

Sampling strategies for selecting general population comparison cohorts

Uffe Heide-Jørgensen
Kasper Adelborg
Johnny Kahlert
Henrik Toft Sørensen
Lars Pedersen

Department of Clinical Epidemiology,
Aarhus University Hospital, Aarhus,
Denmark

Background: For a patient cohort, access to linkable population-based registries permits sampling of a comparison cohort from the general population, thereby contributing to the understanding of the disease in a population context. However, sampling without replacement in random order can lead to immortal time bias by conditioning on the future.

Aim: We compared the following strategies for sampling comparison cohorts in matched cohort studies with respect to time to ischemic stroke and mortality: sampling without replacement in random order; sampling with replacement; and sampling without replacement in chronological order.

Methods: We constructed index cohorts of individuals from the Danish general population with no particular trait, except being alive and without ischemic stroke on the index date. We also constructed index cohorts of persons aged >50 years from the general population. We then applied the sampling strategies to sample comparison cohorts (5:1 or 1:1) from the Danish general population and compared outcome risks between the index and comparison cohorts. Finally, we sampled comparison cohorts for a heart failure cohort using each strategy.

Results: We observed increased outcome risks in comparison cohorts sampled 5:1 without replacement in random order compared to the index cohorts. However, these increases were minuscule unless index persons were aged >50 years. In this setting, sampling without replacement in chronological order failed to sample a sufficient number of comparators, and the mortality risks in these comparison cohorts were lower than in the index cohorts. Sampling 1:1 showed no systematic difference between comparison and index cohorts. When we sampled comparison cohorts for the heart failure patients, we observed a pattern similar to when index persons were aged >50 years.

Conclusion: When index persons were aged >50 years, ie, had high outcome risks, sampling 5:1 without replacement introduced bias. Sampling with replacement or 1:1 did not introduce bias.

Keywords: matched cohort study, survival analysis, population-based registry, observational study

Background

The cohort study is a classic design in epidemiological studies of disease risk or prognosis.¹ Life tables, often used to compute sex- and age-standardized incidence and mortality ratios in cohort studies, can contribute in understanding diseases in a population context.¹ However, life table-based designs are inadequate when more detailed confounder adjustment is required if data on the additional variables is unavailable in the life tables.¹

In epidemiological studies, population registries available in some European countries are used widely to sample and follow comparison cohorts representing the general

Correspondence: Uffe Heide-Jørgensen
Department of Clinical Epidemiology,
Aarhus University Hospital, Olof
Palmes Allé 43-45, DK-8200,
Aarhus N, Denmark
Tel +45 8716 8221
Fax +45 8716 7215
Email uhj@clin.au.dk

population²⁻⁴ based on risk-set sampling.¹ The comparison cohort can be matched to the patient cohort on variables including sex, birth year, and calendar period, ensuring that comparators are alive on the patient's index date.⁵⁻¹⁹ Notice that matching on birth year and calendar period implies matching on age.

Sampling comparison cohorts from the general population provides a more flexible study design than studies relying on population-based life tables,^{12,13} because individuals in the comparison cohort can be assigned the same index date as the patient to whom they are matched. This allows for assessment of characteristics in both the patient and comparison cohorts through linkage to other health care, lifestyle, or socioeconomic registries.^{2,3}

Several strategies can be used to sample a general population comparison cohort for a patient cohort (or other types of index cohorts). One approach is to sample each comparator only once, ie, to sample without replacement,²⁰ avoiding dependency between observations. However, if sampling is done in random (non-chronological) order relative to the order of the index dates of the patients, sampling without replacement can lead to immortal time bias²¹ by conditioning on the future. As an example, consider a study in which it is required to sample a comparison cohort for a patient cohort, matching on sex, birth year, and calendar period. Now consider two patients, p_1 and p_2 , with the same sex and birth year, diagnosed on dates, d_1 and d_2 , respectively, with p_1 being diagnosed before p_2 . If comparators are sampled for p_2 first, then the selected individuals are ineligible to be sampled as comparators for p_1 on d_1 because they survived until d_2 , introducing immortal time bias. As the comparators sampled for p_2 are unavailable as comparators for p_1 , the proportion of potential comparators for p_1 who die before d_2 increase after comparators are sampled for p_2 . Hence, comparators sampled for p_1 are drawn from a set of individuals, who on average die at an earlier date than they would, had we not sampled comparators for p_2 first.

It is possible to circumvent the immortal time bias by sampling with replacement²⁰ or by sampling without replacement in chronological order, ie, given any two patients of the same sex and birth year, p_i and p_j , such that p_i is diagnosed before p_j , the comparators should be selected for p_i before p_j .

In this study, we compare the following three sampling strategies for selecting comparison cohorts, matching on sex, birth year, and calendar period: 1) sampling without replacement in random order, 2) sampling with replacement, and 3) sampling without replacement in chronological order. We compare the strategies with respect to balancing matching

factors, and with respect to time to ischemic stroke and mortality in the sampled comparison cohorts.

We hypothesized that sampling without replacement in random order would produce comparison cohorts in which the outcome rates would be higher than in the general population because of immortal time bias (an illustration of the different sampling strategies is available in the [Supplementary materials](#)).

Methods

Setting and data sources

This study was conducted in Denmark using two nationwide population-based registries: the Civil Registration System (CRS) and the Danish National Patient Registry (DNPR). The CRS contains virtually complete and highly accurate individual-level information on all residents in Denmark, including birth date, sex, vital status, and sequential dates of migration.² Since its establishment in 1968, the CRS has registered cumulatively more than 9.5 million individuals with more than 250 million person-years of follow-up.² CRS data have been used extensively in epidemiological studies, eg, to assess all-cause mortality and to sample comparison cohorts in cohort studies and controls in case-control studies.²

The DNPR has recorded every inpatient hospitalization since 1977 and every outpatient and emergency room visit since 1995.²² For each admission, one primary and potentially several secondary diagnoses are registered, classified according to the International Classification of Diseases, Eighth Revision (ICD-8) until the end of 1993 and Tenth Revision (ICD-10) thereafter.²² Data from the CRS and DNPR can be linked on an individual level using a unique personal identifier assigned to all Danish residents at birth or upon immigration.^{2,22}

Study outline

We undertook both a general population cohort study and a heart failure cohort study.

In the general population cohort study, we constructed index cohorts of individuals from the general population and then sampled comparison cohorts using the three different sampling strategies. Our premise was that if a valid sampling strategy was applied, we would observe no systematic difference between the comparison and the index cohorts with respect to any characteristic. For example, the comparison and index cohorts should be similar with respect to distribution of baseline characteristics, risk and rate differences of outcomes should be zero, and correspondingly, risk and rate ratios should be one.

In the heart failure cohort study, we selected a heart failure patient cohort and sampled comparison cohorts to assess whether choice of sampling strategy affected the associations between heart failure and outcome risks and rates.

In both studies we matched on sex, birth year, and calendar period and followed the cohorts until incident ischemic stroke or death to compute the rates of these outcomes. We selected heart failure as a disease example because it is common among the elderly, allowing us to construct a large patient cohort.⁵ Heart failure is also associated with a poor prognosis with respect to both survival and complications,²³ increasing the likelihood of detecting differences between the strategies. We selected ischemic stroke as an outcome, because it is a common reason for acute hospitalization in the adult population,²⁴ ie, detecting potential differences in risks and rates between cohorts is therefore more likely than if we had chosen a rare disease as an outcome. The study period was from January 1, 1980 to August 31, 2012.

