Stanford | Asia Health Policy Program
Shorenstein APARC

Office 650.723.9741
Fax 650.723.6530

Stanford University
Encina Hall, E301
Stanford, CA 94305-6055

aparc.fsi.stanford.edu

Stanford University
Walter H. Shorenstein Asia-Pacific Research Center
Asia Health Policy Program

Working paper series
on health and demographic change in the Asia-Pacific

Future Projection of the Health and Functional Status of Older People in Japan: A Pseudopanel Microsimulation Model*

Megumi Kasajima, *Corresponding Author Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Tokyo, Japan*
Hideki Hashimoto, *Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Tokyo, Japan*
Sze-Chuan Suen, *Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA*
Brian Chen, *Department of Health Services Policy and Management, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA*
Karen Eggleston, *FSI Shorenstein Asia Pacific Research Center, Stanford University, Stanford, CA, USA*
Jay Bhattacharya, *Center for Primary Care and Outcomes Research, Stanford School of Medicine, Stanford, CA, USA*

Asia Health Policy Program working paper #55

January 16, 2019

https://aparc.fsi.stanford.edu/asiahealthpolicy
For information, contact: Karen N. Eggleston (翁笙和)
Walter H. Shorenstein Asia-Pacific Research Center
Freeman Spogli Institute for International Studies
Stanford University
616 Serra St., Encina Hall E311
Stanford, CA 94305-6055
(650) 723-9072; Fax (650) 723-6530
karene@stanford.edu

Future projection of the health and functional status of older people in Japan: A pseudopanel microsimulation model *

Megumi Kasajima [1], Hideki Hashimoto [2], Sze-Chuan Suen [3], Brian Chen [4], Karen Eggleston [5], and Jay Bhattacharya [6]

January 2019

Abstract

Background: Precise future projection of population health distribution is imperative for designing an efficient healthcare system in rapidly aging countries. Multistate-transition microsimulation models such as the US Future Elderly Model have been developed based on panel data collection, but these data may not be always available. We proposed a pseudopanel method using repeated cross-sectional representative surveys as a complementary approach, and specifically applied the model to Japan's population.

Methods: We calculated birth-cohort and sex-specific prevalence for all combinations of 14 health statuses using microdata from five waves of the Comprehensive Survey of People's Living Conditions. Combining obtained prevalences with vital statistics data, we determined transition probabilities of statuses over time using contingency tables. Assuming that state transition and mortality-exit follow the first-order Markov process, we then designed a virtual Japanese population aged older than 60 years as of 2013 and performed a microsimulation to project disease distributions to 2046 with forward, backward, and external validation tests. Following validation, we compared our projection results with those based on traditional static models.

Results: Our calculated morbidity and mortality rates successfully replicated governmental projections of population pyramids and matched cardiovascular and cancer incidences reported in existing epidemiological studies, supporting the validity of our estimation. Our future projection of stroke and heart disease indicated lower prevalences than expected from static models, presumably because of recent declining trends in disease incidence and fatality.

Conclusions: Our pseudopanel approach provides a valid alternative microsimulation frame for future health projection in aging societies.

Keywords: Pseudopanel approach; microsimulation; forecasting; aging; comorbidities; Japan

JEL No. C53, I1, I12, J11, J14

---

*This study is supported by Ministry of Health, Labour and Welfare in Japan (H26-Chikyukibo-ippan-001) https://www.mhlw.go.jp/english/index.html 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. We express our gratitude to Dr Michiaki Kubo for providing detailed heart disease and stroke incidence rates based on the Hisayama cohort study, and to Professors Toshiharu Ninomiya and Yutaka Kiyohara for their contributions to developing the study. We are grateful to Dr Atsushi Goto for his expert opinion on diabetes, Dr Futoshi Ishii for advising on the protocol used in the Human Mortality Database, and Dr Tomohiro Matsuda for detailed information on the national cancer registry statistics. We thank Mr Kent Byun and Mr Ajay Anand for technical assistance in coding, and Dr Tetsuya Iwamoto for his assistance in the secondary analysis of Japanese governmental microdata. We express our deepest appreciation to Professor Masaru Kitsuregawa and Dr Kazuo Goda who provided the big data platform we used to conduct our simulation study.

[1] Corresponding Author. Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Tokyo, Japan
[2] Department of Health and Social Behavior, School of Public Health, University of Tokyo, Bunkyo-ku, Tokyo, Japan
[3] Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
[4] Department of Health Services Policy and Management, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA
[5] FSI Shorenstein Asia Pacific Research Center, Stanford University, Stanford, CA, USA
[6] Center for Primary Care and Outcomes Research, Stanford School of Medicine, Stanford, CA, USA

Introduction

Rapid population aging is considered a risk to social sustainability in Asian countries including Japan, where the overall population is decreasing, and the proportion of the population aged ≥ 65 years has already reached 27.3% [1]. Increasing demand for old-age pensions, medical services, and long-term care, at a time when revenues to support such services are projected to decrease, threatens the nation's economic future. Population aging also leads to considerable disparity among older people regarding their economic, health, and social resources [2]. Precise 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 statuses usually assume a static and average status for comorbidity prevalence and mortality by age and sex strata [3]; however, such projections may 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 US Future Elderly Model, was developed based on comprehensive information available in large panel datasets such as the Health and Retirement Study [4, 5]. A similar projection model using the existing available panel dataset for Japan has also been constructed [6]. Panel datasets are preferred for modeling highly complex processes in human aging [7] because they provide data on individual transitions of dynamic status in the real world. Longitudinal datasets enable us to disentangle age, cohort, and period effects and help control for unobserved heterogeneity [8].

Microsimulation models have been useful in modeling dynamic population changes in health, but the availability

of panel data for estimating transition probabilities between health states often remains limited. Instead, repeated cross-sectional survey data with consistent sampling frame over time (e.g., nationally-representative surveys of health conditions) are more readily available in many countries. In this study, we proposed a pseudopanel approach using repeated cross-sectional survey data for microsimulations when sufficient panel data are unavailable.

Methods

Data sources

Our model requires repeated cross-sectional data of a closed cohort of the target population (i) collected over at least two waves, which allows us to estimate health state transition probabilities; and (ii) 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 with evenly spaced knots over time.

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 to obtain the prevalence of different combinations of comorbidities and functional disabilities. Because Japan has limited immigration and emigration, the survey provides a consistent sample of a closed cohort of Japanese elderly. This dataset thus meets the necessary conditions we list above.

Approximately 600 000 individuals belonging to 295 000 households were sampled by two-stage cluster sampling

in each wave. We calculated time trends in disease prevalence and functional statuses specific to age-sex-specific cohorts to create a pseudopanel from 2001–2013. 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 [9]. We obtained the monthly age-sex-disease specific probability of death to reflect changing trends in cause-specific mortality over periods, and we standardized the cohort population using data from the 2010 population census.

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 [10–12]. A summary description of correction protocols (extinct cohorts and survivor-ratio methods) is available in the appendix technical document.

Details of the estimation methods for comorbidity and mortality are available in our appendix technical document. In the following sections, we outline the estimation methods that we used.

Health and functional status variables

Fourteen health status variables consisted of self-reported morbidity conditions for 11 chronic diseases,

subjective health status, and two dysfunctional statuses. 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 [13–15]. We indicated "1" if the respondent reported “poor” or “very poor” health. Finally, dysfunctions in activities of daily living (at least one condition among: getting out of bed, bathing, dressing, and eating) and in mobility (difficulty going out) were also counted. We assumed all states of health and functional status were absorbing, or that no recovery was possible once chronic diseases and functional disabilities had developed.

We derived the prevalence estimate of each combination of the listed statuses from the CSLC by sex and 3-year interval birth cohorts. The CSLC asked about comorbid conditions only for those who had regular medical attention for chronic conditions, and the proportion of those with regular medical attention increased with age. We observed a discontinuous increase at age 60 years that corresponded to the legal retirement age in Japan, indicating improved access after retirement rather than sudden change in health status. Based on this finding, we decided to include only those 60 years and over for 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 non-institutionalized 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.

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

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 of statuses over time using *contingency tables*. To build these tables, we first determined the change in prevalence of each combination of comorbidities over time, linked the information with death record statistics to account for death, and used the values to calculate the incidence of each combination of comorbidities.

This procedure relies on five assumptions. First, following previous models [5], we assumed Granger causality. This is because our aim focused on predicting future health statuses, rather than identifying causal pathways.

Second, we assumed, for model simplicity, that all chronic conditions were absorbing states in the Markov process, or that there is no recovery from any chronic conditions in our study. This assumption is further justified by the fact that, for many chronic conditions (such as diabetes) there is no cure. Third, 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 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 also included a base rate of age-sex-specific mortality. Fourth, we assumed that disease prevalence, incidence, and death reached equilibrium over a given time period. Using this assumption, we calculated disease incidence using estimated death probabilities and the change in prevalence over that period. Finally, because of data limitations, we assumed the total death probability was additive across concurrent conditions, e.g., 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. (followed by Prevalence, Mortality, and Incidence subsections)

Determination of health status prevalence

To model changing trends in correlated statuses over waves based on repeated cross-sectional data, Deaton (1985) [17], and others, more recently [18–20], used a dynamic regression-based synthetic cohort model. Researchers usually apply the regression-based approach to a single data source that captures status transitions including exit from the target cohort. Because we had to rely on different data sources to capture comorbidity prevalences and death exit, we could not use this regression-based method. To incorporate death exit while

maintaining covariance between comorbidities obtained from different data sources, we developed a method for estimating transition probabilities between health states formed by patterns of comorbidities for each sex and birth cohort over time using repeated cross-sectional data.

Because CSLC data had a 3-year interval between waves, we smoothed the required numbers for each cell in the contingency tables using the local polynomial method to obtain monthly prevalence. We chose 1 month as the unit of time interval for incidence estimation. We assumed that change in prevalence was defined by entry of incident cases and exit by death during the period 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 (Appendix Figure in the appendix technical document).

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

Mortality estimation

Vital statistics data listing multiple causes of death linked with past comorbidity history is the ideal data to estimate cause-specific case fatality. Unfortunately, the official Japanese vital statistics data contained only a

single cause of death. With this limitation in data availability, 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 fatalities for the 11 chronic diseases we listed, and the additive probabilities of mortality from corresponding comorbidity statuses. Then, 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 birth cohort population. We adopted local polynomial smoothing with four age-year width bands around the kernel to obtain age-specific mortality curves.

Incidence estimation

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

More specifically, assuming d_i and d_j are two arbitrary statuses, 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. We used the values for age-sex-specific conditional incidence rates for 2010, 2011, and 2012 for the future projection.

Simulation

With the estimated sex-age-comorbidity-specific incidence and case-fatality rates, we performed a microsimulation for future projection of disease distributions. We represented health state transition and mortality

exit using a first-order Markov process; specifically, we assumed that disease incidence and mortality at time $t + 1$ depended only on disease comorbidity condition 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 probabilities, 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 60–62-year-olds, using birth-cohort population and mortality rates as of 2011–2013 up to age 60 for the standardized years of age. We assumed the population disease distributions were the same as those of 60, 61, and 62 year-olds in the middle of 2013.

