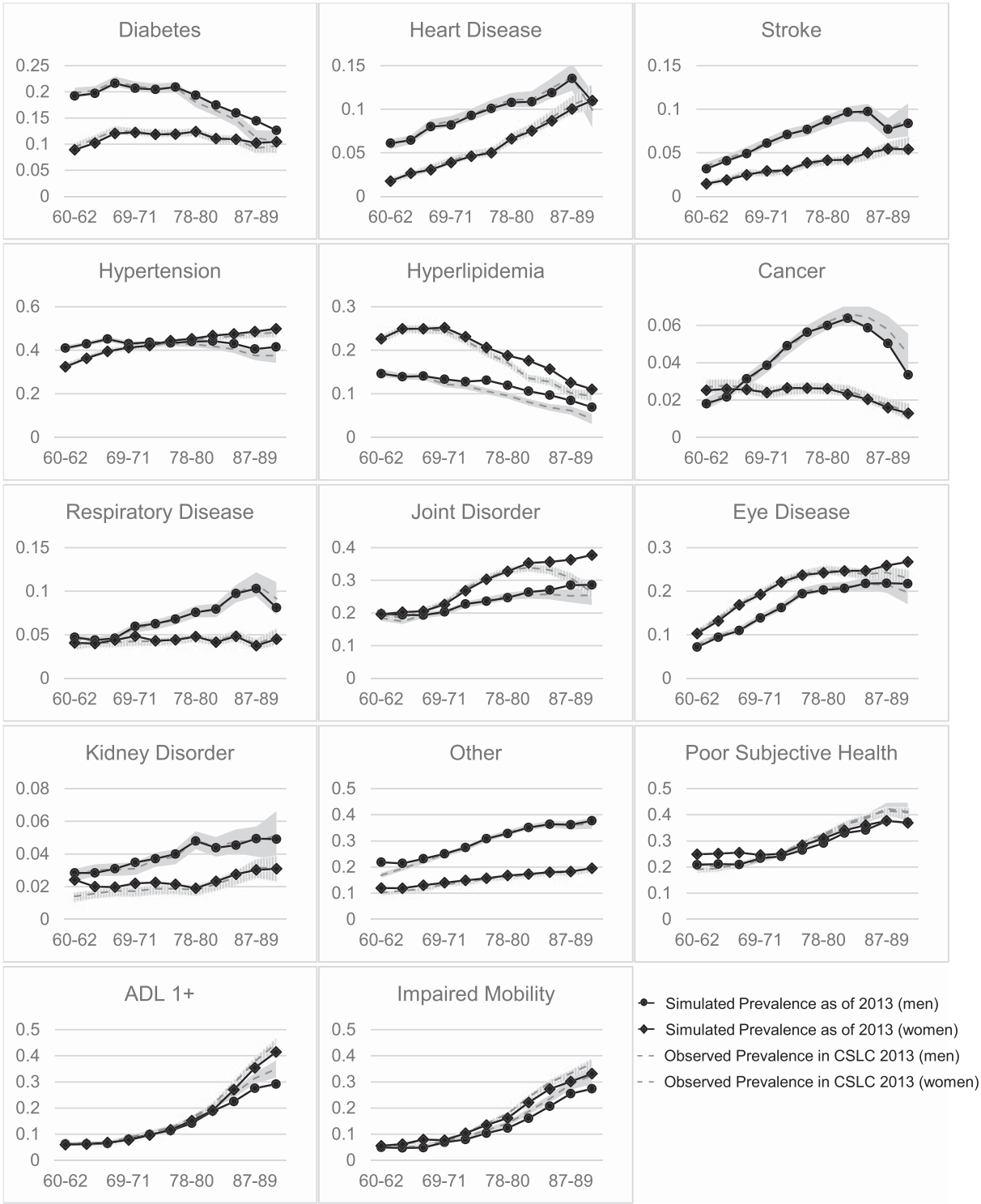


FIGURE 3 The solid line indicates estimations using observations from 2001 as a baseline. The dashed line indicates actually observed data in the Comprehensive Survey of Living Conditions 2013. The range of 95% confidence intervals is shadowed. ADL 1+, at least one condition among dysfunctions in activities of daily living