Cohorts in the general population cohort study

We selected a random sample of 50,000 individuals from the CRS, who were alive in Denmark at some point during the study period, as a source population for the index and comparison cohorts. All Danish residents had the same probability of being sampled. We then constructed an index cohort within the sample as follows: on 1st January of each year in the study period, we recruited individual persons in the source population, who were alive and without ischemic stroke on that date, into the index cohort with a probability of 0.1%. We identified ischemic stroke using primary and secondary diagnoses in all available inpatient and outpatient records in the DNPR (diagnostic codes provided in the [Supplementary materials](#)).

After the index cohort was constructed, we sampled comparison cohorts from the source population, using the three different sampling strategies (Figure 1). Specifically, we sorted the members of the index cohort (referred to as ‘index persons’) either using a random variable drawn from a uniform distribution or, when sampling without replacement in chronological order, by their index dates. We then sampled individuals of the same sex and birth year, who were alive on the index date of the index person and removed the sampled individuals from the sampling pool except when sampling with replacement. Individuals were ineligible to be sampled as comparators on a given date if they had suffered an ischemic stroke or were already included as index persons

before that date. We sampled five comparators for each index person, or as many as possible if less than five were eligible.

Because index persons were characterized by no particular trait except being alive and without ischemic stroke on the index date, we considered the rates of ischemic stroke and mortality observed in the index cohort as true rates against which we could compare the rates observed in the matched comparison cohorts.

For a second analysis, we constructed index cohorts of persons aged >50 years from the general population. We did this by making the probability of being recruited into the index cohorts age-dependent. Thus, we set the probability of being recruited to 0 for those aged 50 years or younger on 1st January in a given year and $[(\text{age in years} - 50)/50]\%$ for those older than 50 years; for example, someone aged 60 years would have a 0.2% probability of being recruited. Again, using the three sampling strategies, we sampled five comparators for each index person, or as many as possible if less than five were eligible, and persons with prevalent ischemic stroke at the index date were ineligible to be included in the index and comparison cohorts.

We repeated this 1,000 times to obtain a measure of the variation of the outcome rates in the different index and comparison cohorts, and we used a new random sample of 50,000 individuals in each iteration.

Cohorts in the heart failure cohort study

To assess whether the sampling strategies would produce different results for a specific patient cohort, we selected a cohort of incident heart failure patients and sampled three comparison cohorts using the different sampling strategies.

To identify the cohort of patients with heart failure, we used all primary and secondary diagnoses in the DNPR inpatient and outpatient records from January 1, 1977, to August 31, 2012. To prevent inclusion of prevalent heart failure patients, we excluded all patients with a diagnosis of heart failure from 1977 through 1979, thus starting the study in 1980. We also excluded heart failure patients with a previous diagnosis of ischemic stroke and patients not living in Denmark at the time of their incident heart failure diagnosis. The remaining patients were included in the study, with the date of their first heart failure diagnosis serving as the index date (Figure 1).

Using each of the three sampling strategies, we sampled five comparators from the CRS for each heart failure patient, or as many as possible if less than five were eligible. Individuals were ineligible to be selected as comparators if they had

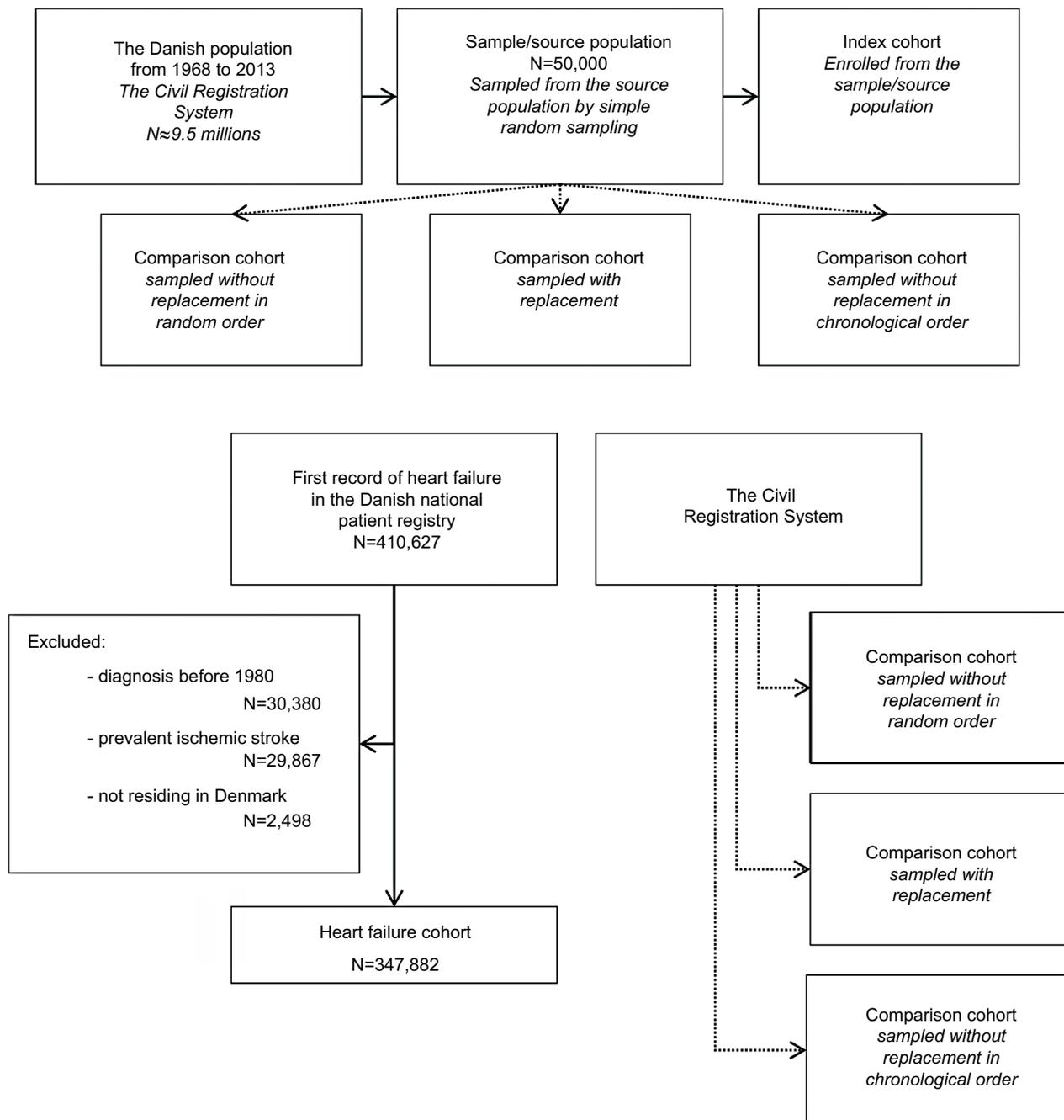


Figure 1 Selection of the index and comparison cohorts within one iteration (top), and selection of the heart failure cohort and comparison cohorts (bottom).
Note: The comparison cohorts were matched to the index/heart failure cohort on sex, birth year, and calendar period.

a prevalent diagnosis of heart failure or ischemic stroke on the index date of the heart failure patient.