Validation of the simulation parameters

We tested the validity of our simulation parameters by forward, external, and backward validation. In the forward validation, 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 [21]. We compared our values for disease incidence with those reported in existing epidemiological cohort studies conducted in Japan for cardiovascular diseases [22] and cancer [23]. We determined age-sex-specific annual

incidence rates for each disease as an average weighted by the frequency of comorbid combinations seen in 2013. Finally, we validated our parameter estimation by backward validation. We performed a simulation with a 2001 cohort based on CSLC 2001 prevalence data as a baseline population to predict the disease prevalence for 2013, and compared the projected results with the actual statistics 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, we calculated disease prevalence based on the static model by multiplying sex cohort-specific prevalence rates by the governmental official population estimates of 2022, 2034, and 2046.

Ethical considerations

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

Results

Forward validation

Figure 1 shows the future projection of the age-sex-specific population for men in blue and for women in red, by 3-year-interval birth cohorts for 2022, 2034, and 2046, using the 2013 population structure as a baseline. Outlined in green are the governmental population projections published by the National Institute of Social Security and

Population Study. Our projected population for 2022 was statistically equivalent to the existing governmental projection of the Japanese population structure. However, we observed a small but significant underestimation of those aged ≥ 75 years in the long-term simulations for 2034 and 2046.

External validation

Figure 2 shows the validated annual incidence rate by age for coronary heart disease (Fig 2(a)) and stroke (Fig 2(b)) in 2013, for men (blue) and women (red). The weighted averages of conditional incidence rates are described by the solid lines, and the range between the 5th and 95th percentiles is shadowed. Green dots plot incidence rates obtained from the published heart and stroke incidences in the Hisayama study from the 1990s [22], which were higher in all age groups compared with our estimations.

In Figure 3, 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, or those who were not in the registry but were retrospectively identified by death certificate [24]. 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.

Backward validation

As shown in Figure 4, we confirmed that our simulation model produced health status prevalences 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 checked prevalences of joint conditions, and confirmed our model replicated up to 3rd order joint distribution of comorbidities in our backward validation.

Trends in mortality and incidence estimation (Supplemental figures describing the results)

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 disease case fatalities for all diseases in our observation period, while 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 incidences of these diseases over time.

Future projection of disease prevalence

Figure 5(a)–(d) presents projected health status prevalences by age and sex. Our projection indicated a lower prevalence of coronary heart disease and stroke than the numbers based on the static model assumption where disease prevalence as of 2013 was multiplied by the projected population published by the National Institute of Population Study (Fig 5(a) and 5(b)). 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% [1], our projection estimated the stroke prevalence at 1.8 million compared with 2.2 million in a static model. Concomitantly, the prevalence of difficulties in activities of daily living and impaired mobility was lower in our projection results compared with those in a static model (Supplemental figures describing the results).

Our cancer and respiratory disease prevalence projections showed a similar structure compared with 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 the 2013 prevalences. However, our projection results indicated that both incidence and case fatality will decline in the future and cancel out, resulting in a similar prevalence projection, but with longer survival of the affected population (Fig 5(c) and 5(d)).

Discussion

In this study, we proposed a pseudopanel approach for future projection of health conditions in older populations based on a microsimulation using repeated cross-sectional representative surveys in Japan. The validity of our estimations of disease incidences and case fatalities was supported by statistically equivalent numbers obtained from existing epidemiological cohort data. Our projection results indicated that future projection of disease prevalence by traditional static models may overestimate/underestimate 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 adopted 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

[25]. We did not rely on this hypothetical adjustment. Instead, our projection empirically estimated improved base mortality rates and lowered incidence/case fatalities for several disease conditions as a driver for improved population longevity. Our results were comparable with other projections using Bayesian modeling without such assumptions [26].

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, while the Hisayama study defined disease diagnosis based on clinical examinations and autopsy findings, which should have detected more asymptomatic cases [22]. Hisayama's report was also based mainly on observations during the 1990s when the incidence of cardiovascular disease was higher compared with the current incidence. With these considerations, we believe that our estimates reflected 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 registered 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 [27]. 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)

[28, 29]. Indeed, when we performed ad hoc re-estimation 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 the registry numbers. Therefore, we believe that actual numbers may lie somewhere between the Cancer Registry numbers and our estimates.

In the backward validation, we slightly overestimated the prevalence of hyperlipidemia and joint disorders, likely because our absorbing assumption (no recovery) may not have reflected the natural course of these conditions. Otherwise, our simulation accurately captured real-world health transitions of older Japanese from 2001–2013. 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 this was because of lower incidence and better survival for the same conditions in the future older population, which may have significant implications for future health policy decisions.

Our proposed pseudopanel approach 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 to panel datasets. Our proposed approach also may be useful when an existing panel suffers nonignorable attrition.

We acknowledge that our method requires refinement. The model could be extended to include a wider list of comorbid conditions including cognitive dysfunction, for example. 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. The model could also be expanded to include estimates of medical and long-term care costs for projected comorbidities. However, it is important to note that current projections already require significant computing resources. We improved the efficiency in incidence calculation using Python as a programming language, but our method still requires a “big data” platform to run the microsimulation with existing statistical packages such as Stata14 (Stata Corp., College Station, TX).

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 successfully incorporate socioeconomic factors into the model in future iterations of our

work [30, 31]. Health affects, and is affected by, socioeconomic conditions [32], and changes in life conditions and available health technologies over time lead to changes in people's likelihood of health, function, comorbidity, and death [33–35]. 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 admit 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 in some combinations of diseases depending on their synergetic or competing impacts on case fatality. Second, we postulated that all chronic conditions are absorbing states 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 60–62-year-old cohorts will be as healthy as those in the 2013 data. Instead, we could have incorporated future time trends for the health statuses of incoming cohorts, as in the original Future Elderly Model [5]. 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. For example, if a new diabetic drug were developed that decreases the probability of developing heart disease by 5%, we could modify our transition probabilities for developing heart disease given diabetes by 5% and run a counterfactual simulation to see the impact of the new drug on the future distribution of diabetes and

heart disease. 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 pseudopanel method could be a useful tool to broaden the potential of microsimulations.

References

1. Cabinet Office. The aging society: current situation and implementation measures FY 2017. 2017. Available from: http://www8.cao.go.jp/kourei/whitepaper/index-w.html. (Accessed January 15, 2018)

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

3. National Institute of Population and Social Security Research. Population Projections for Japan: 2016-2065. Population Research Series No.336. 2017. Available from: http://www.ipss.go.jp/pp-zenkoku/e/zenkoku_e2017/pp_zenkoku2017e.asp. (Accessed December 2, 2017)

4. Goldman DP, Shekelle PG, Bhattacharya J, Hurd M, Joyce GF. Health status and medical treatment of the future elderly. Rand Corp., Santa Monica, CA, 2004. Available from: http://www.rand.org/pubs/technical_reports/TR169.html. (Accessed December 2, 2017)

5. Goldman DP, Lakdawalla D, Michaud P-C, Eibner C, Gailey A, Vaynman I, et al. The Future Elderly Model: Technical documentation. 2015. Available from: https://roybalhealthpolicy.usc.edu/files/2015/05/FEM_techdoc.pdf. (Accessed December 2, 2017)

6. Chen BK, Jalal H, Hashimoto H, Suen SC, Eggleston K, Hurley M, et al. Forecasting trends in disability in a super-aging society: adapting the Future Elderly Model to Japan. J Econ Ageing. 2016;8:42-51.

7. Saksena M, Maldonado N. A dynamic estimation of obesity using NHANES Data: a pseudo-panel approach. Health Econ. 2017;26:e140-e159.

8. Roßmann J, Gummer T. Using paradata to predict and correct for panel attrition. Soc Sci Comp Rev. 2016;34:312-332.

9. Ministry of Health, Labour and Welfare. Outline of Vital Statistics in Japan. Available from: https://www.mhlw.go.jp/english/database/db-hw/outline/index.html. (Accessed August 31, 2018)

10. Wilmoth JR, Andreev K, Jdanov D, Glei DA, Boe C, Bubenheim M, et al. Methods protocol for the human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://www.mortality.org/ [version 31/05/2007]. 2007;9:10-11.

11. National Institute of Population and Social Security Research. Japanese Mortality Database Brief Summary. 2010. Available from: http://www.ipss.go.jp/p-toukei/JMD/briefsummary-en.html. (Accessed August 31, 2018)

12. Terblanche W, Wilson T. An evaluation of nearly-extinct cohort methods for estimating the very elderly populations of Australia and New Zealand. PloS one. 2015;10(4):e0123692.

13. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. J Health Soc Behav. 1997;38:21-37.

14. Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I epidemiologic follow-up study, 1992. Am J Epidemiol. 2000;152:874-883.

15. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self‐rated health question. J Gen Intern Med. 2006;21:267-275.

16. Ministry of Health, Labour and Welfare. Patient survey. 2014. Available from: http://www.mhlw.go.jp/english/database/db-hss/ps.html. (Accessed January 15, 2018).

17. Deaton A. Panel data from time series of cross-sections. J Econom. 1985;30:109-126.

18. Moffitt R. Identification and estimation of dynamic models with a time series of repeated cross-sections. J Econom. 1993;59:99-123.

19. Collado MD. Estimating dynamic models from time series of independent cross-sections. J Econom. 1997;82:37-62.

20. McKenzie DJ. Asymptotic theory for heterogeneous dynamic pseudo-panels. J Econom. 2004;120:235-262.

21. National Institute of Population and Social Security Research. Population Statistics of Japan. 2017. Available from: http://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/db_zenkoku2017/db_zenkoku2017syosaikekka.html. (Accessed August 31, 2018) (In Japanese)

22. Kubo M, Kiyohara Y, Kato I, Tanizaki Y, Arima H, Tanaka K, et al. Trends in the incidence, mortality, and survival rate of cardiovascular disease in a Japanese community: the Hisayama study. Stroke. 2003;34:2349-2354.

23. National Cancer Center. National estimates of cancer incidence based on cancer registries in Japan (1975-2013). Cancer Registry and Statistics. 2015. Available from: https://ganjoho.jp/en/professional/statistics/table_download.html. (Accessed December 2, 2017)

24. National Cancer Center. Monitoring of cancer incidence in Japan 2013. 2017. Available from: https://ganjoho.jp/reg_stat/statistics/brochure/monitoring.html. (Accessed December 2, 2017)

25. National Institute of Population and Social Security Research. Population projections for Japan: 2016-2065. Population research series No. 336. 2017. Available from: http://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/pp29_Report1.pdf. (Accessed December 2, 2017) (In Japanese)

26. Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. The Lancet. 2017;389(10076):1323-35.