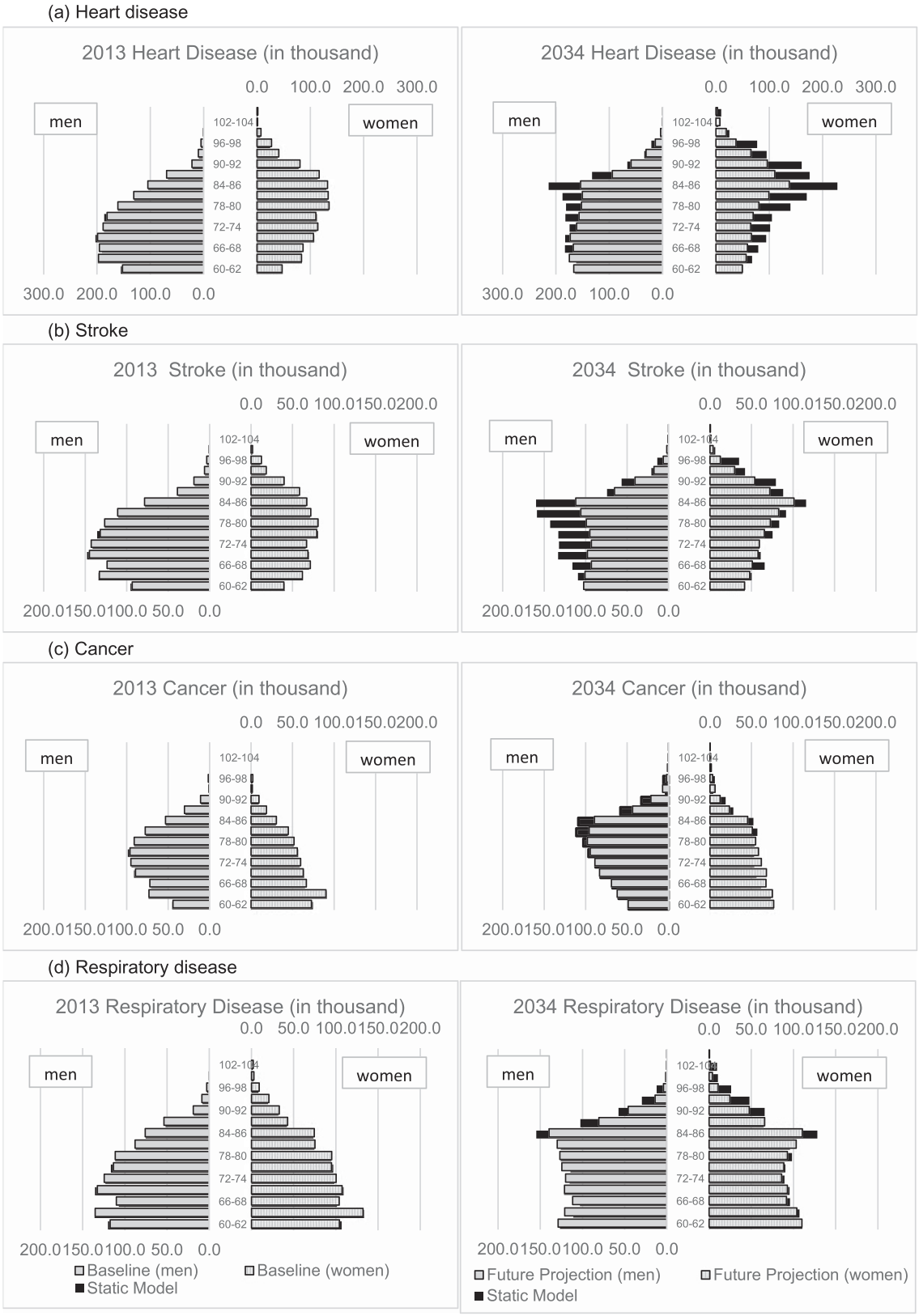


FIGURE 4 The bars in left-hand side describe the prevalence for men, and the bars in right-hand side describe the prevalence for women. The black bars represent estimates based on a static model

projection empirically estimated improved base mortality rates and lower incidence and case fatality rates for several conditions as a driver for improved population longevity. Our results were comparable with other projections using Bayesian modeling without such assumptions (Kontis et al., 2017).

Our estimates of the incidence rates of heart disease and stroke were lower than those reported in previous epidemiological studies. However, a simple comparison may not be plausible because we relied on self-reported health conditions in the CSLC data, and the Hisayama study defined disease diagnosis based on clinical examinations and autopsy findings, which should have detected more asymptomatic cases (Kubo et al., 2003). The Hisayama study was also based mainly on observations during the 1990s when the incidence of cardiovascular disease was higher than currently. In light of these considerations, we believe that our estimates reflect the current number of symptomatic stroke cases.

Cancer incidence published by the National Cancer Institute includes DCO cases to compensate for cancer death, but these are not recorded in the population-based cancer registry system. In Japan, reports of cancer cases are collected through clinics and hospitals and added to the cancer registry database (Japanese Association of Cancer Registries, 2010). When a cancer death is reported, the death record is matched with a cancer case in the cancer registry database. In cases of no matching registration, the cancer death is treated as a DCO case. In this study, because DCO cases were considered dead at the time of case identification, the related survivor time was treated as zero, which created an upward bias in the incidence estimates given the calculated prevalence (i.e., prevalence = incidence × average disease length; Brenner et al., 2016; Brenner & Holleczek, 2011). Indeed, when we performed ad hoc reestimation of cancer incidence allowing a DCO of 8.8% (or death by cancer from noncancer preconditions), the percentage reported in the 2013 National Cancer Registry, we confirmed that our estimate matched that of the registry. Therefore, we believe that the actual figure lies somewhere between the Cancer Registry number and our estimate.

In the backward validation, we slightly overestimated the prevalence of hyperlipidemia and joint disorders, probably because our absorbing assumption (no recovery) may not reflect the natural course of these conditions. Otherwise, our simulation accurately captured real-world health transitions of older Japanese from 2001 to 2013. Despite decreasing trends in incidence and mortality for most diseases, we project an increase in prevalence of multiple chronic conditions (and longer survival with disease) in Japan's near future because of the increasing absolute numbers of older people and improved survival of those with multiple conditions.

It is important to carefully discuss the differences between the results based on our dynamic model and those based on existing static models, which simply depict the average status of comorbidity prevalence by age and sex strata while assuming constant rates over time. Because new incoming cohorts had lower risks for stroke and coronary heart disease, our estimates of future prevalence of these conditions were located outside the 95% confidence interval range of the static model estimates. Although the estimated prevalence of respiratory conditions and cancer was similar between the dynamic and static models, our model suggested that this was because of lower incidence and better survival for the same conditions in the future older population, which may have different implications for future health policy decisions.

Our proposed multistate transition contingency table method with repeated cross-sectional data provides a complementary method with existing multistate transition models based on a panel-data structure. Repeated cross-sectional datasets are widely available, and our approach may be useful especially for those with limited availability of panel data. Our proposed approach also may be useful when an existing panel suffers from nonignorable attrition.

We acknowledge that our method requires further refinement. The model could be extended to include a wider list of comorbid conditions including cognitive limitations. In addition to age, sex, and health status, the model could also incorporate more detailed stratification by education level and/or other socioeconomic status indices to clarify social disparity in health in older people after retirement age. Our model also does not include risk factors such as body mass index, smoking habits, or exercise, which could improve the accuracy of the model. The model could also be expanded to include estimates of medical and long-term care costs.

Our simulation model may help assess the impact of policy and technological innovations on disease prevalence using counterfactual simulations. For instance, the model may help policy makers make informed decisions on policy reform by projecting the implications of different policy changes. Our model also has great potential to identify diverse and dynamic associations between health, the economy, and social conditions among older populations when we incorporate socioeconomic factors into future iterations (Shimizutani, Oshio, & Fujii, 2014; Stowasser, Heiss, McFadden, & Winter, 2011). Health affects, and is affected by, socioeconomic conditions (World Health Organization, 2008); changes in living conditions and available health technologies over time lead to changes in health, function, and likelihood of death (Ma et al., 2007; Tang & Kurashina, 1987; Wang, Weber, & Graham, 2015). Our model may help clarify the

implications for health and social disparities among older people with diverse sets of sociobehavioral, clinical, and economic risk factors.

Despite the promising benefits, we acknowledge that our proposed approach has limitations. First, we assumed an additive increase in mortality risk in those with multiple comorbidities. However, this assumption may overestimate or underestimate mortality risks for some combinations of diseases, depending on their synergetic or competing impacts on case fatality. Second, we postulated that all chronic conditions are absorbing in the Markov process; however, some symptomatic conditions, such as knee pain, included in joint disorders, may be reversible. Third, we fixed mortality and incidence parameters as of the most recent years of observation to minimize uncertainty and to obtain the most conservative result. For the same reason, we assumed that future incoming 60- to 62-year-old cohorts will be as healthy as those in the 2013 data. Instead, we could have incorporated future time trends for the health states of incoming cohorts, as in the original Future Elderly Model (Goldman et al., 2015). Fourth, our model does not consider future changes in technology that may improve morbidity and mortality. With sufficient knowledge of how certain changes in technology could change our estimated parameters, we could run counterfactual analyses of the impact of new technology on the distribution of morbidity and mortality in the future. Finally, we did not present confidence interval estimates in our simulation results. Bootstrapping confidence interval estimates is an option but one that requires overwhelming computing time.

Despite these challenges, developing a multistate transitional microsimulation model is a promising endeavor to open new horizons for policy evaluation and discussion regarding aging societies. The method also furthers our knowledge of the dynamic interactions between a diverse set of risk factors among heterogeneous older populations. Our proposed multistate transition contingency table method could be a useful tool to broaden the potential of microsimulations.