Follow-up and statistical analyses

We followed members of each cohort for a maximum of 10 years from the index date, until occurrence of ischemic stroke, death, emigration, or August 31, 2012, whichever came first. It was possible for a person in a comparison

cohort to become an index person/heart failure patient during follow-up. In this case, the person was enrolled into the index/heart failure cohort; however, the person was not censored from the relevant comparison cohort on the date of becoming an index person/heart failure patient, imitating an intention-to-treat type of analysis.²⁵ Had we chosen to censor comparators who became index persons during follow-up in the general population study, we would expect to see no

difference in outcome rates as the censoring would be non-informative. In the heart failure study, censoring comparators who developed heart failure during follow-up would likely decrease the outcome rates in the comparison cohorts as the censoring would be informative, and we would expect the censored comparators to be at a higher risk of outcomes than the non-censored comparators.

We recorded data on the size of individual cohorts, median age and proportion of males, and number of comparators who became index persons/heart failure patients during follow-up in the respective comparison cohorts. We also computed duration of follow-up, number of persons who developed ischemic stroke or died during follow-up, rates of ischemic stroke and mortality, and the ratio of unique individuals to cohort size in the comparison cohorts sampled with replacement.

In the general population cohort study, we summarized these statistics by their median and interquartile range across all 1,000 iterations.

We further estimated the cumulative incidence²⁶ of ischemic stroke and death in each cohort, treating one outcome as a competing event to the other. In the general population cohort study, we used the cumulative incidences to estimate risk differences of ischemic stroke and death between the comparison and the index cohorts for every month of follow-up. Based on the 1,000 iterations, we computed the 2.5th, 5th, 10th, 25th, 50th (median), 75th, 90th, 95th, and 97.5th percentiles of the risk differences and plotted medians, along with bands of symmetric percentiles, against time, for each sampling strategy. Similarly, in the general population cohort study, we computed hazard ratios of ischemic stroke and death using the index cohorts as the reference and presented the log-hazard ratios by box plots. If a valid strategy was used, log-hazard ratios should be normally distributed around zero, while hazard ratios should be log-normally distributed around one.

In the heart failure cohort study, we also computed hazard ratios of each outcome comparing the heart failure cohort to each comparison cohort. However, as the proportionality assumption was violated, we instead presented incidence rate ratios. We also recorded the number of matches sampled for each heart failure patient and computed incidence rate ratios of ischemic stroke and death after restricting the analysis to patients for whom the individual sampling strategy identified five comparators.

Sensitivity analyses

In the general population study we conducted two sensitivity analyses. First, we sampled one rather than five comparators

for each index person in the index cohorts of persons aged >50 years. Second, we constructed index cohorts by recruiting anyone alive on 1st January in a given year with a chance of 0.5% rather than 0.1%.

In the heart failure cohort study, we also performed a sensitivity analysis by sampling one to one.

All analyses were carried out using SAS version 9.4 (Cary, NC, USA). The study was approved by the Danish Data Protection Agency (record number: 1-16-02-268-14). No approval from an ethics committee or informed consent from patients is required for registry studies in Denmark.

Results

General population cohort study

When anyone in the source population was recruited into the index cohort with a probability of 0.1%, the median size of the index cohorts was 1,000 persons, and the comparison cohorts were approximately five times larger. Generally, the median age and proportion of males were similar in the index and comparison cohorts (Table 1). This indicates that all three sampling strategies were successful in constructing comparison cohorts that resembled the index cohorts with respect to the matching factors.

When we compared comparison cohorts sampled without replacement in random order to the index cohorts, we observed a consistently positive median risk difference of death (Figure 2). However, the magnitude of the median risk difference was minuscule throughout the follow-up and peaked at a value of just 0.15% (data not shown). Thus, overall we found no substantial difference between the index cohorts and the comparison cohorts sampled without replacement in random order (Table 1, Figures 2 and 3).

We found no systematic deviation from the index cohorts with respect to any outcome measure for either of the other two sampling strategies (Table 1, Figures 2 and 3).

When we constructed index cohorts of people aged >50 years, the median size of the index cohort was almost 1,000 persons, there was a minority of males, and the median age was in the early seventies (Table 2). In the comparison cohorts, we were generally unable to find a sufficient number of matches, particularly when sampling without replacement in chronological order (Table 2). We found that sampling without replacement in random order led to increased outcome rates compared to the index cohorts resulting in positive median risk differences and log-hazard ratios (Table 2, Figures 4 and 5). Furthermore, we observed negative risk differences of death when we compared comparison cohorts sampled without replacement in chronological order to the index cohorts (Figure 4).

Table 1 Characteristics of the index and comparison cohorts

	Index cohort	Comparison cohort sampled without replacement in random order	Comparison cohort sampled with replacement	Comparison cohort sampled without replacement in chronological order
Cohort size, N	1,000 (979–1,018)	5,000 (4,893–5,088)	5,000 (4,893–5,090)	5,000 (4,893–5,088)
Male proportion, %	49.4 (48.3–50.4)	49.3 (48.2–50.4)	49.3 (48.3–50.4)	49.3 (48.3–50.4)
Median age, years	37.5 (36.7–38.2)	37.5 (36.7–38.2)	37.5 (36.7–38.2)	37.5 (36.7–38.2)
Number of comparators enrolled in the index cohort during follow-up		40 (36–44)	38 (33–42)	36 (32–41)
Ratio of unique persons to cohort size, %			93.5 (93.2–93.8)	
Follow-up, person-years	7,886 (7,702–8,054)	39,290 (38,475–40,136)	39,404 (38,599–40,206)	39,144 (38,324–39,956)
Mean follow-up, person-years	7.89 (7.82–7.96)	7.87 (7.81–7.92)	7.89 (7.83–7.95)	7.83 (7.78–7.89)
Number of ischemic strokes during follow-up	15 (12–18)	76 (70–83)	75 (69–82)	75 (69–82)
Incidence rate of ischemic stroke per 100,000 person-years	189.7 (157.8–224.8)	194.4 (179.1–211.3)	190.4 (174.9–206.7)	192.6 (175.8–210.0)
Number of deaths during follow-up	84 (78–90)	424 (403–447)	417 (394–441)	417 (394–440)
Mortality rate per 100,000 person-years	1,068.9 (989.5–1,143.9)	1,083.0 (1,026.4–1,134.7)	1,058.6 (1,003.5–1,114.9)	1,065.0 (1,009.2–1,122.9)

Note: Numbers presented are the median (interquartile range) of the specified statistic measured over 1,000 iterations.