27. Japanese Association of Cancer Registries. Cancer Registry in Japan second edition. 2010. Available from: http://www.jacr.info/publicication/document/CRIJ_eng.pdf. (Accessed August 31, 2018)

28. Brenner H, Holleczek B. 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. 2011.

29. Brenner H, Castro FA, Eberle A, Emrich K, Holleczek B, Katalinic A, et al. Death certificate only proportions should be age adjusted in studies comparing cancer survival across populations and over time. European Journal of Cancer. 2016;52:102-8.

30. Shimizutani S, Oshio T, Fujii M. Option value of work, health status, and retirement decisions in Japan: Evidence from the Japanese Study on Aging and Retirement (JSTAR). In: Wise DA, editor. Social security programs and retirement around the world: disability insurance programs and retirement. Chicago: University of Chicago Press; 2014. pp. 497-535.

31. Stowasser T, Heiss F, McFadden D, Winter J. "Healthy, Wealthy and Wise?" Revisited: An analysis of the causal pathways from socioeconomic status to Health. In: Wise DA, editor. Investigations in the economics of aging. Chicago: University of Chicago Press; 2011. pp. 267-317.

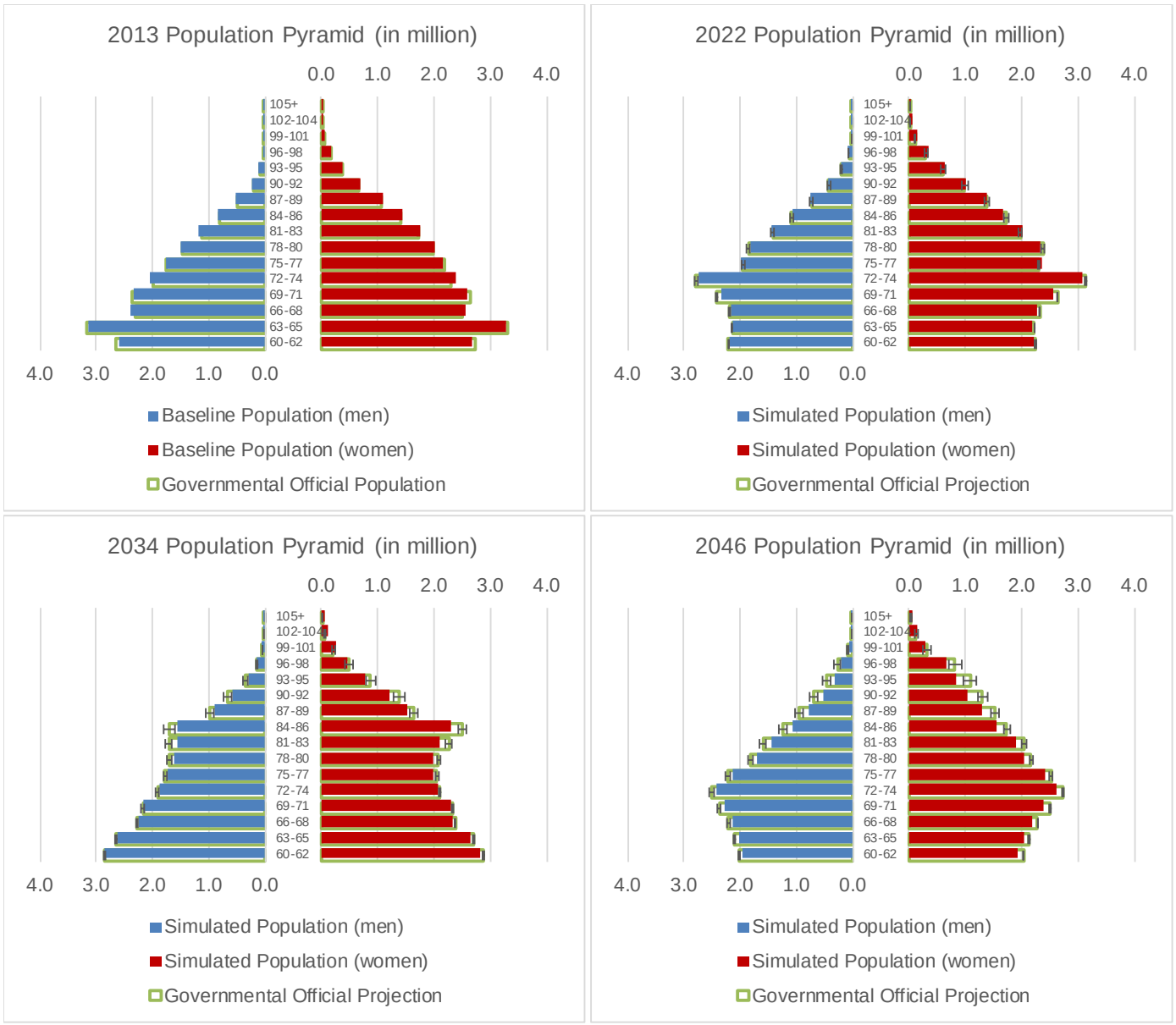
32. World Health Organization. Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health Final Report. World Health Organization. 2008;33.

33. Tang T, Kurashina S. 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. Stat Med. 1987;6:709-726.

34. Ma E, Takahashi H, Mizuno A, Okada M, Yamagishi K, Iso H. Stratified age-period-cohort analysis of stroke mortality in Japan, 1960 to 2000. J Stroke Cerebrovasc Dis. 2007;16:91-102.

35. Wang C, Weber A, Graham DY. Age, period, and cohort effects on gastric cancer mortality. Dig Dis Sci. 2015;60:514-523.

Fig 1. Simulated population pyramid compared with Japanese governmental official projections in 2013, 2022, 2034, and 2046.



Simulated population pyramid (solid line); Japanese governmental official projections (outlined in green); Japanese governmental projections with low-mortality and high-mortality assumptions (error bars).

Fig 2. Estimated age-specific annual incidence rates per 100 000 persons for coronary heart disease and stroke.

Fig 2(a) Heart disease

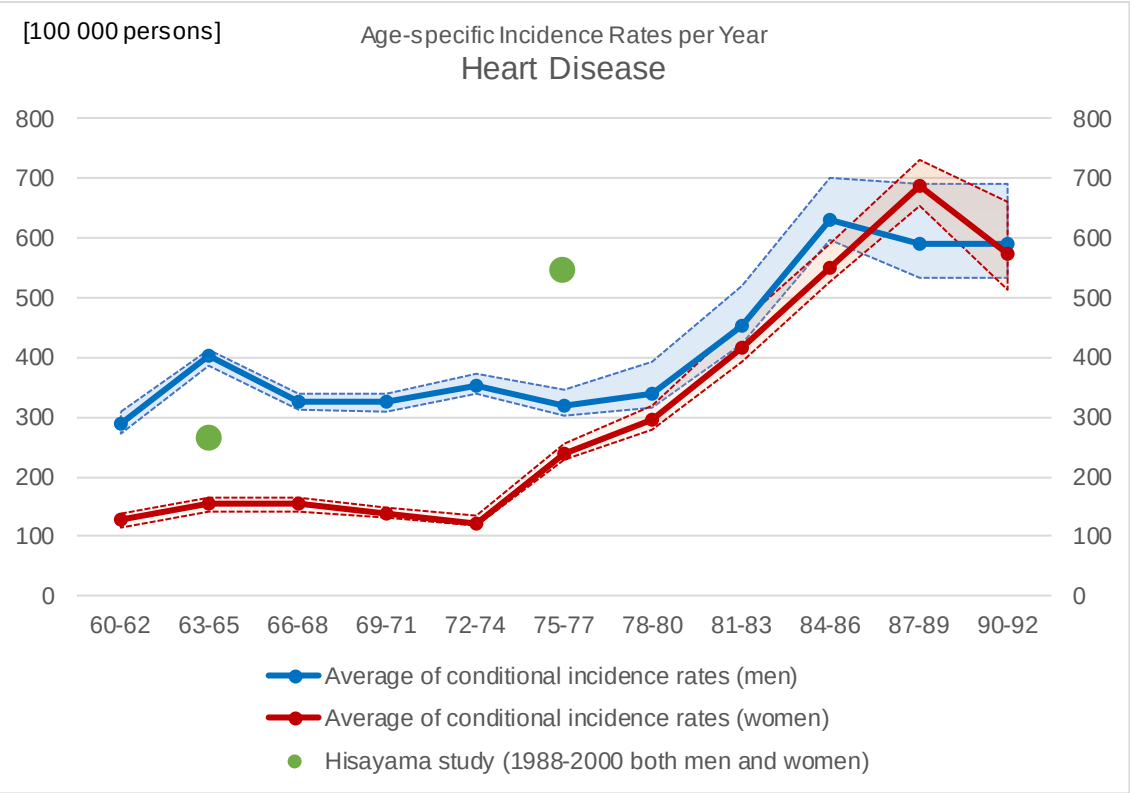
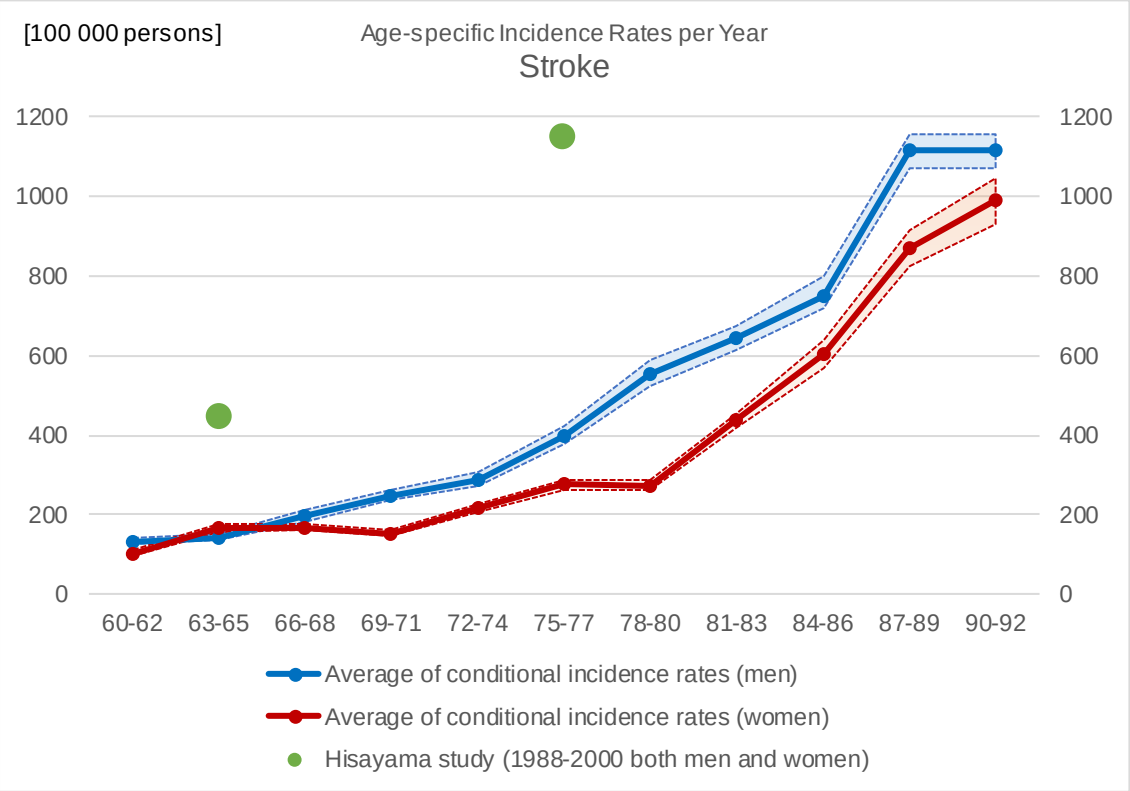
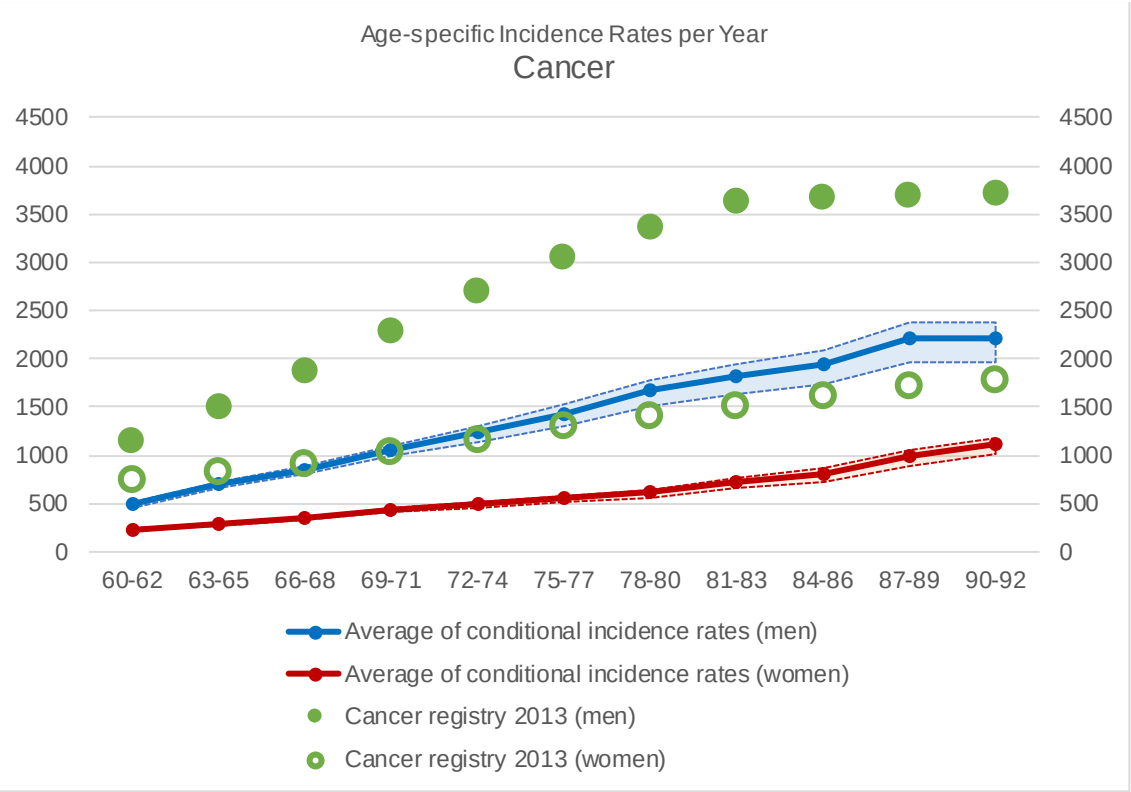