ACKNOWLEDGEMENTS

We express our gratitude to Michiaki Kubo for providing detailed heart disease and stroke incidence rates based on the Hisayama cohort study. We are grateful to Atsushi Goto for his expert opinion on diabetes prevalence in Japan, Futoshi Ishii for advising on the protocol used in the Human Mortality Database, and Tomohiro Matsuda for detailed information on the national cancer registry statistics. We thank Kent Byun and Ajay Anand for technical assistance in coding and Tetsuya Iwamoto for his assistance in the secondary analysis of Japanese governmental microdata. We express our deepest appreciation to Masaru Kitsuregawa and Kazuo Goda who provided the big data platform we used to conduct our simulation study.

CONFLICT OF INTEREST

We declare that the authors have no conflicts of interest.

DETAILS OF FUNDING SOURCES THAT SUPPORTED THE WORK

We are grateful for support from the National Institutes on Health and, in particular, from the National Institute on Aging (P30 AG17253) for the conduct of the study. This study is also supported by the Ministry of Health, Labour and Welfare in Japan (H26-Chikyukibo-ippan-001; https://www.mhlw.go.jp/english/index.html), the Ministry of Education, Culture, Sports, and Technology (Grant in Aid for Scientific Research (A) No.18H04070), and Cabinet of Office, Japan (ImPACT program 2015-PM16-02-01; http://www.jst.go.jp/impact/en/program/16.html). The use of governmental micro data was officially approved for the funded projects. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ETHICAL STATEMENT

The need for ethical approval was waived because our study involved secondary analysis of anonymous data, under governmental-use approval.

ORCID

Megumi Kasajima https://orcid.org/0000-0003-3804-8328

REFERENCES

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, *112*(1), 3–56.

Brenner, H., Castro, F. A., Eberle, A., Emrich, K., Holleczek, B., Katalinic, A., … Cancer Survival Workgroup, G. E. K. I. D. (2016). Death Certificate Only proportions should be age adjusted in studies comparing cancer survival across populations and over time. *European Journal of Cancer*, *52*, 102–108.

Brenner, H., & Holleczek, B. (2011). Deriving valid population–based cancer survival estimates in the presence of nonnegligible proportions of cancers notified by death certificates only. *Cancer Epidemiology and Prevention Biomarkers*, *20*(12), 2480–2486.

Brunet, R. C., & Struchiner, C. J. (1999). A non-parametric method for the reconstruction of age-and time-dependent incidence from the prevalence data of irreversible diseases with differential mortality. *Theoretical Population Biology*, *56*(1), 76–90.

Cabinet Office. (2017). The Aging Society: Current situation and implementation measures FY 2017. Retrieved from http://www8.cao.go.jp/kourei/whitepaper/index-w.html

Chen, B. K., Jalal, H., Hashimoto, H., Suen, S. C., Eggleston, K., Hurley, M., … Bhattacharya, J. (2016). Forecasting trends in disability in a super-aging society: Adapting the Future Elderly Model to Japan. *The Journal of the Economics of Ageing*, *8*, 42–51. https://doi.org/10.1016/j.jeoa.2016.06.001

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine*, *21*(3), 267–275.

Dinno A. (2017). Mean-equivalence *t* tests. Stata software package. Retrieved from https://www.alexisdinno.com/stata/tost.html

Goldman, D. P., Lakdawalla, D., Michaud, P.-C., Eibner, C., Gailey, A., Vaynman, I., & Clair, P. S. (2015). The Future Elderly Model: Technical documentation. Retrieved from https://roybalhealthpolicy.usc.edu/files/2015/05/FEM_techdoc.pdf

Goldman, D. P., Hurd, M., Shekelle, P. G., Newberry, S. J., Panis, C. W. A., Shang, B., Bhattacharya, J., Joyce, G. F., & Lakdawalla, D. (2004). Health status and medical treatment of the future elderly: Final report, TR-169-CMS, Santa Monica, *CA: RAND (2004)*.

Hallett, T. B., Zaba, B., Todd, J., Lopman, B., Mwita, W., Biraro, S., … Network, A. (2008). Estimating incidence from prevalence in generalised HIV epidemics: Methods and validation. *PLoS Medicine*, *5*(4), e80.

Ichimura, H., Shimizutani, S., & Hashimoto, H. (2009). *JSTAR first results 2009 report*. Research Institute of Economy, Trade and Industry (RIETI), Discussion Paper Series 09-E-047, 1-305.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*(1), 21–37. Retrieved from. http://www.jstor.org/stable/2955359

Idler, E. L., Russell, L. B., & Davis, D. (2000). Survival, functional limitations, and self-rated health in the NHANES I epidemiologic follow-up study, 1992. *American Journal of Epidemiology*, *152*(9), 874–883.

Japanese Association of Cancer Registries. (2010). Cancer Registry in Japan second edition. Retrieved from http://www.jacr.info/publicication/document/CRIJ_eng.pdf

Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K., & Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: Projections with a Bayesian model ensemble. *The Lancet*, *389*(10076), 1323–1335.

Kubo, M., Kiyohara, Y., Kato, I., Tanizaki, Y., Arima, H., Tanaka, K., … Iida, M. (2003). Trends in the incidence, mortality, and survival rate of cardiovascular disease in a Japanese community: The Hisayama study. *Stroke*, *34*(10), 2349–2354. https://doi.org/10.1161/01.STR.0000090348.52943.A2

Ma, E., Takahashi, H., Mizuno, A., Okada, M., Yamagishi, K., & Iso, H. (2007). Stratified age-period-cohort analysis of stroke mortality in Japan, 1960 to 2000. *Journal of Stroke and Cerebrovascular Diseases*, *16*(3), 91–102. https://doi.org/10.1016/j.jstrokecerebrovasdis.2006.11.006

Michaud, P.-C., & Van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics*, *27*(5), 1312–1325.

Ministry of Health, Labour and Welfare. Outline of vital statistics in Japan. Retrieved from https://www.mhlw.go.jp/english/database/db-hw/outline/index.html

National Cancer Center. (2015). National estimates of cancer incidence based on cancer registries in Japan (1975-2013). *Cancer Registry and Statistics*. Retrieved from https://ganjoho.jp/en/professional/statistics/table_download.html

National Cancer Center. (2017). Monitoring of cancer incidence in Japan 2013. Retrieved from https://ganjoho.jp/reg_stat/statistics/brochure/monitoring.html (In Japanese)

National Institute of Population and Social Security Research. (2017a). Population projections for Japan: 2016–2065. *Population Research Series, No.336*.

National Institute of Population and Social Security Research. (2017b). Population statistics of Japan. Retrieved from http://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/db_zenkoku2017/db_zenkoku2017syosaikekka.html (In Japanese)

Roßmann, J., & Gummer, T. (2016). Using paradata to predict and correct for panel attrition. *Social Science Computer Review*, *34*(3), 312–332.