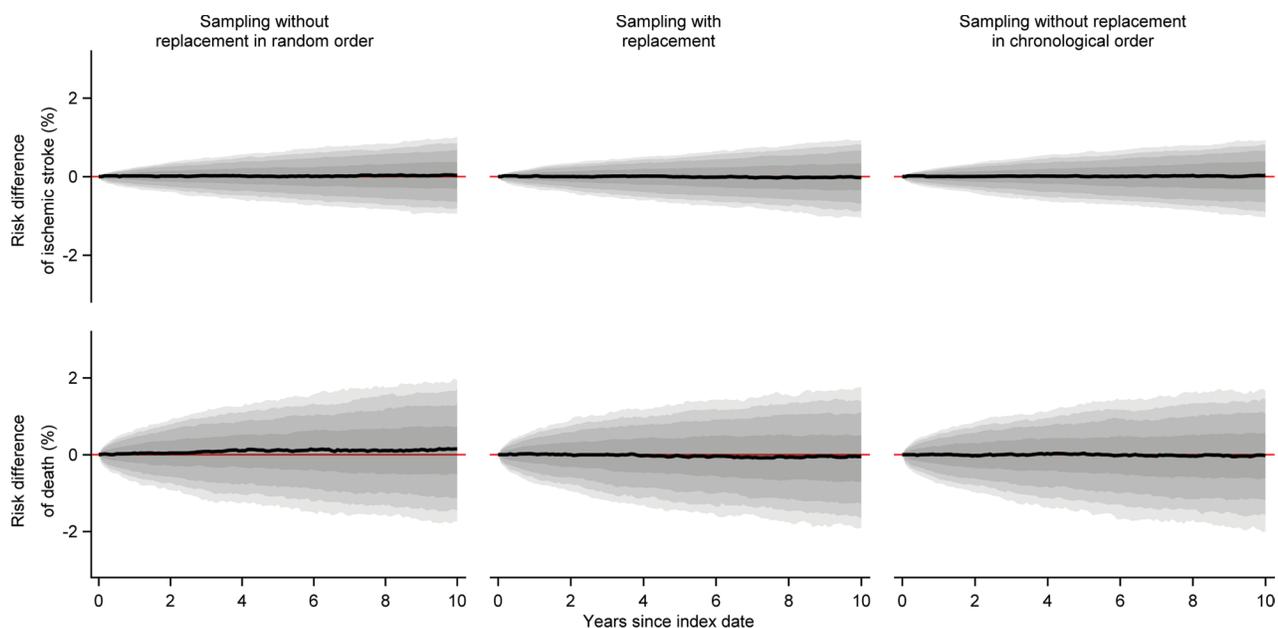


Figure 2 Medians of estimated monthly risk differences of ischemic stroke and death comparing each of the comparison cohorts to the index cohorts, along with bands covering the interquartile range, 10th–90th, 5th–95th, and 2.5th–97.5th percentiles.

When we sampled one comparator for each member of the index cohort of persons aged >50 years, we observed no substantial difference between the sampling strategies, ie, all comparison cohorts generally resembled the index cohorts (Table 3, Figures 6 and 7).

In the sensitivity analysis where we set the probability of being recruited into the index cohort to 0.5% for anyone alive in the source population on 1st January of each year, a median

of 4,737 persons from the source population were recruited into the index cohort, and as a result the comparison cohorts became very large relative to the source population of 50,000 individuals (Table S1). We found that the comparison cohorts resembled the index cohorts with respect to sex and age distributions; however, sampling without replacement in chronological order resulted in smaller comparison cohorts than the other strategies (Table S1). With regards to outcomes, we found that

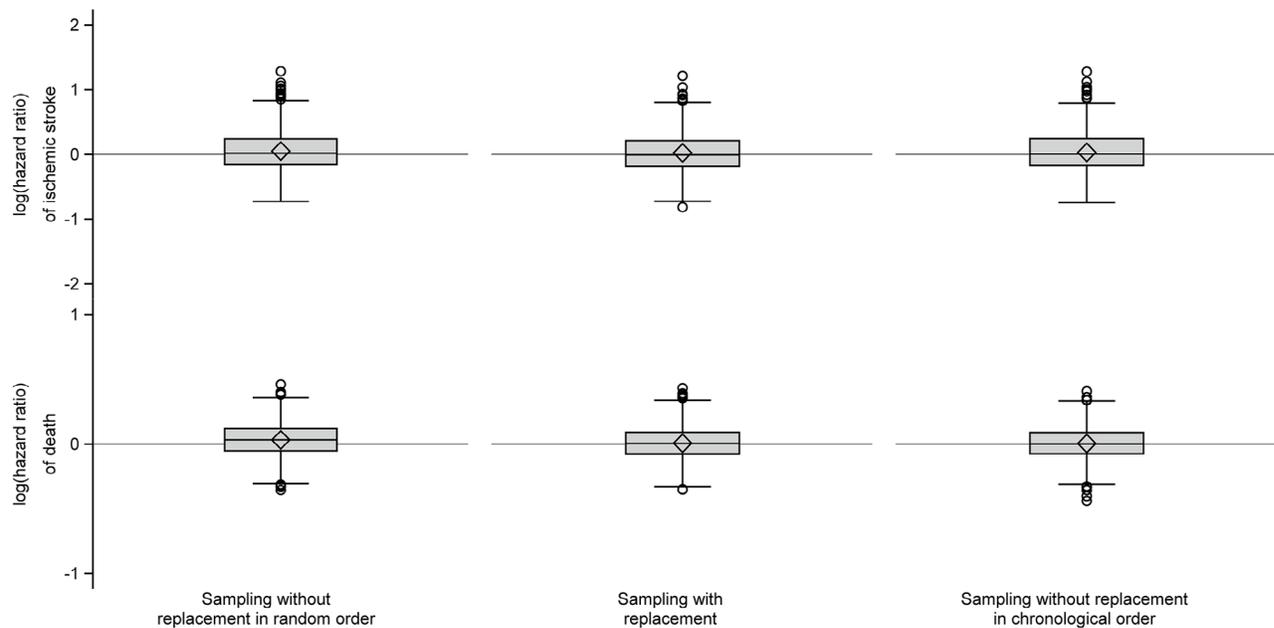


Figure 3 Box plots of log(hazard ratios) of ischemic stroke and death comparing the individual comparison cohorts to the index cohorts.

Notes: The elements of the box plot are as follows. Diamond, mean; line in the box, median; box, interquartile range (IQR), ie, range from first quartile (Q1) to third quartile (Q3); lower (upper) line outside of the box, minimum value above $Q1 - 1.5 \times IQR$ (maximum value below $Q3 + 1.5 \times IQR$); circles, outliers.

Table 2 Characteristics of the index cohorts of persons aged >50 years and their five to one sampled comparison cohorts

	Index cohort	Comparison cohort sampled without replacement in random order	Comparison cohort sampled with replacement	Comparison cohort sampled without replacement in chronological order
Cohort size, N	974 (954–994)	4,822 (4,723–4,925)	4,847 (4,749–4,947)	4,772 (4,676–4,867)
Male proportion, %	42.6 (41.6–43.7)	42.6 (41.5–43.7)	42.6 (41.5–43.7)	42.6 (41.6–43.7)
Median age, years	72.2 (71.9–72.6)	72.1 (71.8–72.4)	72.2 (71.9–72.5)	71.9 (71.7–72.3)
Number of comparators enrolled in the index cohort during follow-up		160 (150–171)	137 (128–147)	136 (127–145)
Ratio of unique persons to cohort size, %			82.2 (81.7–82.7)	
Follow-up, person-years	6,190 (6,038–6,349)	29,733 (29,014–30,387)	30,884 (30,181–31,634)	30,649 (29,928–31,407)
Mean follow-up, person-years	6.36 (6.28–6.44)	6.16 (6.10–6.22)	6.38 (6.32–6.43)	6.43 (6.37–6.48)
Number of ischemic strokes during follow-up	52 (47–57)	282 (270–295)	262 (248–275)	257 (245–270)
Incidence rate of ischemic stroke per 100,000 person-years	842.0 (761.6–925.1)	948.1 (904.8–993.1)	849.2 (805.7–889.5)	838.2 (802.8–877.0)
Number of deaths during follow-up	347 (334–359)	1,824 (1,773–1,883)	1,712 (1,666–1,764)	1,658 (1,617–1,705)
Mortality rate per 100,000 person-years	5,583.9 (5,405.1–5,792.9)	6,137.0 (5,999.1–6,291.8)	5,542.8 (5,409.7–5,679.5)	5,409.0 (5,287.9–5,523.1)

Note: Numbers presented are the median (interquartile range) of the specified statistics measured over 1,000 iterations.

sampling without replacement in random order led to positive risk differences and log-hazard ratios of both ischemic stroke and death compared to the index cohorts (Figures S1 and S2). Comparison cohorts sampled with replacement or without replacement in chronological order did not deviate substantially from the index cohort (Table S1, Figures S1 and S2).