Fig 2(b) Stroke



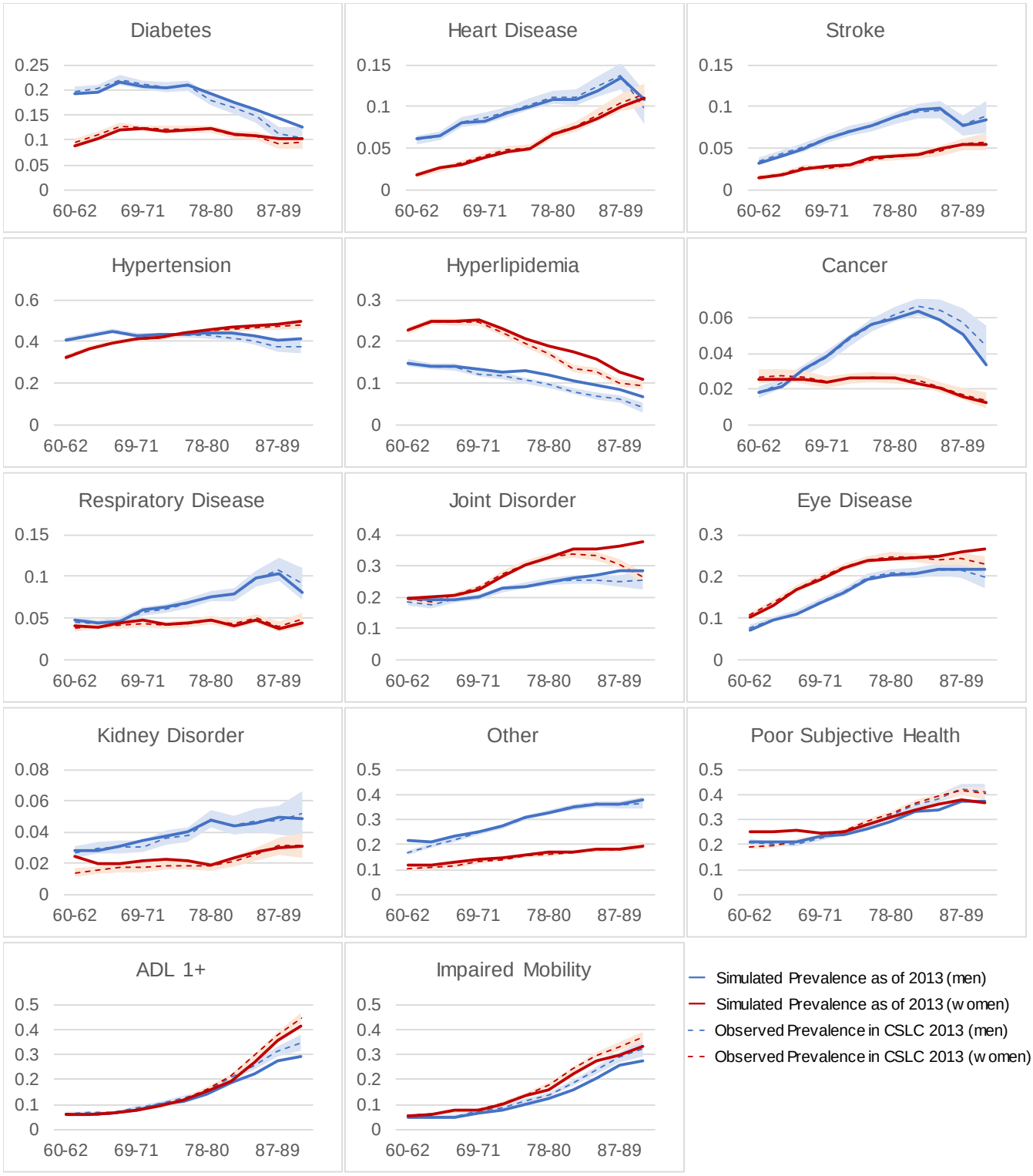
The range between the 5th and 95th percentiles is shadowed. Green plots indicate epidemiological observations derived from the reference [22].

Fig 3. Estimated age-specific annual incidence rates for cancer per 100 000 persons.



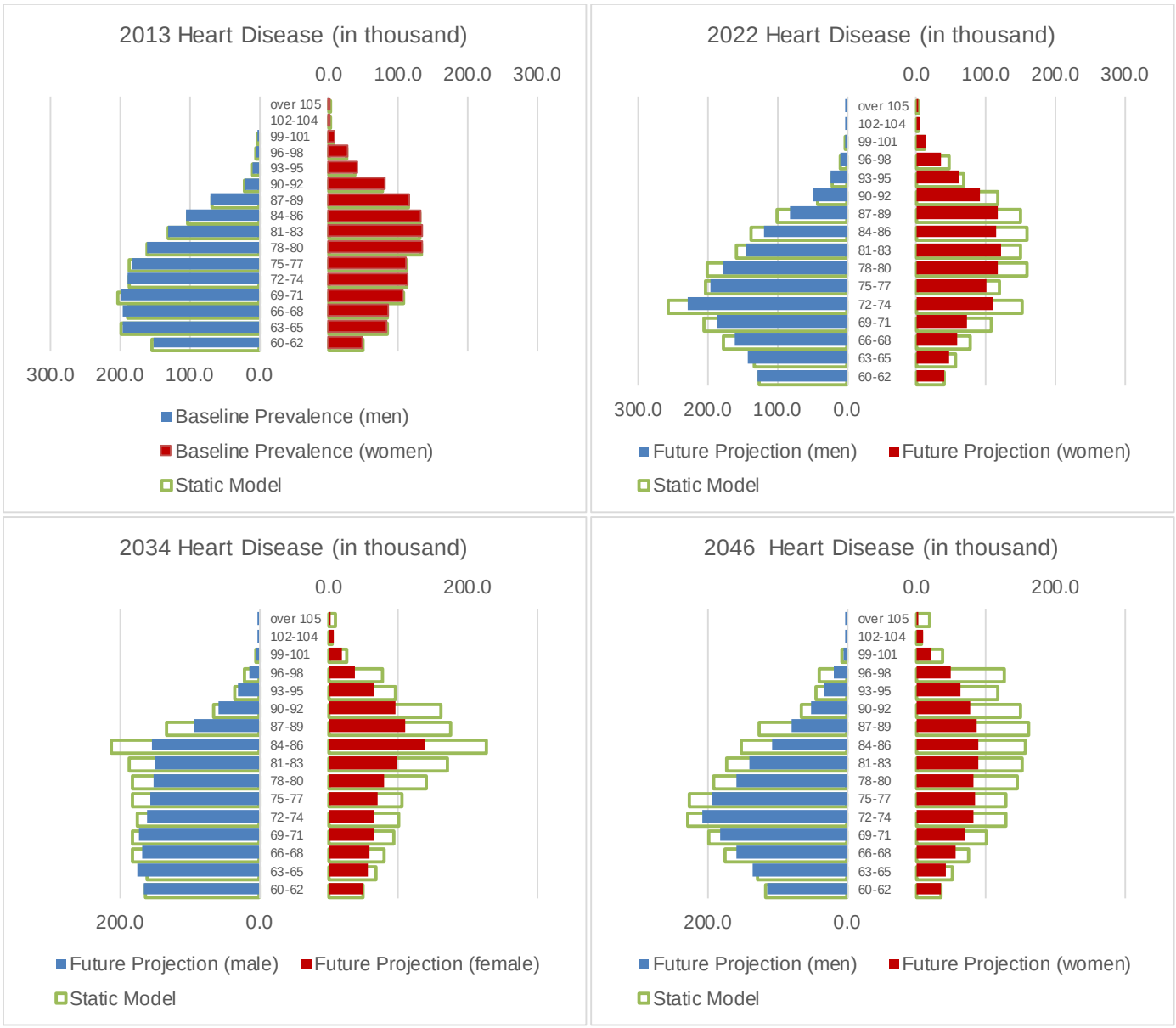
The range between the 5th and 95th percentiles is shadowed. Green plots indicate the numbers published in the Japanese national cancer registry (reference [23]).