Saksena, M., & Maldonado, N. (2017). A dynamic estimation of obesity using Nhanes data: A pseudo-panel approach. *Health Economics*, *26*(12), e140–e159.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680.

Shimizutani, S., Oshio, T., & Fujii, M. (2014). Option value of work, health status, and retirement decisions in Japan: Evidence from the Japanese Study on Aging and Retirement (JSTAR). In *Social security programs and retirement around the world: Disability insurance programs and retirement* (pp. 497–535). Chicago, IL: University of Chicago Press.

Stowasser, T., Heiss, F., McFadden, D., & Winter, J. (2011). "Healthy, Wealthy and Wise?" revisited: An analysis of the causal pathways from socioeconomic status to health. In *Investigations in the Economics of Aging* (pp. 267–317). Chicago, IL: University of Chicago Press.

Tang, T., & Kurashina, S. (1987). Age, period and cohort analysis of trends in mortality from major diseases in Japan, 1955 to 1979: Peculiarity of the cohort born in the early Showa Era. *Statistics in Medicine*, *6*(6), 709–726.

Wang, C., Weber, A., & Graham, D. Y. (2015). Age, period, and cohort effects on gastric cancer mortality. *Dig Dis Sci*, *60*(2), 514–523. https://doi.org/10.1007/s10620-014-3359-0

World Health Organization. (2008). Closing the gap in a generation: Health equity through action on the social determinants of health: Commission on Social Determinants of Health Final Report 33. Geneva, Switzerland: WHO Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kasajima M, Hashimoto H, Suen S, et al. Future projection of the health and functional status of older people in Japan: A multistate transition microsimulation model with repeated cross-sectional data. *Health Economics*. 2021;30(S1):30–51. https://doi.org/10.1002/hec.3986

APPENDIX TECHNICAL DOCUMENT A.

A.1. | Adjustment of Data Source Information

A.1.1. | Population adjustment in demographics

The number of older individuals in the population census data is often underreported because of institutionalization and other reasons, while death reports are likely complete. Therefore, the number of deaths often exceeds the population size in advanced-age segments. We corrected estimates for the older population (> 80 years of age) by following protocols recommended by the Human Mortality Database project, an international collaborative project for demographic statistics, as this database has been widely adopted by many national institutions including in Japan and Australia (National Institute of Population and Social Security Research, 2010; Terblanche & Wilson, 2015; Wilmoth et al., 2007).

Because of incomplete or missing responses from census data, populations of those aged ≥ 80 years was determined using Vital Statistics (death records) microdata from 2000–2014 using extinct cohorts and survivor ratios. The protocol for this method is publicly available from the website of the Human Mortality Database Project (Wilmoth et al., 2007)).

The extinct cohorts method determines the number of survivors retrospectively by summing all counts of deaths of extinct generations for the period, under a "no immigrant" assumption. For example, as the birth cohorts born in 1898 or earlier reached extinction in 2014, their population sizes as of the year 2000 should be equal to the cumulative death counts during the years 2000–2014.

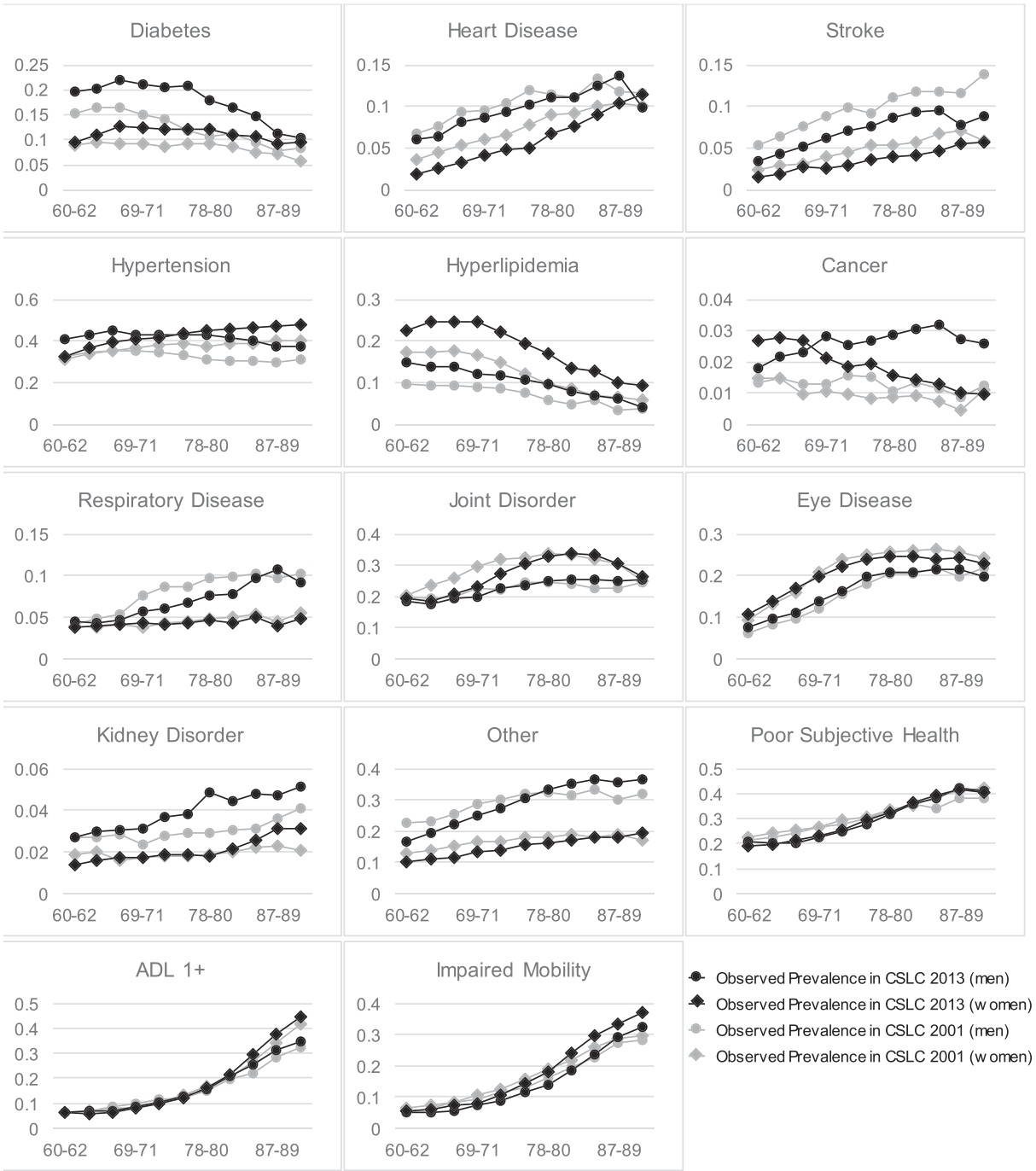
The survivor ratio method is a modified extinct cohorts method applied to pre-extinct cohorts. By estimating a proper survival ratio, one can reconstruct a past population by adding the estimated number of survivors to the accumulated death counts.