Heart failure cohort study

In the heart failure cohort (N=347,882), the median age was 77 years and a slight majority of patients were male (Table 4). When we sampled a comparison cohort without replacement in random order, the median age was slightly lower than in the heart failure cohort (Table 4), as a substantial number of

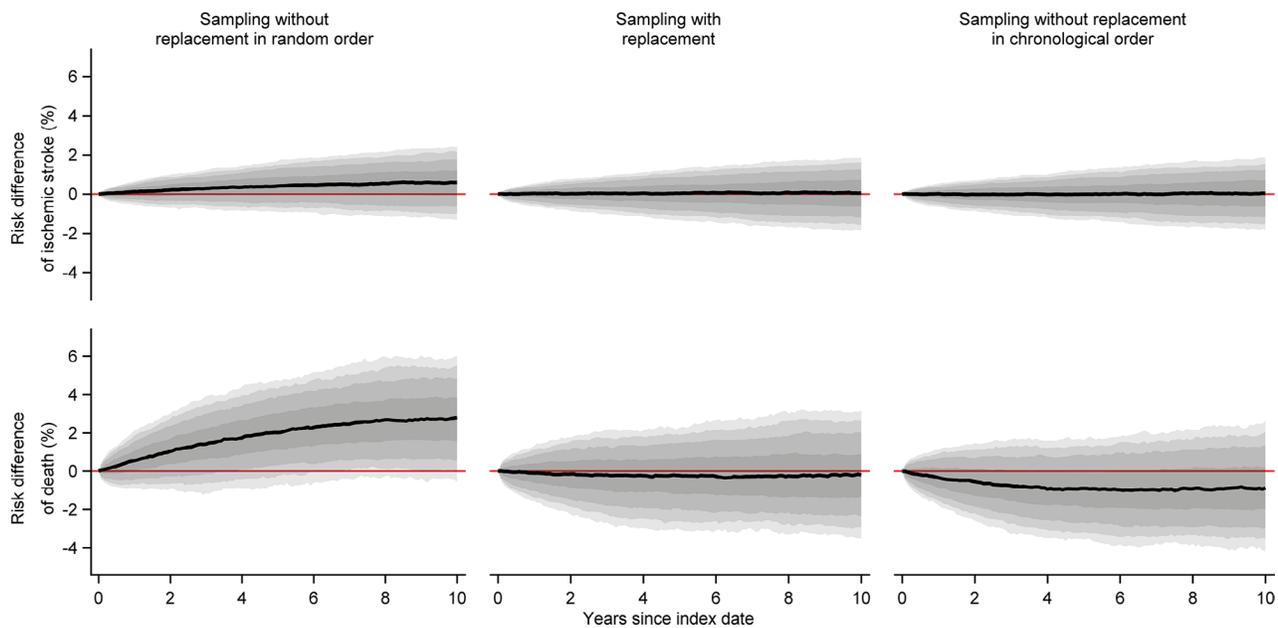


Figure 4 Medians of estimated monthly risk differences of ischemic stroke and death comparing each of the five to one sampled comparison cohorts to the index cohorts of persons aged >50 years, along with bands covering the interquartile range, 10th–90th, 5th–95th, and 2.5th–97.5th percentiles.

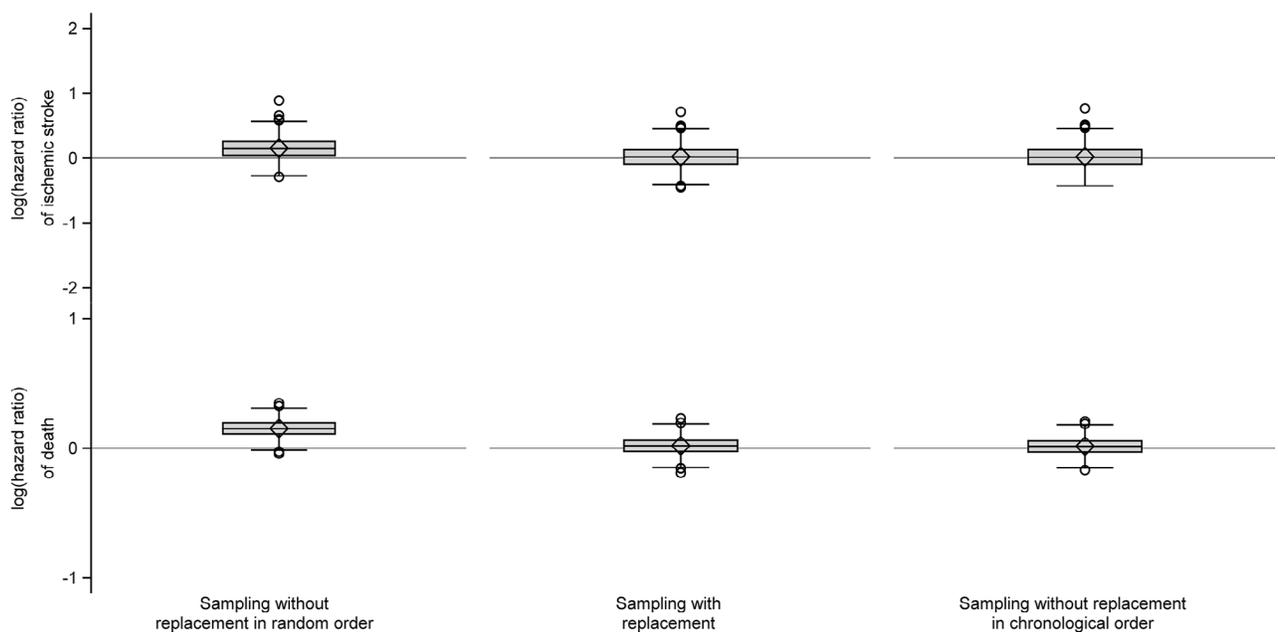


Figure 5 Box plots of log(hazard ratios) of ischemic stroke and death comparing the individual five to one sampled comparison cohort to the index cohorts of persons aged >50 years.

Notes: The elements of the box plot are as follows. Diamond, mean; line in the box: median; box, interquartile range (IQR), ie, range from first quartile (Q1) to third quartile (Q3); lower (upper) line outside of the box, minimum value above $Q1 - 1.5 \times IQR$ (maximum value below $Q3 + 1.5 \times IQR$); circles, outliers.

heart failure patients were left without a match (Table S2). When we sampled without replacement in chronological order, almost one-fifth of the heart failure patients had no match (Table S2), resulting in a comparison cohort whose age distribution was substantially lower than in the heart failure

cohort (Table 4). These problems did not occur when we sampled with replacement (Table 4) or one to one (Table S3).