Fig 4. Estimated prevalence of health statuses compared with observed statuses.



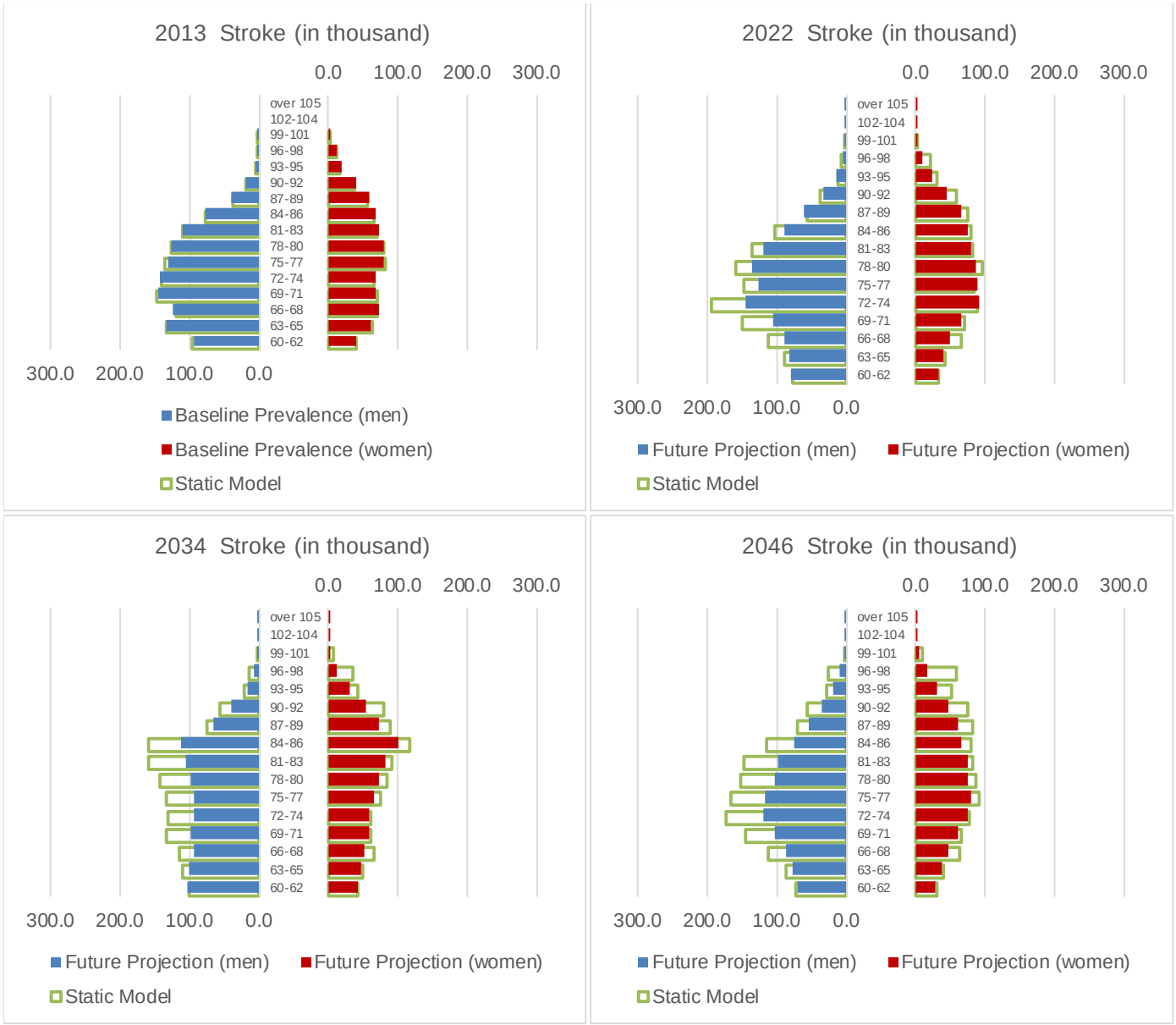
The solid line indicates estimations using observations from 2001 as a baseline. The dashed line indicates actually-observed data in the Comprehensive Survey of Living Conditions 2013. The range of 95% confidence intervals is shadowed. ADL 1+, at least one condition among dysfunctions in activities of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

Fig 5(a). Simulated prevalence for heart disease (2013–2046).



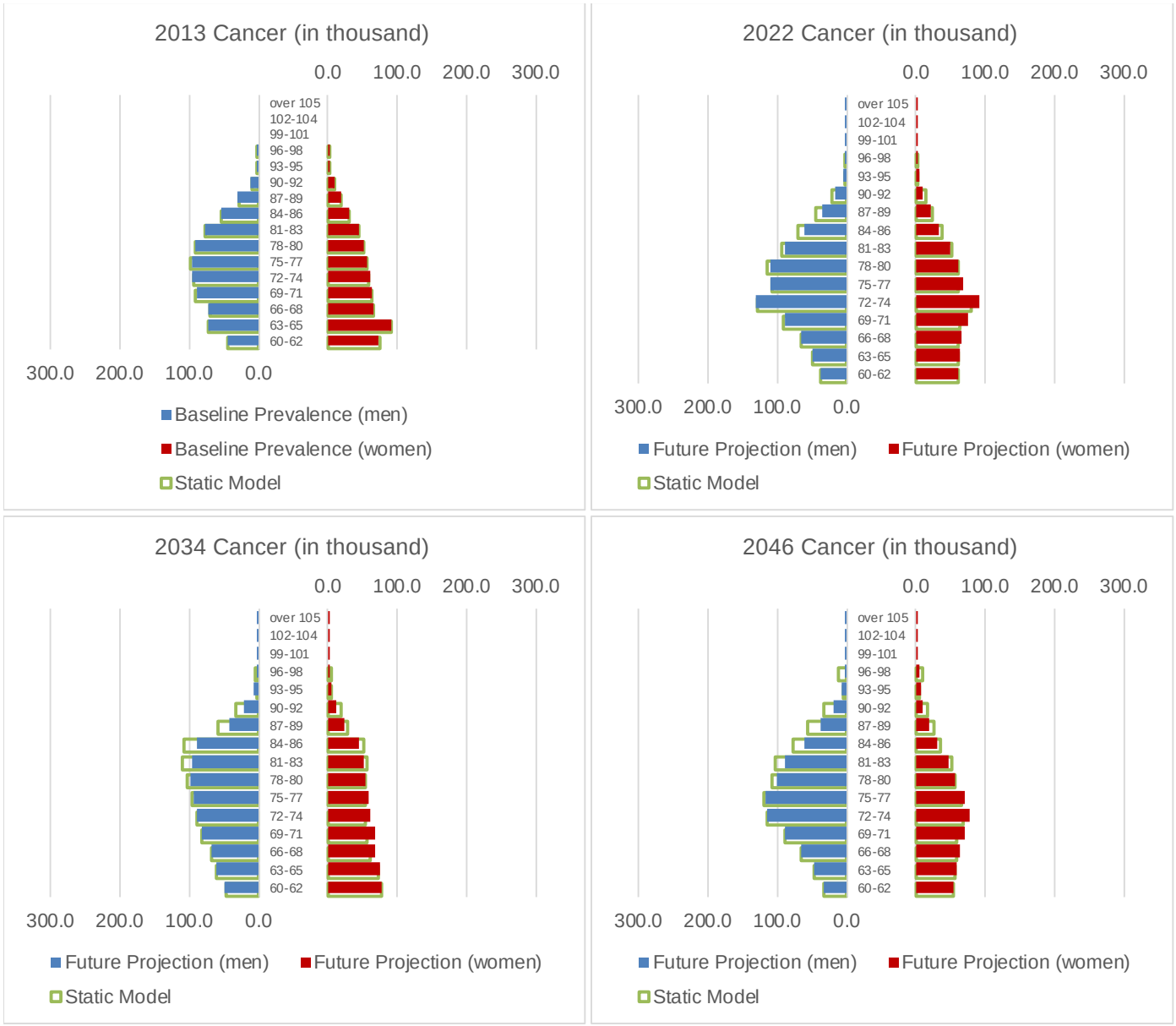
Blue bars describe the future prevalence for men, and red bars describe the future prevalence for women. The green lines represent estimates based on a static model.

Fig 5(b). Simulated prevalence for stroke (2013–2046).