A.1.2. | Data selection for prevalence estimates of comorbidity

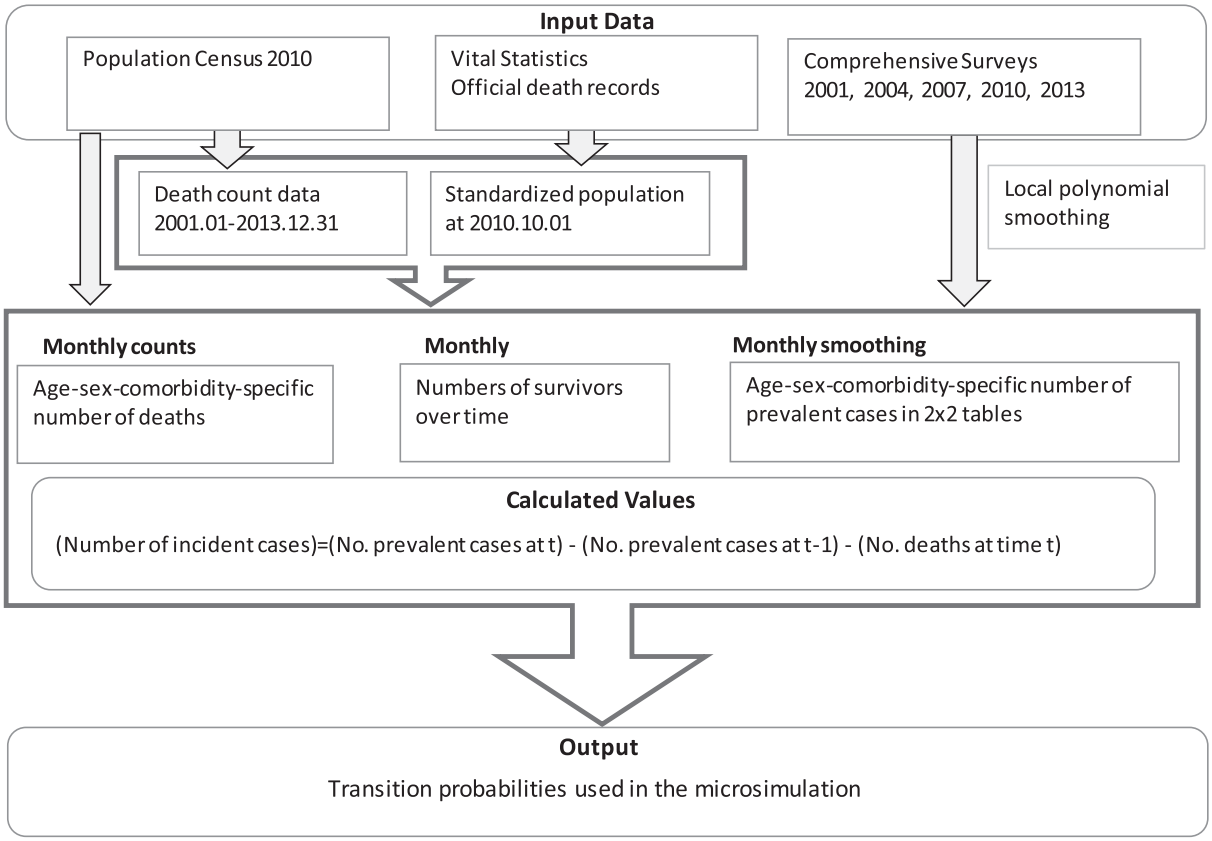
We derived the prevalence estimate of each combination of the listed disease statuses from the CSLC by sex and 3-year interval birth cohorts (Appendix Fig. 1). The CSLC asked about comorbid conditions only for individuals who received regular medical attention for chronic conditions. The proportion of individuals receiving regular medical attention increased with age. We observed a discontinuous increase in disease prevalence at age 60 years, which corresponds to the legal retirement age in Japan, suggesting improved access to medical care after retirement rather than sudden changes in health status. Based on this finding, we decided to include only individuals ≥ 60 years for our prevalence estimates. We also set 90 years as an upper age range for men, and 95 years for women because the CSLC asks for comorbidity information only from noninstitutionalized adults, and the proportion of those hospitalized and/or institutionalized exceeded nonignorable levels (e.g., 8%) at the cutoff age points. For those over the cutoff age, we adopted the same prevalence ratios as for the cutoff age.

A.1.3. | Inflation of numbers of cancer prevalence

When we compared the self-reported disease prevalence obtained from the CSLC data with the hospital-record-based numbers derived from the Japanese Patient Survey (Ministry of Health, Labour and Welfare, 2014) for our data validity check, we found that disease prevalences were comparable between the two data sources except for cancer, for which we observed approximately 30% under-reporting in the CSLC data. To address the under-reporting, we generated additional cancer cases in the CSLC sample by assigning uniform random numbers to a disease-free subpopulation (e.g., those who had no comorbidities) of each sex- and birth-specific cohort and recategorized those with the largest



APPENDIX FIGURE 1 Prevalence estimate from the CSLC data for 2001 and 2013 by sex and 3-year interval birth cohorts with synthetic panel datasets using 2 × 2 contingency tables 2.



APPENDIX FIGURE 2 Work flow of incidence determination using a contingency table method

randomly generated numbers into cancer status until the number of cancer cases matched those in the Japanese Patient Survey.

We use 2 × 2 tables, keeping exact numbers of frequencies standardized for sex-specific birth-cohort populations for all possible combinations of comorbidities as described in Note , which follows. An alternative method to the exact method is to estimate comorbidity status by generating a joint normal distribution based on the prevalence of comorbidity statuses of multiple diseases. Although this can be applied directly to panel data, its application to synthetic panel data requires an additional assumption regarding the covariance of comorbidities. Instead, we propose using the contingency table method, which does not require such a strong assumption.

In this contingency table approach, we retain the additive assumption on case fatality for comorbid status, i.e.; the model is calibrated such that the case fatality rate for each combination of comorbidities is equal to the sum of the case fatality rates for each individual morbidity status (see details in Note).

We also note that we use monthly data for cohort dynamics rather than yearly or longer time periods because some disease conditions (e.g., cancer) have a rapid turnover. Using a 1-month interval, we can safely assume that during the period, a cohort can be considered closed, and any change in the prevalence of comorbidity conditions over periods can be attributed to a dynamic equilibrium of new entries (incidence) and exits (case fatality). Although vital statistics were available on a monthly basis, the original data in the Comprehensive Survey was collected every three years. Therefore, we smoothed the 3-year interval prevalence to estimate monthly prevalence, and we decomposed the data into entry and exit data to obtain the number of incident cases.

To determine the comorbidity-specific incidence rates, we solve for equilibria of equations obtained using multiple 2 × 2 tables (see Note 3 for details). We estimate the conditional incidence rates by taking the weighted average of all possible incidence patterns calculated from the corresponding 2 × 2 tables for each comorbidity status.