Regarding outcomes, we found that the risks and rates of ischemic stroke and death in the comparison cohort sampled without replacement in random order were higher than those

Table 3 Characteristics of the index cohorts of persons aged >50 years and their one to one sampled comparison cohorts

	Index cohort	Comparison cohort sampled without replacement in random order	Comparison cohort sampled with replacement	Comparison cohort sampled without replacement in chronological order
Cohort size, N	974 (954–994)	972 (952–993)	973 (952–993)	972 (952–992)
Male proportion, %	42.6 (41.6–43.7)	42.6 (41.5–43.7)	42.6 (41.5–43.7)	42.6 (41.5–43.7)
Median age, years	72.2 (71.9–72.6)	72.2 (71.9–72.5)	72.2 (71.9–72.5)	72.2 (71.9–72.5)
Number of comparators enrolled in the index cohort during follow-up		28 (24–32)	28 (24–32)	28 (24–32)
Ratio of unique persons to cohort size, %			95.9 (95.5–96.4)	
Follow-up, person-years	6,190 (6,038–6,349)	6,164 (6,016–6,328)	6,183 (6,036–6,341)	6,176 (6,036–6,347)
Mean follow-up, person-years	6.36 (6.28–6.44)	6.35 (6.27–6.42)	6.36 (6.28–6.44)	6.36 (6.29–6.43)
Number of ischemic strokes during follow-up	52 (47–57)	53 (48–57)	52 (47–57)	52 (48–57)
Incidence rate of ischemic stroke per 100,000 person-years	842.0 (761.6–925.1)	851.3 (770.9–937.3)	837.2 (759.1–925.8)	844.8 (773.1–924.8)
Number of deaths during follow-up	347 (334–359)	348 (335–361)	345 (333–358)	346 (333–359)
Mortality rate per 100,000 person-years	5,583.9 (5,405.1–5,792.9)	5,632.8 (5,428.5–5,841.8)	5,576.9 (5,392.8–5,795.6)	5,590.4 (5,392.1–5,778.1)

Note: Numbers presented are the median (interquartile range) of the specified statistics measured over 1,000 iterations.

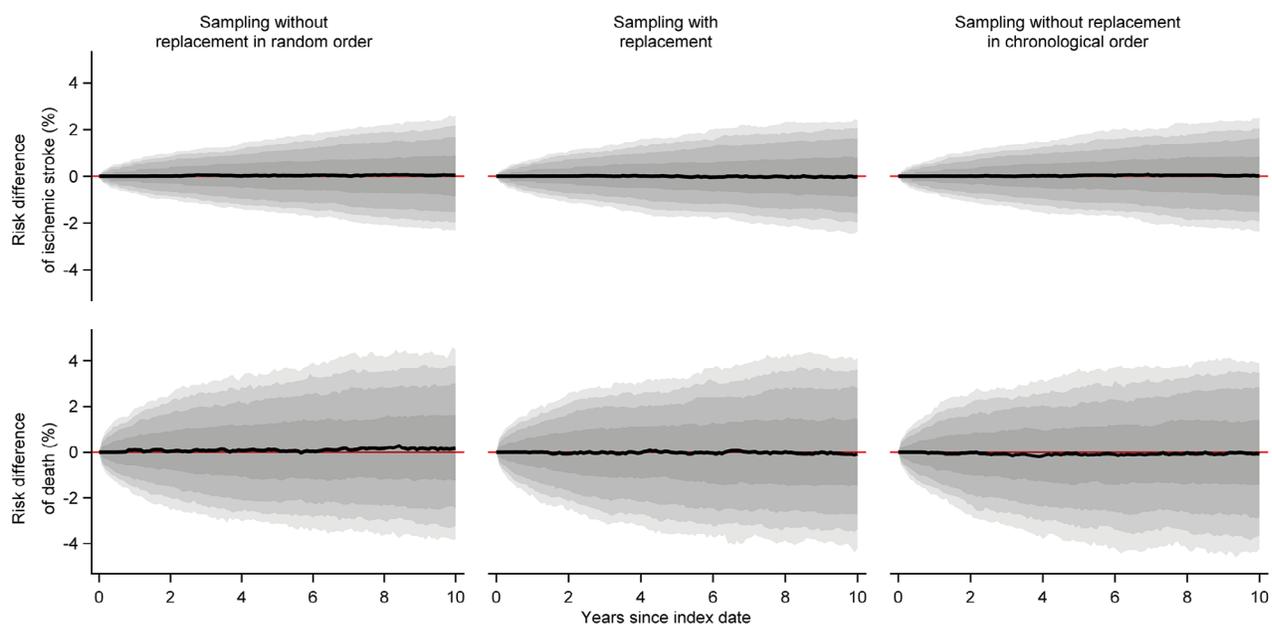


Figure 6 Medians of estimated monthly risk differences of ischemic stroke and death comparing each of the one to one sampled comparison cohorts to the index cohorts of persons aged >50 years, along with bands covering the interquartile range, 10th–90th, 5th–95th, and 2.5th–97.5th percentiles.

observed in the other comparison cohorts (Table 4, Figure 8). Furthermore, the risks and rates of outcomes in the comparison cohort sampled without replacement in chronological order were lower than in the comparison cohort sampled with replacement (Table 4, Figure 8).

However, when we sampled one to one, the differences between all comparison cohorts decreased, though the

mortality rate was still elevated in the comparison cohort sampled without replacement in random order (Table S3, Figure S3). The results from the comparison cohorts sampled with replacement or without replacement in chronological order became almost identical (Table S3, Figure S3).

In the analysis where we restricted to patients for whom the individual sampling strategy identified five comparators,

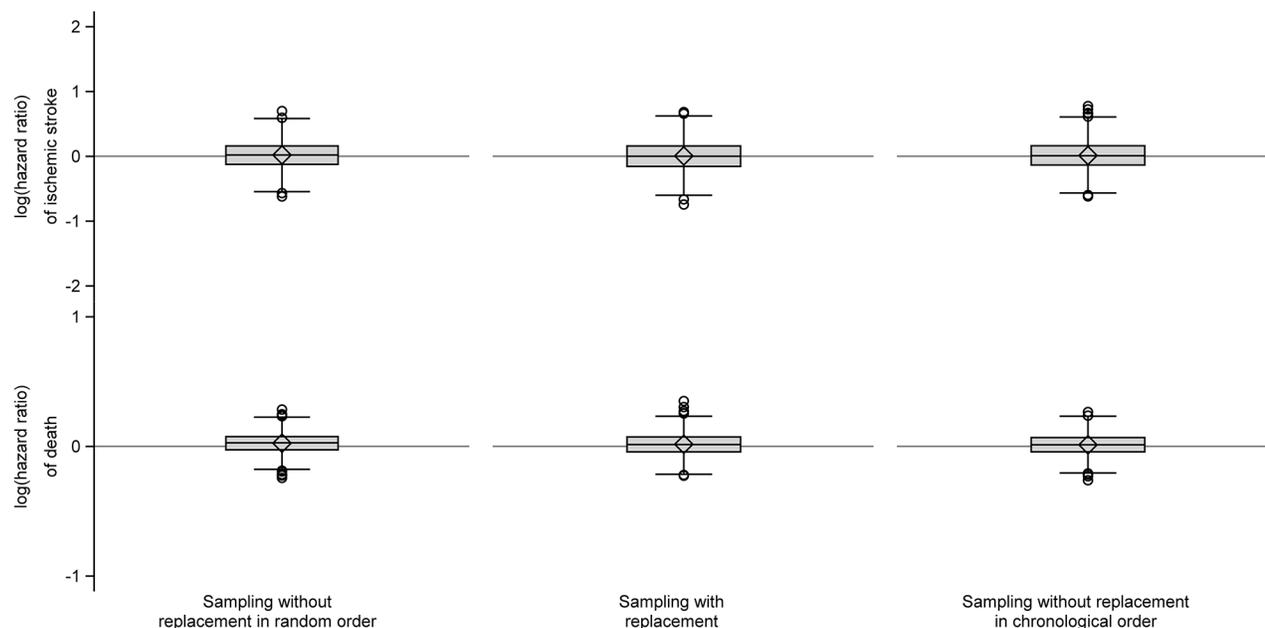


Figure 7 Box plots of log(hazard ratios) of ischemic stroke and death comparing the individual one to one sampled comparison cohort to the index cohorts of persons aged >50 years.