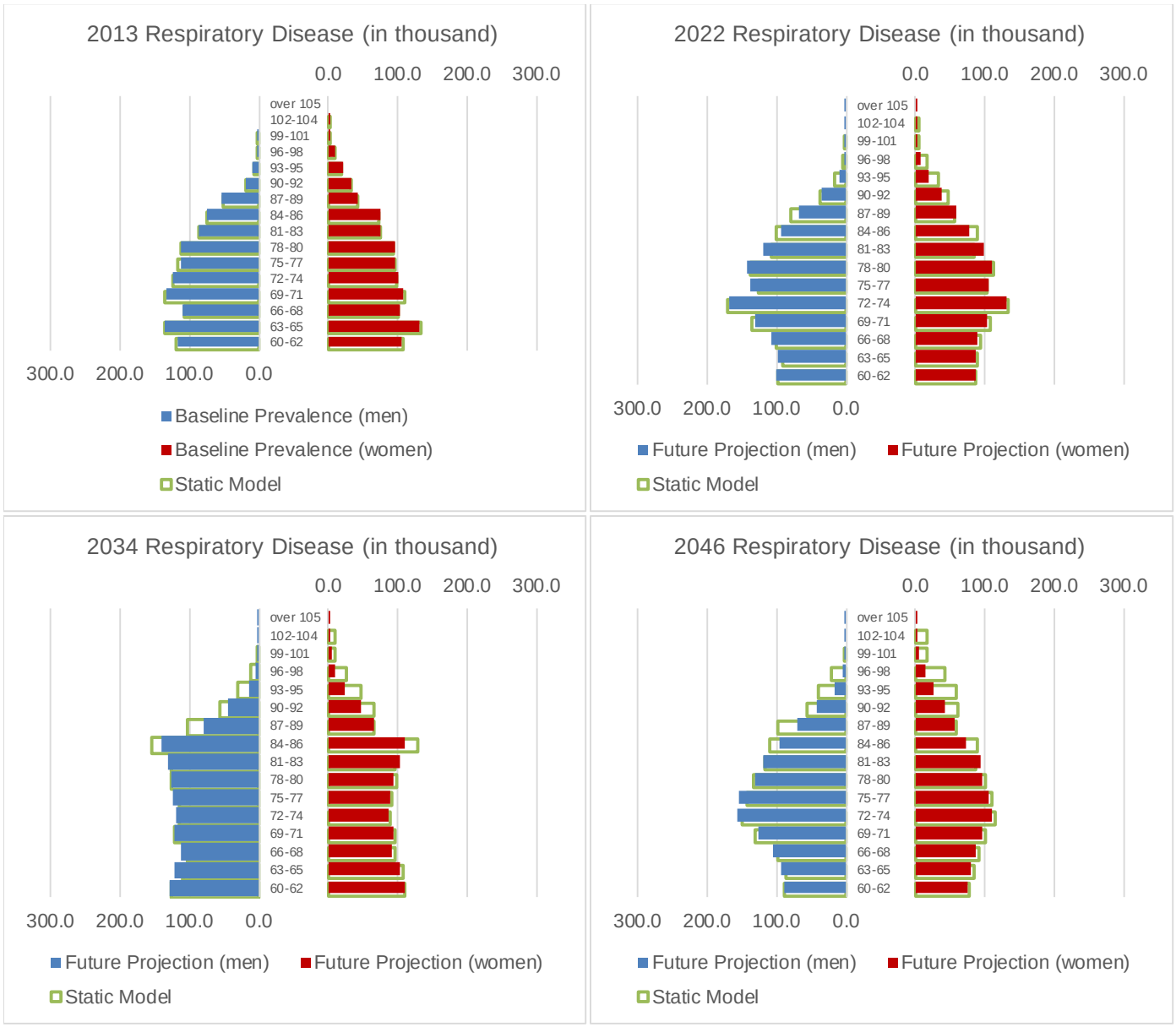
Blue bars describe the future prevalence for men, and the red bars describe the future prevalence for women. The green lines represent estimates based on a static model.

Fig 5(c). Simulated prevalence for cancer (2013–2046).



Blue bars describe the future prevalence for men, and the red bars describe the future prevalence for women. The green lines represent estimates based on a static model.

Fig 5(d). Simulated prevalence for respiratory disease (2013–2046).



Blue bars describe the future prevalence for men, and the red bars describe the future prevalence for women. The green lines represent estimates based on a static model.

Appendix Technical Document

Contingency table method for estimating conditional disease incidence with pseudopanel datasets

Summary

This technical note describes the details of our method to estimate conditional disease incidence probabilities with pseudopanel datasets using 2 × 2 contingency tables (Appendix Figure 1).

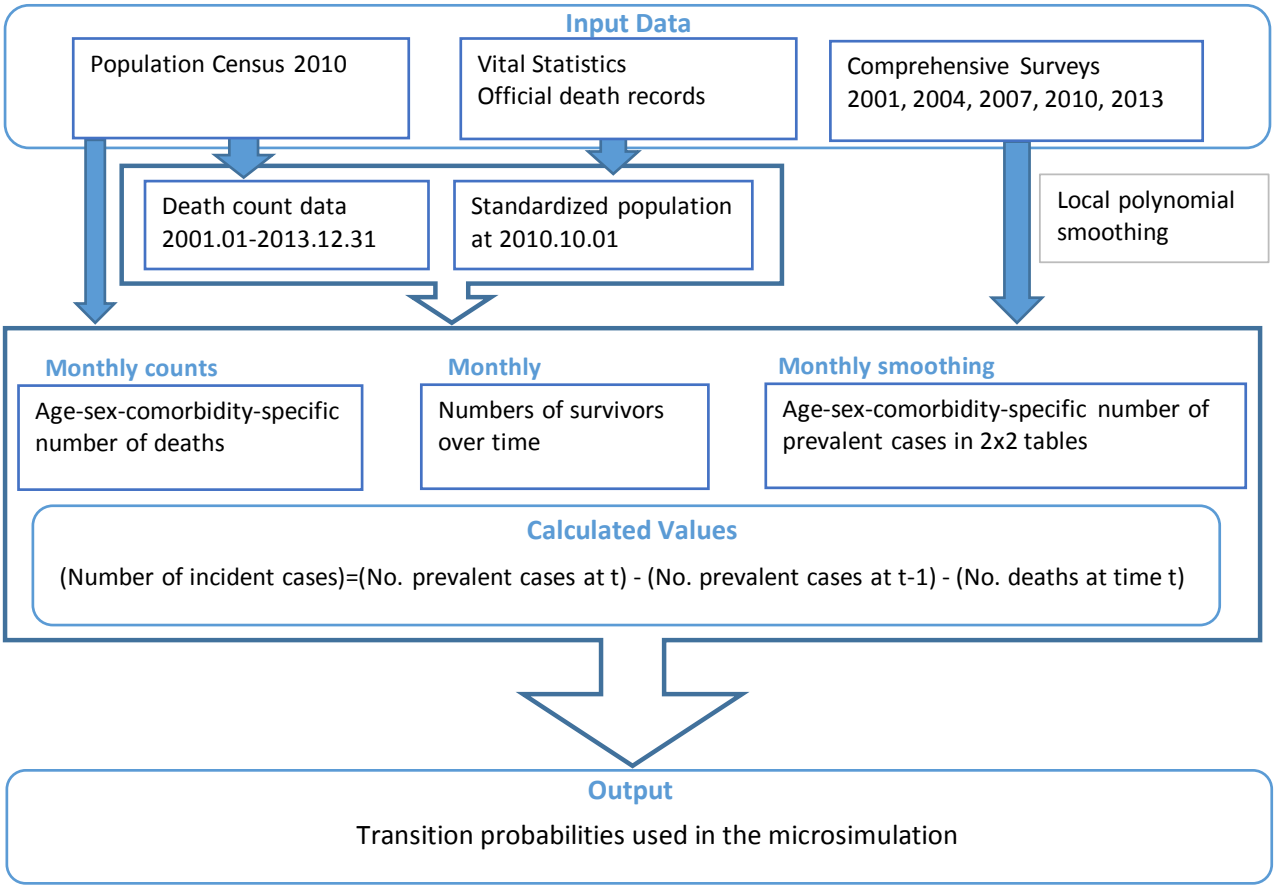
We use 2 × 2 tables, keeping exact numbers of frequencies standardized for sex-specific birth-cohort populations for all possible combinations of comorbidities as described in Section 1, which follows. An alternative method to the exact method is to estimate comorbidity status by generating a joint normal distribution based on the prevalence of comorbidity statuses of multiple diseases. Although this can be applied directly to panel data, its application to pseudopanel 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 Section 2).

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 Section 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.

Appendix Figure. Work flow of incidence determination using a contingency table method



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, and 19 birth cohorts (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.

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 d1 (diabetes in the following example) as α_1, the case fatality rate attributable to d2 (heart disease in the following example) as α_2, and the mortality rate for other conditions as α_{other}^{d1d2}.

Appendix Table 1. 2 × 2 tables of population and mortality rates

Table 1-(a)

	Population	d1: 0	d1: 1
d2	0	$pop_t(0,0)$	$pop_t(1,0)$
	1	$pop_t(0,1)$	$pop_t(1,1)$

Table 1-(b)

	Mortality rate	d1: 0	d1: 1
d2	0	α_{other}^{d1d2}	$\alpha_1 + \alpha_{other}^{d1d2}$
	1	$\alpha_2 + \alpha_{other}^{d1d2}$	$\alpha_1 + \alpha_2 + \alpha_{other}^{d1d2}$

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

(Observed mortality from diabetes in the vital statistics data) = $\alpha_1 * \{Pop_t(1,0) + Pop_t(1,1)\}$

(Observed mortality from heart disease in the vital statistics data) = $\alpha_2 * \{Pop_t(0,1) + Pop_t(1,1)\}$

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

$= \alpha_{other}^{d1d2} * \{Pop_t(0,0) + Pop_t(1,0) + Pop_t(0,1) + Pop_t(1,1)\}$

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.

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 2, we obtain the following four incidence numbers:

- Incidence of diabetes from $pop_t(0,0)$
- Incidence of heart disease from $pop_t(0,0)$
- Incidence of diabetes from $pop_t(0,1)$
- Incidence of heart disease from $pop_t(1,0)$

Appendix Table 2. Population flow (red arrows) because of disease incidence in equations (1)–(4) in the (d1,d2) 2 × 2 table

Eq(1)

		d1 = 0	d1 = 1
d2	0	$pop_t(0,0)$ →	$pop_t(1,0)$
	1	↓ $pop_t(0,1)$	$pop_t(1,1)$

Eq(2)

		d1 = 0	d1 = 1
d2	0	$pop_t(0,0)$	$pop_t(1,0)$ ↓
	1	$pop_t(0,1)$ →	$pop_t(1,1)$

Eq(3)

		d1 = 0	d1 = 1
d2	0	$pop_t(0,0)$ ↓	$pop_t(1,0)$
	1	$pop_t(0,1)$ →	$pop_t(1,1)$

Eq(4)

		d1 = 0	d1 = 1
d2	0	$pop_t(0,0)$ →	$pop_t(1,0)$ ↓
	1	$pop_t(0,1)$	$pop_t(1,1)$

The incidence numbers satisfy the following relationships if we assume closed cohorts:

<u>Population flow from [(d1,d2) = (0,0)]</u>

Incidence of heart disease from [(d1,d2) = (0,0)] + Incidence of diabetes from [(d1,d2) = (0,0)]

= Survivors in [(d1,d2) = (0,0)] − Prevalence of having the condition [(d1,d2) = (0,0)] at (t + 1)

$= pop_t(0,0) * (1 - \alpha_{other}^{d1d2}) - pop_{t+1}(0,0)$ Eq(1)

<u>Population flow into [(d1,d2) = (1,1)]</u>

Incidence of heart disease from [(d1,d2) = (1,0)] + Incidence of diabetes from [(d1,d2) = (0,1)]

= Prevalence of having the condition [(d1,d2) = (1,1)] at (t + 1) − Survivors in [(d1,d2) = (1,1)]

$= pop_{t+1}(1,1) - pop_t(1,1)*(1 - \alpha_{other}^{d1d2} - \alpha_1 - \alpha_2)$ Eq(2)

<u>Population flow in/out [(d1,d2) = (0,1)]</u>

Incidence of heart disease from [(d1,d2) = (0,0)] − Incidence of diabetes from [(d1,d2) = (0,1)]

= Prevalence of having the condition [(d1,d2) = (0,1)] at (t + 1) − Survivors in [(d1,d2) = (0,1)]

$= pop_{t+1}(0,1) - pop_t(0,1) \times (1 - \alpha_{other}^{d1d2} - \alpha_2)$ Eq(3)

<u>Population flow in/out [(d1,d2) = (1,0)]</u>

Incidence of diabetes from [(d1,d2) = (0,0)] − Incidence of heart disease from [(d1,d2) = (1,0)]

= Prevalence of having the condition [(d1,d2) = (1,0)] at (t + 1) − Survivors in [(d1,d2) = (1,0)]

$= pop_{t+1}(1,0) - pop_t(1,0) \times (1 - \alpha_{other}^{d1d2} - \alpha_1)$ Eq(4)

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 (d1) from [(d1,d2)=(0,0)] condition in the (d1,d2) − 2 × 2 table" by $inc_{d1}^{d1d2}(0,0)$, and denote "the incidence rate of heart disease (d2) from [(d1,d2) = (0,0)] condition in the (d1,d2) – 2 × 2 table" by $inc_{d2}^{d1d2}(0,0)$. Then we can rewrite Eq(1) as:

$(inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0)) = 1 - \frac{pop_{t+1}(0,0)}{pop_t(0,0) * (1 - \alpha_{other}^{d1d2})}$. Eq(1)′

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 3.

Appendix Table 3. Set of 26 Eq(1)′s to determine the incidence rates of diabetes (d1) and heart disease (d2)

Eq(1)′s including diabetes incidence	Eq(1)′s including heart disease (d2)
$(inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0))$	$(inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0))$
$(inc_{d1}^{d1d3}(0,0) + inc_{d3}^{d1d3}(0,0))$	$(inc_{d2}^{d2d3}(0,0) + inc_{d3}^{d2d3}(0,0))$
$(inc_{d1}^{d1d4}(0,0) + inc_{d4}^{d1d4}(0,0))$	$(inc_{d2}^{d2d4}(0,0) + inc_{d4}^{d2d4}(0,0))$
$(inc_{d1}^{d1d5}(0,0) + inc_{d5}^{d1d5}(0,0))$	$(inc_{d2}^{d2d5}(0,0) + inc_{d5}^{d2d5}(0,0))$
$(inc_{d1}^{d1d6}(0,0) + inc_{d6}^{d1d6}(0,0))$	$(inc_{d2}^{d2d6}(0,0) + inc_{d6}^{d2d6}(0,0))$
$(inc_{d1}^{d1d7}(0,0) + inc_{d7}^{d1d7}(0,0))$	$(inc_{d2}^{d2d7}(0,0) + inc_{d7}^{d2d7}(0,0))$
$(inc_{d1}^{d1d8}(0,0) + inc_{d8}^{d1d8}(0,0))$	$(inc_{d2}^{d2d8}(0,0) + inc_{d8}^{d2d8}(0,0))$
$(inc_{d1}^{d1d9}(0,0) + inc_{d9}^{d1d9}(0,0))$	$(inc_{d2}^{d2d9}(0,0) + inc_{d9}^{d2d9}(0,0))$
$(inc_{d1}^{d1d10}(0,0) + inc_{d10}^{d1d10}(0,0))$	$(inc_{d2}^{d2d10}(0,0) + inc_{d10}^{d2d10}(0,0))$
$(inc_{d1}^{d1d11}(0,0) + inc_{d11}^{d1d11}(0,0))$	$(inc_{d2}^{d2d11}(0,0) + inc_{d11}^{d2d11}(0,0))$
$(inc_{d1}^{d1d12}(0,0) + inc_{d12}^{d1d12}(0,0))$	$(inc_{d2}^{d2d12}(0,0) + inc_{d12}^{d2d12}(0,0))$
$(inc_{d1}^{d1d13}(0,0) + inc_{d13}^{d1d13}(0,0))$	$(inc_{d2}^{d2d13}(0,0) + inc_{d13}^{d2d13}(0,0))$
$(inc_{d1}^{d1d14}(0,0) + inc_{d14}^{d1d14}(0,0))$	$(inc_{d2}^{d2d14}(0,0) + inc_{d14}^{d2d14}(0,0))$

Arithmetically subtracting the sum of the elements in the right column from the sum of the elements in the left column in Appendix Table 3, we obtain the following equations:

$$(inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0)) + (inc_{d1}^{d1d3}(0,0) + inc_{d3}^{d1d3}(0,0)) + (inc_{d1}^{d1d4}(0,0) + inc_{d4}^{d1d4}(0,0))$$
$$+ (inc_{d1}^{d1d5}(0,0) + inc_{d5}^{d1d5}(0,0)) + (inc_{d1}^{d1d6}(0,0) + inc_{d6}^{d1d6}(0,0)) + (inc_{d1}^{d1d7}(0,0) + inc_{d7}^{d1d7}(0,0))$$
$$+ (inc_{d1}^{d1d8}(0,0) + inc_{d8}^{d1d8}(0,0)) + (inc_{d1}^{d1d9}(0,0) + inc_{d9}^{d1d9}(0,0)) + (inc_{d1}^{d1d10}(0,0) + inc_{d10}^{d1d10}(0,0))$$
$$+ (inc_{d1}^{d1d11}(0,0) + inc_{d11}^{d1d11}(0,0)) + (inc_{d1}^{d1d12}(0,0) + inc_{d12}^{d1d12}(0,0)) + (inc_{d1}^{d1d13}(0,0) + inc_{d13}^{d1d13}(0,0))$$
$$+ (inc_{d1}^{d1d14}(0,0) + inc_{d14}^{d1d14}(0,0)) - (inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0)) - (inc_{d2}^{d2d3}(0,0) + inc_{d3}^{d2d3}(0,0))$$
$$- (inc_{d2}^{d2d4}(0,0) + inc_{d4}^{d2d4}(0,0)) - (inc_{d2}^{d2d5}(0,0) + inc_{d5}^{d2d5}(0,0)) - (inc_{d2}^{d2d6}(0,0) + inc_{d6}^{d2d6}(0,0))$$
$$- (inc_{d2}^{d2d7}(0,0) + inc_{d7}^{d2d7}(0,0)) - (inc_{d2}^{d2d8}(0,0) + inc_{d8}^{d2d8}(0,0)) - (inc_{d2}^{d2d9}(0,0) + inc_{d9}^{d2d9}(0,0))$$
$$- (inc_{d2}^{d2d10}(0,0) + inc_{d10}^{d2d10}(0,0)) - (inc_{d2}^{d2d11}(0,0) + inc_{d11}^{d2d11}(0,0)) - (inc_{d2}^{d2d12}(0,0) + inc_{d12}^{d2d12}(0,0))$$
$$- (inc_{d2}^{d2d13}(0,0) + inc_{d13}^{d2d13}(0,0)) - (inc_{d2}^{d2d14}(0,0) + inc_{d14}^{d2d14}(0,0))$$
$$= (inc_{d1}^{d1d3}(0,0) + inc_{d1}^{d1d4}(0,0) + inc_{d1}^{d1d5}(0,0) + inc_{d1}^{d1d6}(0,0) + inc_{d1}^{d1d7}(0,0) + inc_{d1}^{d1d8}(0,0)$$

$+ inc_{d1}^{d1d9}(0,0) + inc_{d1}^{d1d10}(0,0) + inc_{d1}^{d1d11}(0,0) + inc_{d1}^{d1d12}(0,0) + inc_{d1}^{d1d13}(0,0) + inc_{d1}^{d1d14}(0,0))$
$- (inc_{d2}^{d2d3}(0,0) + inc_{d2}^{d2d4}(0,0) + inc_{d2}^{d2d5}(0,0) + inc_{d2}^{d2d6}(0,0) + inc_{d2}^{d2d7}(0,0) + inc_{d2}^{d2d8}(0,0)$
$+ inc_{d2}^{d2d9}(0,0) + inc_{d2}^{d2d10}(0,0) + inc_{d2}^{d2d11}(0,0) + inc_{d2}^{d2d12}(0,0) + inc_{d2}^{d2d13}(0,0) + inc_{d2}^{d2d14}(0,0))$
$= 12 (inc_{d1}(0) - inc_{d2}(0))$

where $inc_{d1}(0)$ denotes the equilibrium of the incidence rate of diabetes from disease-free conditions, and $inc_{d2}(0)$ denotes the equilibrium of the incidence rate of heart disease from disease-free conditions.

The first row of Appendix Table 3 can be approximated by $inc_{d1}^{d1d2}(0,0) + inc_{d2}^{d1d2}(0,0) = inc_{d1}(0) + inc_{d2}(0)$, which determines the solutions for Eq(1)', $inc_{d1}^{d1d2}(0,0)$, and $inc_{d2}^{d1d2}(0,0)$. The solution to Eq(1)' sequentially provides the remainder of the solutions in the system of equations (1)–(4), $inc_{d1}^{d1d2}(0,1)$, and $inc_{d2}^{d1d2}(1,0)$.

Population adjustment in demographics

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 (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.

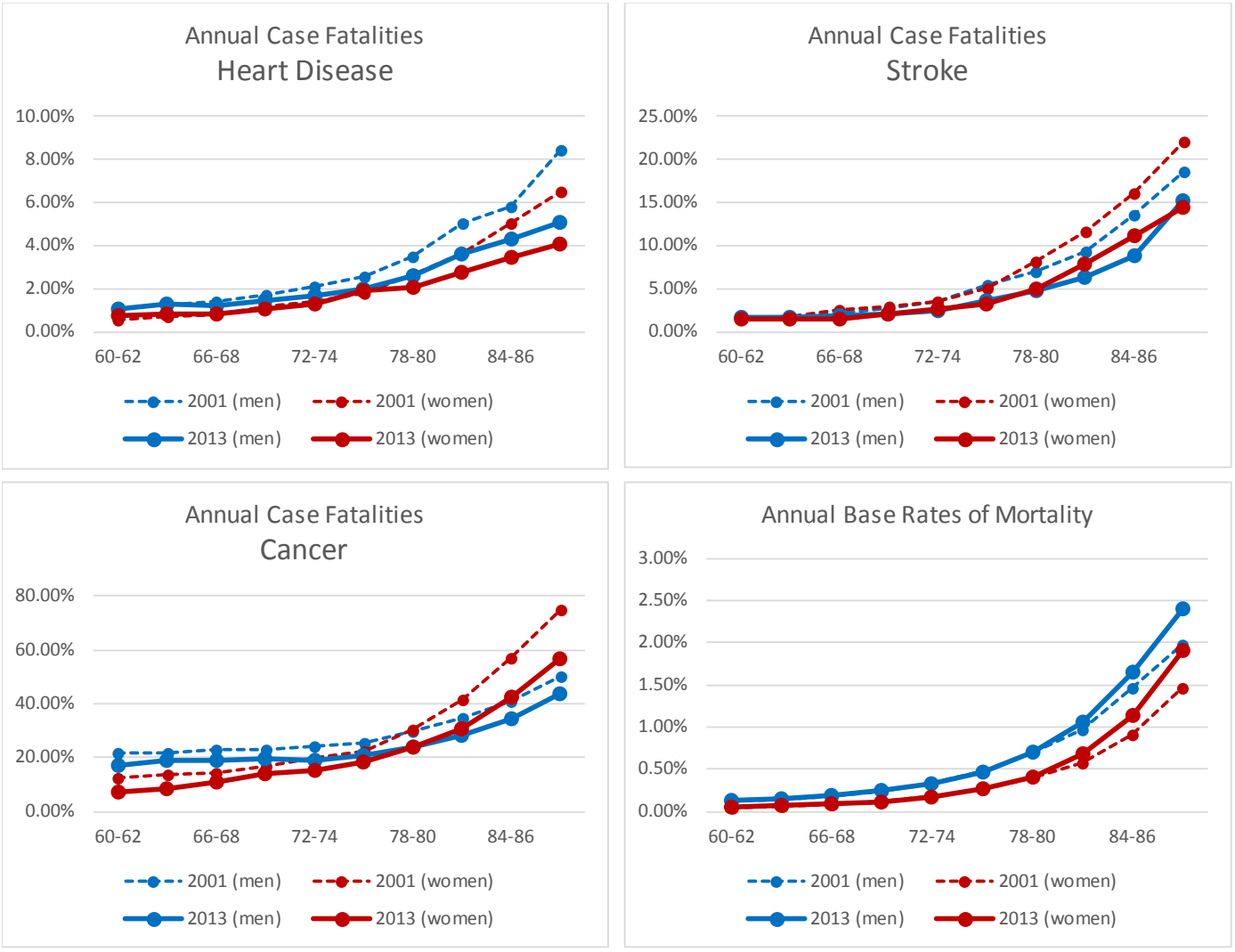
Reference

Wilmoth JR, Andreev K, Jdanov D, Glei DA, Boe C, Bubenheim M, et al. Methods protocol for the human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://www.mortality.org/ [version 31/05/2007]. 2007;9:10-11.

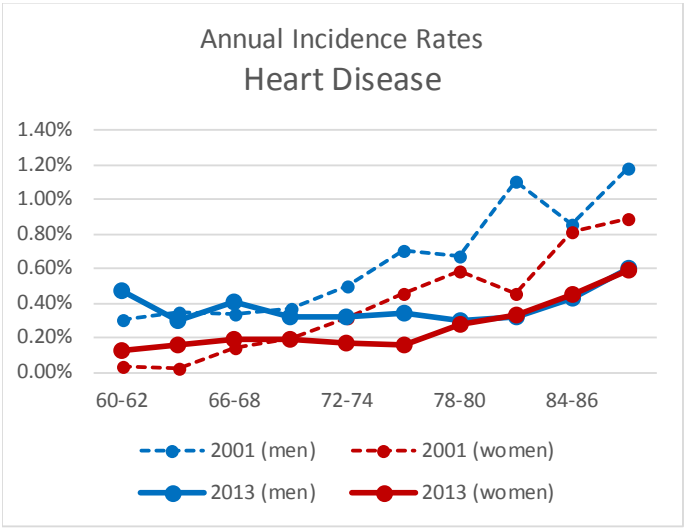
Supplemental figures describing the results

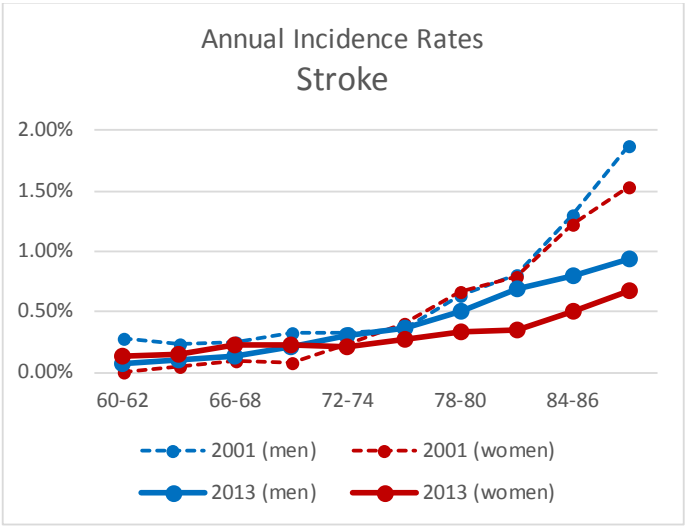
Parameters for backward validation delivered time trends for age-sex-specific incidence and mortality. We observed a gradual decline in disease case fatalities for all diseases in our observation period, while the baseline mortality rates increased slightly at ≥ 80 years of age (Appendix Fig. 1). Incidence 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 incidences of these diseases over years (Appendix Fig. 2). Despite decreasing incidence trends for mortality and most diseases, the prevalence of chronic conditions will increase in older Japanese populations because of extremely rapid population aging.

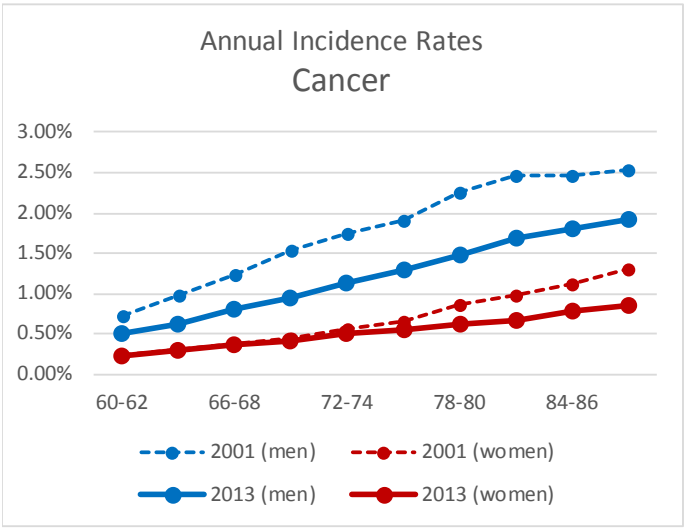
Appendix Figure 1. Estimated case fatality in percentiles for 2001 and 2013



Appendix Figure 2. Estimated incidence rates in percentiles for 2001 and 2013

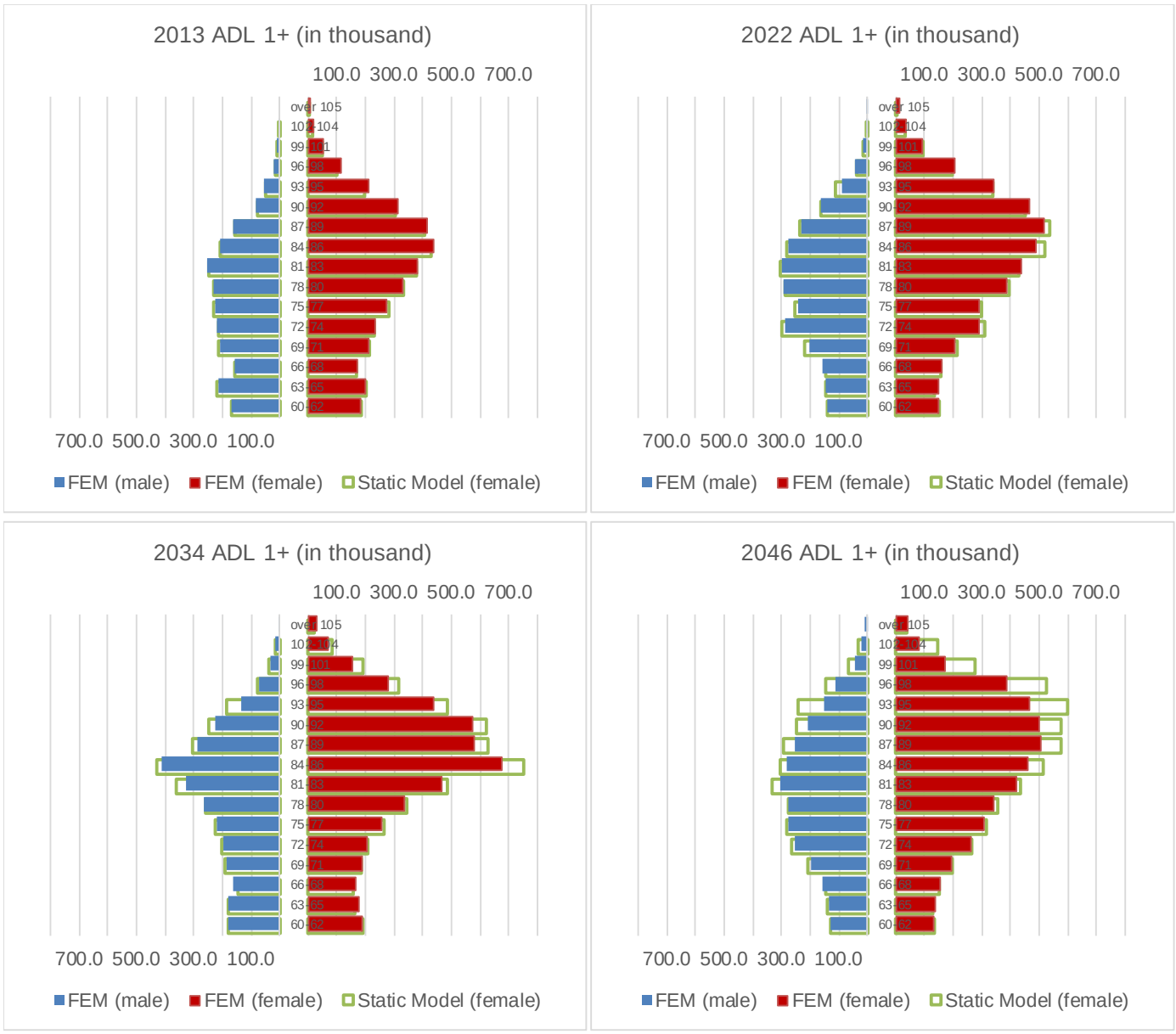






The prevalence of difficulties with activities of daily living and mobility concomitantly declined with trends for stroke prevalence. Our model predicted a 7.5-million activities of daily living dysfunction prevalence in 2034, which is a 1.8 million increase from 2013, while the static model predicted 8.1 million in 2034 (Appendix Fig. 3). Our model forecasted a 7.0-million prevalence for mobility dysfunction in 2034, which is an increase of 1.5 million from 2013 (Appendix Fig. 4).

Appendix Figure 3. Simulated prevalence for difficulties in activities of daily living (2013–2046)



Note: ADL 1+, at least one condition among dysfunctions in activities of daily living.

Appendix Figure 4. Simulated prevalence for impaired mobility (2013–2046)