A.2. | Note 1. 2 × 2 table creation

To create a 2 × 2 table, we first take two diseases and consider 0 or 1 status for each disease where 0 stands for not diagnosed and 1 stands for diagnosed. Next, we distribute the birth-sex cohort population into four cells based on the prevalence of two diseases and the frequency of two concurrent diseases. Because we have 91 (= 14 × 13/2) possible combinations of disease conditions, two gender groups (*s=1,2*), and 19 birth cohorts (*c=1,2, ... ,19*; 1903–1905 birth cohort, 1906–1908 birth cohort, continuing to the 1957–1959 birth cohort), we create 3458 (= 91 × 2 × 19) 2 × 2 tables for each survey wave. Using five waves from the Comprehensive Surveys for 2001, 2004, 2007, 2010, and 2013, we smoothed the required numbers for each cell in the contingency tables using the local polynomial method to obtain monthly prevalence.

For two arbitrary statuses, d_i and d_j, the 2 × 2 table at time t contains the initial prevalence numbers at the beginning of the month for four comorbidity patterns $(d_i, d_j) = (0,0), (1,0), (0,1), (1,1)$. Then, between time t and $t + 1$, the cohort population decreases by the number of the deceased population. The population with condition $(d_i, d_j) = (0,0)$ decreases by the base mortality rate depending on individuals' age and sex. The remaining three cells have additive mortality risks attributable to diseases describing individuals' health conditions at time t, and the population in the cells decreases by the corresponding mortality rates. The remaining populations are the numbers of survivors in each cell at the end of the month for that closed cohort population. We compared the numbers of survivors in the four cells in the 2 × 2 table with the estimated prevalence numbers in the corresponding cells of the subsequent month's table (for time $t + 1$). Differences are attributed to changes in health status because of disease incidence during the month. We computed 364 differences (four differences in 91 2 × 2 tables) over time for each sex-birth cohort using this process.

To convert differences into the status incidence cases under the 14-dimensional health condition vectors, we built systems of difference equations to estimate the conditional incidence probabilities (see the appendix technical document for details). Consequently, we obtained a total of 114 688 (= $14 \times (2^{14}/2)$) monthly conditional incidence probabilities as solutions for each sex, birth cohort, and month.

Finally, we translated the sex-birth cohort-specific conditional incidence rates into age-sex-specific conditional incidence rates in the following two steps: after translating the monthly rates into annual rates for the consecutive 13 years, we regressed the pooled incidence estimates for each of the 114 688 combinations of comorbidities for age, age squared, and birth-cohort dummy variables to incorporate cohort-specific fixed effects, for men and women separately.

A.3. | Note 2. Age-sex-disease-specific mortality rates

In this section, we estimate mortality rates (Appendix Table 1-(b)) using monthly 2 × 2 tables (Appendix Table 1-(a)). Let us denote the case fatality rate attributable to disease 1 (diabetes in the following example) as $\alpha_{1(c,s,t)}$, the case fatality rate attributable to disease 2 (heart disease in the following example) as $\alpha_{2(c,s,t)}$, and the mortality rate for other conditions as $\alpha^{(1,2)}_{base(c,s,t)}$.

TABLE 1 2 × 2 tables of population and mortality rates

Table 1-(a)		**d_1**	
Population		0	1
d_2	0	$pop^{(1,2)}_{c,s,t}(0,0)$	$pop^{(1,2)}_{c,s,t}(1,0)$
	1	$pop^{(1,2)}_{c,s,t}(0,1)$	$pop^{(1,2)}_{c,s,t}(1,1)$

Table 1-(b)		**d_1**	
Mortality rate		0	1
d_2	0	$\alpha^{(1,2)}_{base(c,s,t)}$	$\alpha_{1(c,s,t)} + \alpha^{(1,2)}_{base(c,s,t)}$
	1	$\alpha_{2(c,s,t)} + \alpha^{(1,2)}_{base(c,s,t)}$	$\alpha_{1(c,s,t)} + \alpha_{2(c,s,t)} + \alpha^{(1,2)}_{base(c,s,t)}$

Under the assumption of additive mortality rates, the following equations hold:

(Observed mortality from diabetes in the vital statistics data) $= \alpha_{1(c,s,t)} * \{pop^{(1,2)}_{c,s,t}(1,0) + pop^{(1,2)}_{c,s,t}(1,1)\}$

(Observed mortality from heart disease in the vital statistics data) $= \alpha_{2(c,s,t)} * \left\{pop^{(1,2)}_{c,s,t}(0,1) + pop^{(1,2)}_{c,s,t}(1,1)\right\}$

(Observed mortality from diseases other than diabetes or heart disease)

$$= \alpha^{(1,2)}_{base(c,s,t)} * \left\{ pop^{(1,2)}_{c,s,t}(0,0) + pop^{(1,2)}_{c,s,t}(1,0) + pop^{(1,2)}_{c,s,t}(0,1) + pop^{(1,2)}_{c,s,t}(1,1) \right\}.$$

We attribute additional mortality exits from the poor subjective health cell to mental health conditions. Because of data limitations, we assume that impaired mobility and dysfunctions in activities of daily living do not independently raise mortality risk. Therefore, regardless of dysfunctions in activities of daily living and mobility, the probability of mortality depends on subjective health and the 11 diseases. Definitions of cause specific death with International Classification of Death cause version10 (ICD-10) were listed in Appendix Table 2.

TABLE 2 List of categories in ICD-10 for calculation of cause-specific mortality from vital statistics.

Disease categories	ICD-10
Diabetes	E10-E14
Coronary heart diseases	I20-I25
Stroke	I60-I69
Hypertension	I10, I11, I12, I13, I15
Hyperlipidemia	E78
Cancer	C00 - C97
All respiratory diseases	J10-J22, J40-J47, J60-J70, J80-J84, J99, A15-A16
Joint disorders (rheumatoid arthritis, collagen vascular disease)	M05-M08, M10-M14, M15-M19, M40-M54
Eye diseases	H25-H28, H30-H36, H40-H42
Kidney disorders	N00-N07, N10-N15, N17-N19
Others	I00-I09, I26-I52, K00-K99
Liver	B15-B19, K70-K77
Ulcer	K25-K27, K29
Prostatic hyperplasia	N40
Mental disorders	F20-F48, X60-X84

A.4. | Note 3. Conditional incidence using 2 × 2 tables

In the next step, to estimate the conditional incidence rates, we count the monthly incidence as the difference between the number of survivors and the prevalence number in the subsequent period. Using the example of the diabetes-heart disease table as in Appendix Table 3, we obtain the following four incidence numbers:

TABLE 3 Population flow (gray arrows) because of disease incidence in equations (1)–(4) in the (d_1,d_2) 2 × 2 table

Eq(1)

d_2 \ d_1	0	1
0	$pop^{(1,2)}_{c,s,t}(0,0)$ →	$pop^{(1,2)}_{c,s,t}(1,0)$
1	$pop^{(1,2)}_{c,s,t}(0,1)$	$pop^{(1,2)}_{c,s,t}(1,1)$