Notes: The elements of the boxplot are as follows. Diamond, mean; line in the box, median; box, interquartile range (IQR), ie, range from first quartile (Q1) to third quartile (Q3); lower (upper) line outside of the box, minimum value above $Q1 - 1.5 \times IQR$ (maximum value below $Q3 + 1.5 \times IQR$); circles: outliers.

Table 4 Characteristics of the heart failure cohort and its comparison cohorts

	Heart failure cohort	Comparison cohort sampled without replacement in random order	Comparison cohort sampled with replacement	Comparison cohort sampled without replacement in chronological order
Cohort size, N	347,882	1,589,687	1,739,401	1,395,256
Male sex, N (%)	183,176 (52.7)	825,608 (51.9)	915,871 (52.7)	723,353 (51.8)
Median age (IQR), years	76.5 (68.0–83.3)	75.3 (67.0–81.9)	76.5 (68.0–83.3)	73.6 (65.6–79.3)
Unique persons, N (%)		1,105,918 (63.58)		
Number of comparators who developed heart failure during follow-up, N (%)		3,572 (0.2)	3,433 (0.2)	2,693 (0.2)
Follow-up, person-years	1,102,609	8,097,781	10,501,626	9,182,655
Median follow-up (IQR), person-years	2.0 (0.3–5.2)	4.5 (1.8–9.1)	6.2 (2.8–10.0)	7.4 (3.5–10.0)
Number of ischemic strokes during follow-up	21,554	143,710	147,314	119,644
Incidence rate of ischemic stroke per 1,000 person-years (95% CI)	17.6 (17.5–17.8)	14.3 (14.2–14.4)	10.1 (10.0–10.1)	8.6 (8.5–8.6)
Incidence rate ratio of ischemic stroke (95% CI)		1.23 (1.22–1.24)	1.75 (1.73–1.77)	2.05 (2.03–2.07)
Incidence rate ratio of ischemic stroke (95% CI) after restricting the analysis to strata with five matches		1.18 (1.16–1.20)	1.75 (1.72–1.78)	1.86 (1.83–1.89)
Number of deaths during follow-up	251,408	800,937	753,535	521,064
Mortality rate per 1,000 person-years (95% CI)	228.0 (227.5–228.5)	98.9 (98.7–99.1)	71.8 (71.6–71.9)	56.7 (56.6–56.9)
Mortality rate ratio (95% CI)		2.31 (2.30–2.31)	3.18 (3.17–3.19)	4.02 (4.00–4.03)
Mortality rate ratio (95% CI) after restricting the analysis to strata with five matches		2.20 (2.19–2.21)	3.18 (3.16–3.19)	3.53 (3.51–3.54)

Abbreviations: CI, confidence interval; IQR, interquartile range.

the differences in incidence rate ratios decreased when sampling without replacement in chronological order compared to sampling with replacement. However, the differences

remained substantial (Table 4). Furthermore, when sampling without replacement in random order, the mortality rate ratio decreased (Table 4).

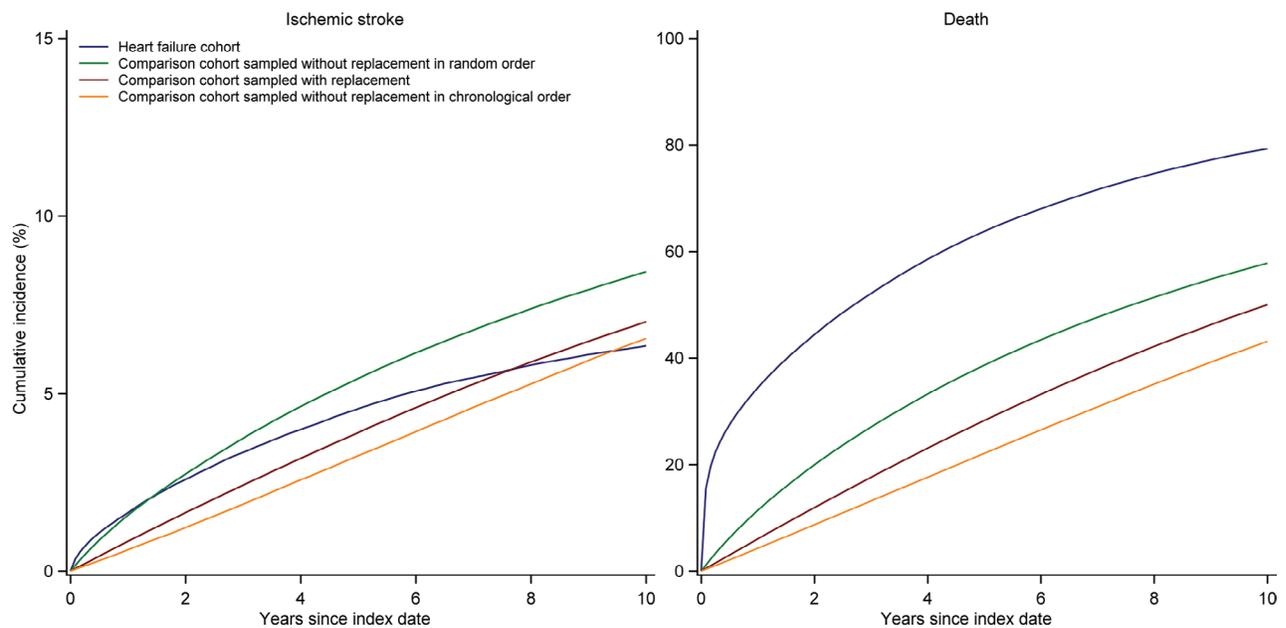


Figure 8 Cumulative incidence of ischemic stroke and death in the heart failure cohort and the comparison cohorts sampled using different strategies.

Discussion

In this study, we found that the three sampling strategies generated comparison cohorts with outcome rates similar to that observed in the index cohorts representing the general population, unless the index persons were aged >50 years or the comparison cohorts were very large compared to the source population. Thus, our hypothesis, that sampling without replacement in random order would result in comparison cohorts with increased outcome rates compared to the general population, was only confirmed in some settings in the general population cohort study. In the heart failure cohort study, we found that different sampling strategies generated substantially different results, both for fatal and non-fatal outcomes. However, differences between comparison cohorts for both the index cohort with persons aged >50 years and the heart failure cohort attenuated greatly when sampling one to one rather than five to one.

The null result in the first analysis of the general population cohort study suggests that even if the comparison cohort includes as many as 10% of the source population, the choice of sampling strategy may not in itself impact the results. That said, the sampling strategy may be important if comparators are at a substantial risk of getting the outcome. This was exemplified in our study by constructing index cohorts of people aged >50 years.

All three sampling strategies led to comparison cohorts with different characteristics when the index cohort comprised people aged >50 years. This was reflected in the heart failure

cohort study, in which sampling without replacement failed to sample an adequate number of comparators, particularly when sampling in chronological order. As researchers typically exclude patients for whom no match can be found,^{27,28} it is an important observation that the age distribution was skewed toward lower age in the comparison cohort sampled without replacement in chronological order, compared with the heart failure cohort. It follows that if patients without matches were removed from the cohort under study, the remaining patients would constitute a selected (younger) subcohort.^{23,24} We attribute the lower outcome rates observed when sampling without replacement in chronological order, compared to sampling with replacement, to the age difference between members of the two comparison cohorts.