Eq(2)

d_2 \ d_1	0	1
0	$pop^{(1,2)}_{c,s,t}(0,0)$	$pop^{(1,2)}_{c,s,t}(1,0)$
1	$pop^{(1,2)}_{c,s,t}(0,1)$ →	$pop^{(1,2)}_{c,s,t}(1,1)$

Eq(3)

d_2 \ d_1	0	1
0	$pop^{(1,2)}_{c,s,t}(0,0)$	$pop^{(1,2)}_{c,s,t}(1,0)$
1	$pop^{(1,2)}_{c,s,t}(0,1)$ →	$pop^{(1,2)}_{c,s,t}(1,1)$

Eq(4)

d_2 \ d_1	0	1
0	$pop^{(1,2)}_{c,s,t}(0,0)$ →	$pop^{(1,2)}_{c,s,t}(1,0)$
1	$pop^{(1,2)}_{c,s,t}(0,1)$	$pop^{(1,2)}_{c,s,t}(1,1)$

- Incidence of diabetes from $pop^{(1,2)}_{c,s,t}(0,0)$
- Incidence of heart disease from $pop^{(1,2)}_{c,s,t}(0,0)$
- Incidence of diabetes from $pop^{(1,2)}_{c,s,t}(0,1)$

- Incidence of heart disease from $pop^{(1,2)}_{c,s,t}(1,0)$

The incidence numbers satisfy the following relationships if we assume closed cohorts:
Population flow from $[(d_1,d_2)=(0,0)]$
Incidence of heart disease from $[(d_1,d_2)=(0,0)]$ + Incidence of diabetes from $[(d_1,d_2)=(0,0)]$

$$= \text{Survivors in } [(d_1,d_2)=(0,0)] - \text{Prevalence of having the condition } [(d_1,d_2)=(0,0)] \text{ at } (t+1)$$

$$= pop^{(1,2)}_{c,s,t}(0,0) * \left(1-\alpha^{(1,2)}_{base(c,s,t)}\right) - pop^{(1,2)}_{c,s,t+1}(0,0) \quad \text{(Eq1)}$$

Population flow into $[(d_1,d_2)=(1,1)]$
Incidence of heart disease from $[(d_1,d_2)=(1,0)]$ + Incidence of diabetes from $[(d_1,d_2)=(0,1)]$

$$= \text{Prevalence of having the condition } [(d_1,d_2)=(1,1)] \text{ at } (t+1) - \text{Survivors in } [(d_1,d_2)=(1,1)]$$

$$= pop^{(1,2)}_{c,s,t+1}(1,1) - pop^{(1,2)}_{c,s,t}(1,1) * \left(1-\alpha^{(1,2)}_{base(c,s,t)} - \alpha_{1(c,s,t)} - \alpha_{2(c,s,t)}\right) \quad \text{(Eq2)}$$

Population flow in/out $[(d_1,d_2)=(0,1)]$
Incidence of heart disease from $[(d_1,d_2)=(0,0)]$ − Incidence of diabetes from $[(d_1,d_2)=(0,1)]$

$$= \text{Prevalence of having the condition } [(d_1,d_2)=(0,1)] \text{ at } (t+1) - \text{Survivors in } [(d_1,d_2)=(0,1)]$$

$$= pop^{(1,2)}_{c,s,t+1}(0,1) - pop^{(1,2)}_{c,s,t}(0,1) \times \left(1-\alpha^{(1,2)}_{base(c,s,t)} - \alpha_{2(c,s,t)}\right) \quad \text{(Eq3)}$$

Population flow in/out $[(d_1,d_2)=(1,0)]$
Incidence of diabetes from $[(d_1,d_2)=(0,0)]$ − Incidence of heart disease from $[(d_1,d_2)=(1,0)]$

$$= \text{Prevalence of having the condition } [(d_1,d_2)=(1,0)] \text{ at } (t+1) - \text{Survivors in } [(d_1,d_2)=(1,0)]$$

$$= pop^{(1,2)}_{c,s,t+1}(1,0) - pop^{(1,2)}_{c,s,t}(1,0) \times \left(1-\alpha^{(1,2)}_{base(c,s,t)} - \alpha_{1(c,s,t)}\right) \quad \text{(Eq4)}$$

This system of equations cannot be uniquely solved because of a lack of constraint conditions. However, we can find the equilibrium of the solutions for Eq(1)−Eq(4) using a relevant set of multiple 2 × 2 tables. Let us denote "the incidence rate of diabetes (disease 1) from $[(d_1,d_2)=(0,0)]$ condition in the (d_1,d_2) − 2 × 2 table" by $incidence^{(1,2)}_{1(c,s,t)}(0,0)$, and denote "the incidence rate of heart disease (disease 2) from $[(d_1,d_2)=(0,0)]$ condition in the (d_1,d_2) – 2 × 2 table" by $incidence^{(1,2)}_{2(c,s,t)}(0,0)$. Then we can rewrite Eq(1) as:

$$incidence^{(1,2)}_{1(c,s,t)}(0,0) + incidence^{(1,2)}_{2(c,s,t)}(0,0) = 1-\alpha^{(1,2)}_{base(c,s,t)} - \frac{pop^{(1,2)}_{c,s,t+1}(0,0)}{pop^{(1,2)}_{c,s,t}(0,0)}. \quad \text{(Eq1)}$$

We average the values on the right side of Eq(1)′ for 12 consecutive months in a certain year. Because we have 91 2 × 2 tables, we obtain 91 patterns of monthly averages for Eq(1)′ for each sex, cohort, and year. To separate the elements on the left side of Eq(1)′, we use the multiple 2 × 2 tables listed in Appendix Table 4.

TABLE 4 Set of 26 Eq(1)′s to determine the incidence rates of diabetes and heart disease

Eq(1)′s including diabetes incidence (disease 1)	Eq(1)′s including heart disease (disease 2)
$incidence^{(1,2)}_{1(c,s,t)}(0,0)+incidence^{(1,2)}_{2(c,s,t)}(0,0)$	$incidence^{(1,2)}_{1(c,s,t)}(0,0)+incidence^{(1,2)}_{2(c,s,t)}(0,0)$
$incidence^{(1,3)}_{1(c,s,t)}(0,0)+incidence^{(1,3)}_{3(c,s,t)}(0,0)$	$incidence^{(2,3)}_{2(c,s,t)}(0,0)+incidence^{(2,3)}_{3(c,s,t)}(0,0)$
$incidence^{(1,4)}_{1(c,s,t)}(0,0)+incidence^{(1,4)}_{4(c,s,t)}(0,0)$	$incidence^{(2,4)}_{2(c,s,t)}(0,0)+incidence^{(2,4)}_{4(c,s,t)}(0,0)$
$incidence^{(1,5)}_{1(c,s,t)}(0,0)+incidence^{(1,5)}_{5(c,s,t)}(0,0)$	$incidence^{(2,5)}_{2(c,s,t)}(0,0)+incidence^{(2,5)}_{5(c,s,t)}(0,0)$
$incidence^{(1,6)}_{1(c,s,t)}(0,0)+incidence^{(1,6)}_{6(c,s,t)}(0,0)$	$incidence^{(2,6)}_{2(c,s,t)}(0,0)+incidence^{(2,6)}_{6(c,s,t)}(0,0)$
$incidence^{(1,7)}_{1(c,s,t)}(0,0)+incidence^{(1,7)}_{7(c,s,t)}(0,0)$	$incidence^{(2,7)}_{2(c,s,t)}(0,0)+incidence^{(2,7)}_{7(c,s,t)}(0,0)$
$incidence^{(1,8)}_{1(c,s,t)}(0,0)+incidence^{(1,8)}_{8(c,s,t)}(0,0)$	$incidence^{(2,8)}_{2(c,s,t)}(0,0)+incidence^{(2,8)}_{8(c,s,t)}(0,0)$
$incidence^{(1,9)}_{1(c,s,t)}(0,0)+incidence^{(1,9)}_{9(c,s,t)}(0,0)$	$incidence^{(2,9)}_{2(c,s,t)}(0,0)+incidence^{(2,9)}_{9(c,s,t)}(0,0)$
$incidence^{(1,10)}_{1(c,s,t)}(0,0)+incidence^{(1,10)}_{10(c,s,t)}(0,0)$	$incidence^{(2,10)}_{2(c,s,t)}(0,0)+incidence^{(2,10)}_{10(c,s,t)}(0,0)$
$incidence^{(1,11)}_{1(c,s,t)}(0,0)+incidence^{(1,11)}_{11(c,s,t)}(0,0)$	$incidence^{(2,11)}_{2(c,s,t)}(0,0)+incidence^{(2,11)}_{11(c,s,t)}(0,0)$
$incidence^{(1,12)}_{1(c,s,t)}(0,0)+incidence^{(1,12)}_{12(c,s,t)}(0,0)$	$incidence^{(2,12)}_{2(c,s,t)}(0,0)+incidence^{(2,12)}_{12(c,s,t)}(0,0)$
$incidence^{(1,13)}_{1(c,s,t)}(0,0)+incidence^{(1,13)}_{13(c,s,t)}(0,0)$	$incidence^{(2,13)}_{2(c,s,t)}(0,0)+incidence^{(2,13)}_{13(c,s,t)}(0,0)$
$incidence^{(1,14)}_{1(c,s,t)}(0,0)+incidence^{(1,14)}_{14(c,s,t)}(0,0)$	$incidence^{(2,14)}_{2(c,s,t)}(0,0)+incidence^{(2,14)}_{14(c,s,t)}(0,0)$

Arithmetically subtracting the sum of the elements in the right column from the sum of the elements in the left column in Appendix Table 4, we obtain the following equations:

$$\sum_{\substack{k\neq 1\\ k=1,\ldots,14}}\left(incidence^{(1,k)}_{1(c,s,t)}(0,0)+incidence^{(1,k)}_{k(c,s,t)}(0,0)\right)-\sum_{\substack{k\neq 2\\ k=1,\ldots,14}}\left(incidence^{(2,k)}_{2(c,s,t)}(0,0)+incidence^{(2,k)}_{k(c,s,t)}(0,0)\right)$$

$$\fallingdotseq 12\left(incidence_{1(c,s,t)}\left(\vec{0}\right)-incidence_{2(c,s,t)}\left(\vec{0}\right)\right)$$

where $incidence_{1(c,s,t)}\left(\vec{0}\right)$ denotes the equilibrium of the incidence rate of diabetes from disease-free condition (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14})$=\vec{0}$, and $incidence_{2(c,s,t)}\left(\vec{0}\right)$ denotes the equilibrium of the incidence rate of heart disease from disease-free condition.

The first row of Appendix Table 4 can be approximated by $incidence^{(1,2)}_{1(c,s,t)}(0,0)+incidence^{(1,2)}_{2(c,s,t)}(0,0)\fallingdotseq incidence_{1(c,s,t)}\left(\vec{0}\right)+incidence_{2(c,s,t)}\left(\vec{0}\right)$, which determines the solutions for Eq(1)', $incidence^{(1,2)}_{1(c,s,t)}(0,0)$, and $incidence^{(1,2)}_{2(c,s,t)}(0,0)$. The solution to Eq(1)′ sequentially provides the remainder of the solutions in the system of equations (1)–(4), $incidence^{(1,2)}_{1(c,s,t)}(0,1)$, and $incidence^{(1,2)}_{2(c,s,t)}(1,0)$.

From the estimated monthly incidence rates using 91 2 × 2 tables above, we calculated the conditional incidence probabilities under the condition (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14}). We converted the incidence rates in the 2 × 2 table form to conditional incidence probabilities of disease k in the 14 dimensional health status form by the weighted average:

$$\frac{\sum_{l \neq k} incidence^{(k,l)}_{k(c,s,t)}(d_k, d_l) \times pop^{(k,l)}_{c,s,t}(d_k, d_l)}{\sum_{l \neq k} pop^{(k,l)}_{c,s,t}(d_k, d_l)}_{l=1,\ldots,14}.$$

For example, to calculate the conditional incidence probability of diabetes (k=1) under the condition (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14})=(0,0,0,1,0,0,0,1,0,0,0,0,0,0), we took the weighted average of incidence rates of diabetes in 13 2 × 2 tables, such as $incidence^{(1,2)}_{1(c,s,t)}(0,0)$, $incidence^{(1,3)}_{1(c,s,t)}(0,0)$, $incidence^{(1,4)}_{1(c,s,t)}(0,1)$, $incidence^{(1,5)}_{1(c,s,t)}(0,0)$, $incidence^{(1,6)}_{1(c,s,t)}(0,0)$, $incidence^{(1,7)}_{1(c,s,t)}(0,0)$, $incidence^{(1,8)}_{1(c,s,t)}(0,1)$, $incidence^{(1,9)}_{1(c,s,t)}(0,0)$, $incidence^{(1,10)}_{1(c,s,t)}(0,0)$, $incidence^{(1,11)}_{1(c,s,t)}(0,0)$, $incidence^{(1,12)}_{1(c,s,t)}(0,0)$, $incidence^{(1,13)}_{1(c,s,t)}(0,0)$, and $incidence^{(1,14)}_{1(c,s,t)}(0,0)$.

REFERENCE

Ministry of Health, Labour and Welfare. (2014). Patient Survey. Retrieved from http://www.mhlw.go.jp/english/database/db-hss/ps.html

National Institute of Population and Social Security Research. (2010). Japanese Mortality Database Brief Summary. Retrieved from http://www.ipss.go.jp/p-toukei/JMD/briefsummary-en.html

Terblanche, W., & Wilson, T. (2015). An evaluation of nearly-extinct cohort methods for estimating the very elderly populations of Australia and New Zealand. *PLoS One, 10*(4), e0123692.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., ... Vachon, P. (2007). Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007], 9*, 10-11.