Several studies have been published in which a comparison cohort has been sampled from the general population.⁵⁻¹⁹ While a few studies reported sampling with replacement,^{5,19} the sampling strategy was not described in others.⁶⁻¹⁸

In addition to the sampling strategy, several other important factors must be considered when conducting a matched cohort study.¹ The sampling ratio, ie, the number of comparators sampled for each patient, is an important consideration, as highlighted by the sensitivity analyses in which we changed the sampling ratio. If subgroup analyses are conducted, it may be necessary to sample a greater number of comparators to ensure their availability for all index persons than if no subgroup analyses are conducted. Furthermore, to get

precise estimates, more comparators are needed if the outcome of interest is rare rather than frequent. However, it is questionable if there is much to gain in statistical power when the sampling ratio exceeds four.²⁹ Also, if the patient cohort is large compared to the general population, conditional on matching factors, it may not be feasible to sample many comparators for each index person. In previous studies, the sampling ratio has varied between 4¹⁷ and 100.^{6,13}

Another aspect to consider is the matching factors. In addition to sex, birth year, and calendar period, comparison cohorts have been matched on factors such as residency^{16,17} or a specific comorbidity.¹⁷ However, matching on some confounders may reduce rather than improve efficiency.³⁰ Furthermore, increasing the number of matching factors increases the risk of finding no eligible comparators, because the matching factors function as eligibility criteria for the comparison cohort.

Our study has several limitations. Although it showed differences between comparison cohorts selected by different sampling strategies in some settings, it did not present means of quantifying the size of these differences in general. Furthermore, we chose heart failure to illustrate the sampling strategies because this disease is common among the elderly,²³ and old age is generally associated with high rates of ischemic stroke and mortality. For this reason, the differences in outcome rates between comparison cohorts may be extreme, compared to differences that would be observed in smaller patient cohorts with less frequent outcomes in the general population. Also, in the heart failure cohort study, we did not know the true outcome rates in the general population, unlike in the general population study. Moreover, there might be differences between sampling strategies which we did not detect, eg, we did not assess baseline covariates beyond the matching factors.

Conclusion

When the comparators were at a substantial risk of getting the outcome, the sampling strategy impacted study results. Thus, sampling without replacement in random order led to comparison cohorts with inflated outcome rates in such settings, while sampling without replacement in chronological order led to an unsuccessful matching when sampling five to one. Therefore, sampling with replacement may be advisable in matched cohort studies when the outcome of interest is common among comparators. Alternatively, one to one matching may be applicable.

Acknowledgments

Funding: This paper was supported by the Program for Clinical Research Infrastructure (PROCRIN) established by the

Lundbeck Foundation and the Novo Nordisk Foundation and administered by the Danish Regions. PROCRIN had no role in the conception and design of the study, nor in the collection, analyses, or interpretation of the data, nor in the writing of the manuscript and the decision to submit for publication.

Disclosure

The authors report no conflicts of interest in this work.

References

- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
- Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549.
- Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekblom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009;24(11):659–667.
- Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–836.
- Adelborg K, Horváth-Puhó E, Ording A, Pedersen L, Toft Sørensen H, Henderson VW. Heart failure and risk of dementia: a Danish nationwide population-based cohort study. *Eur J Heart Fail*. 2017;19(2):253–260.
- Broesby-Olsen S, Farkas DK, Vestergaard H, et al. Risk of solid cancer, cardiovascular disease, anaphylaxis, osteoporosis and fractures in patients with systemic mastocytosis: a nationwide population-based study. *Am J Hematol*. 2016;91(11):1069–1075.
- Corraini P, Ording AG, Henderson VW, et al. Cancer, other comorbidity, and risk of venous thromboembolism after stroke: a population-based cohort study. *Thromb Res*. 2016;147:88–93.
- Faurschou M, Ahlström MG, Lindhardsen J, Baslund B, Obel N. Impact of pre-existing co-morbidities on mortality in granulomatosis with polyangiitis: a cohort study. *Rheumatology*. 2016;55(4):649–653.
- Faurschou M, Ahlström MG, Lindhardsen J, Obel N, Baslund B. Risk of diabetes mellitus among patients diagnosed with giant cell arteritis or granulomatosis with polyangiitis: comparison with the general population. *J Rheumatol*. 2017;44(1):78–83.
- Folkestad L, Hald JD, Canudas-Romo V, et al. Mortality and causes of death in patients with osteogenesis imperfecta: a register-based nationwide cohort study. *J Bone Miner Res*. 2016;31(12):2159–2166.
- Hallager S, Brehm Christensen P, Ladelund S, et al. Mortality rates in patients with chronic hepatitis C and cirrhosis compared with the general population: a Danish cohort study. *J Infect Dis*. 2017;215(2):192–201.
- Iversen LH, Pedersen L, Riis A, Friis S, Laurberg S, Sørensen HT. Population-based study of short- and long-term survival from colorectal cancer in Denmark, 1977–1999. *Br J Surg*. 2005;92(7):873–880.
- Jensen-Fangel S, Pedersen L, Pedersen C, et al. Low mortality in HIV-infected patients starting highly active antiretroviral therapy: a comparison with the general population. *AIDS*. 2004;18(1):89–97.
- Leganger J, Søborg MK, Mortensen LQ, Gregersen R, Rosenberg J, Burcharth J. Association between diverticular disease and Ehlers-Danlos syndrome: a 13-year nationwide population-based cohort study. *Int J Colorectal Dis*. 2016;31(12):1863–1867.
- Lindahl Norberg A, Montgomery SM, Bottai M, Heyman M, Hovén EI. Short-term and long-term effects of childhood cancer on income from employment and employment status: a national cohort study in Sweden. *Cancer*. 2017;123(7):1238–1248.
- Lohse N, Hansen AB, Pedersen G, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med*. 2007;146(2):87–95.

17. Matthews A, Langan SM, Douglas IJ, Smeeth L, Bhaskaran K. Phosphodiesterase type 5 inhibitors and risk of malignant melanoma: matched cohort study using primary care data from the UK clinical practice research datalink. *PLoS Med.* 2016;13(6):e1002037.
18. Mellekjaer L, Papadopoulos FC, Pukkala E, et al. Cancer incidence among patients with anorexia nervosa from Sweden, Denmark and Finland. *PLoS One.* 2015;10(5):e0128018.
19. Svensson E, Horváth-Puhó E, Stockholm MG, Sørensen HT, Henderson VW, Borghammer P. Appendectomy and risk of Parkinson's disease: a nationwide cohort study with more than 10 years of follow-up. *Mov Disord.* 2016;31(12):1918–1922.
20. Lohr SL. *Sampling: Design and Analysis*. 2nd ed. Boston, MA: Brooks/Cole; 2010.
21. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf.* 2007;16(3):241–249.
22. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449–490.
23. Roger VL. Epidemiology of heart failure. *Circ Res.* 2013;113(6):646–659.
24. Writing Group Members, Mozaffarian D, Benjamin EJ, et al. Heart disease and stroke statistics – 2016 update: a report from the American Heart Association. *Circulation.* 2016;133(4):e38–e360.
25. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res.* 2013;22(1):70–96.
26. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: Wiley; 2002.
27. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol.* 2013;42(3):860–869.
28. Sjölander A, Greenland S. Ignoring the matching variables in cohort studies – when is it valid and why? *Stat Med.* 2013;32(27):4696–4708.
29. Daly LE, Bourke GJ, Bourke GJ. *Interpretation and Uses of Medical Statistics*. 5th ed. Malden, MA: Blackwell Science; 2000.
30. Greenland S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol.* 1990;131(1):151–159.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress