

## Supplementary Material 1:

[An animation](#) illustrating the three sampling strategies matching (five to one) on sex, birth year, and calendar time.

Each black pin represents an index person. They are all born in the same year and all are of the same sex. Their positions on the timescale reflect their respective index dates. Each horizontal grey line represents a person born in the same year and who is the same sex as the index persons. The grey line spans the time a given person is available in the population registry, and as such, the time interval during which the person can be part of a risk set. The risk set of a given index person is represented by the grey lines which intersect the red line. The blue lines represent the individuals selected as comparators for a given index person. In the panel on the left, sampling is done without replacement in random order; in the center panel, sampling is done with replacement; and in the right-hand panel, sampling is done without replacement in chronological order.

We notice that the set of grey lines is constant throughout each step of the sampling procedure when sampling is done with replacement. When sampling is done without replacement the density of the grey lines increases with decreasing follow-up as the sampling procedure progresses. When sampling is done in random order, the density of the grey lines across the red line, *i.e.*, within the risk set, tends to be concentrated increasingly among those with short follow-up, as the sampling procedure progresses. When sampling is done in chronological order there is no clear tendency of a gradient in the density of the grey lines across the red line, *i.e.*, the risk set is not an increasingly selected population of persons with short follow-up. However, late in the sampling procedure the risk sets become small or empty, making it impossible to find a sufficient number of matches for every index person.

**Supplementary Material 2:**

*Diagnostic codes used to identify heart failure and ischemic stroke in the Danish National Patient Registry*

	ICD-8#	ICD-10#
Heart failure	42709 42710 42711 42719 42899 78249	DI50 DI110 DI130 DI132
Ischemic stroke	433 434	DI63 DI64

#Including all subcodes.

### Supplementary Material 3:

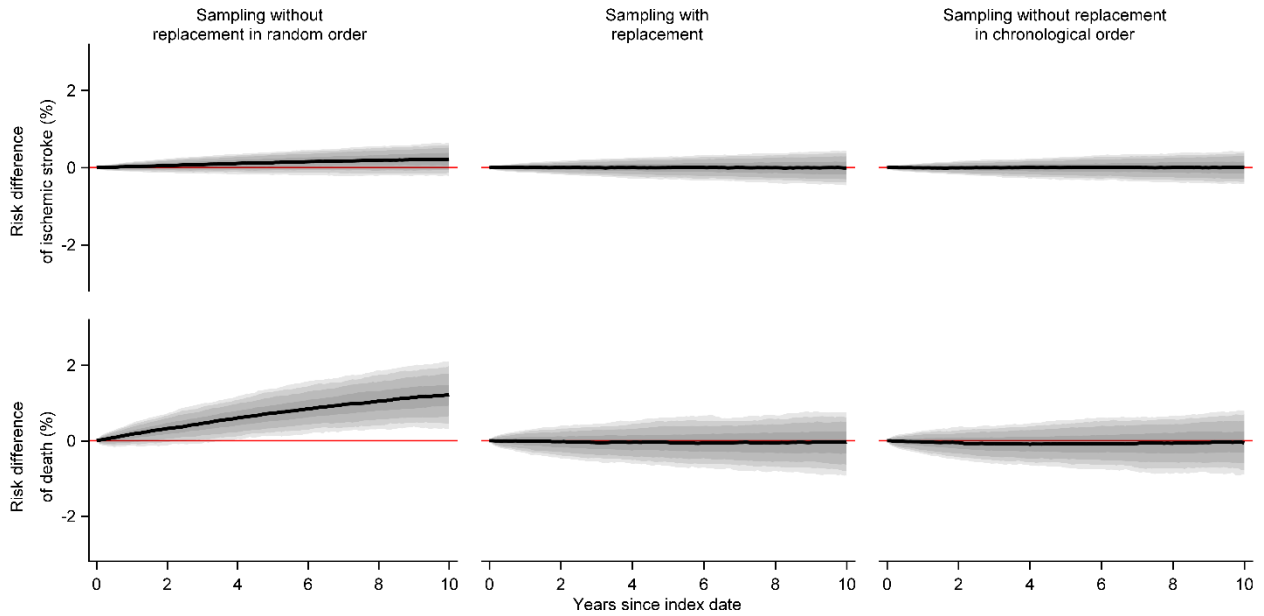
#### Supplementary Tables and Figures.

The table and figures below are based on situations in which the probability of being recruited into the index cohort, if alive on 1 January any given year, was 0.5%.

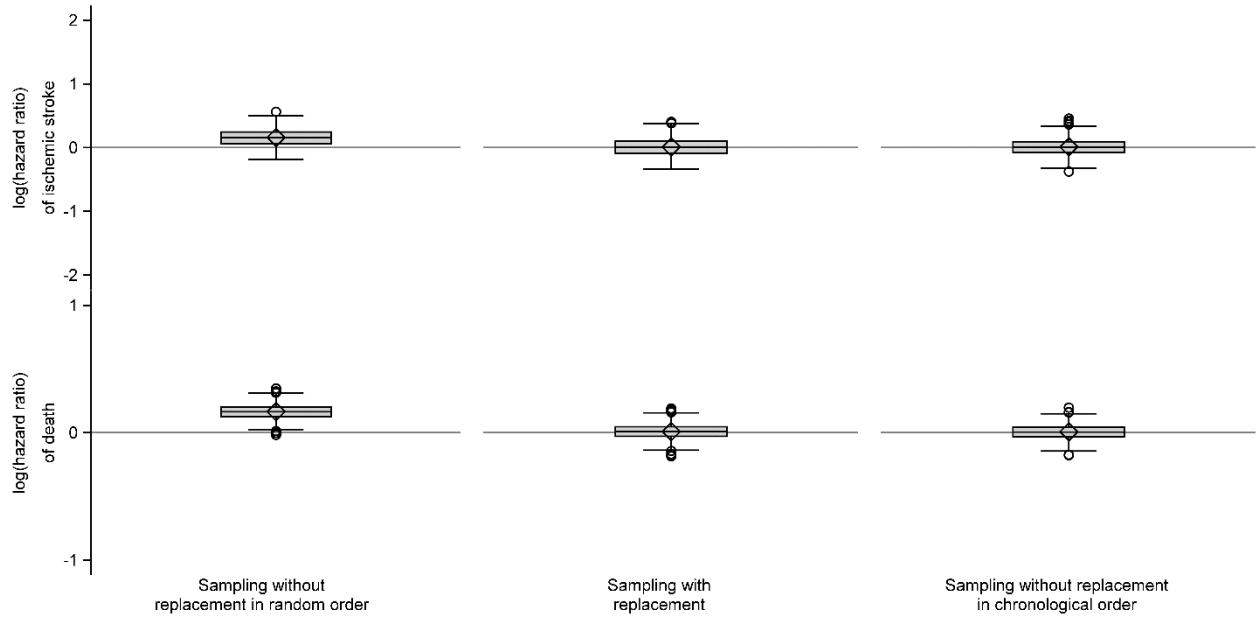
Supplementary Table 1. Characteristics of the index and comparison cohorts. Numbers presented are the median (interquartile range) of the specified statistics measured over 1,000 iterations.

	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	4,737 (4,691 - 4,779)	23,648 (23,420 - 23,863)	23,669 (23,442 - 23,886)	23,539 (23,322 - 23,751)
Male proportion (%)	49.4 (48.9 - 49.8)	49.4 (48.9 - 49.8)	49.4 (48.9 - 49.8)	49.4 (48.9 - 49.9)
Median age, years	37.1 (36.8 - 37.5)	37.1 (36.8 - 37.4)	37.1 (36.8 - 37.5)	36.9 (36.6 - 37.3)
Number of comparators enrolled in the index cohort during follow-up		1,138 (1,108 - 1,172)	866 (840 - 891)	847 (824 - 871)
Ratio of unique persons to cohort size (%)			73.5 (73.3 - 73.8)	
Follow-up, person-years	37,564 (37,197 - 37,964)	184,228 (182,485 - 186,090)	187,843 (185,924 - 189,839)	184,371 (182,589 - 186,250)
Mean follow-up, person-years	7.93 (7.90 - 7.96)	7.79 (7.76 - 7.82)	7.94 (7.91 - 7.96)	7.84 (7.81 - 7.86)
Number of ischemic strokes during follow-up	71 (65 - 77)	395 (381 - 411)	355 (338 - 370)	349 (336 - 363)
Incidence rate of ischemic stroke per 100,000 person-years	189.1 (173.6 - 204.4)	214.6 (207.1 - 222.7)	188.8 (180.3 - 197.0)	188.8 (181.7 - 196.8)
Number of deaths during follow-up	400 (387 - 413)	2222 (2164 - 2279)	1990 (1941 - 2037)	1951 (1905 - 1997)
Mortality rate per 100,000 person-years	1,064.4 (1,031.6 - 1,100.1)	1,207.6 (1,177.8 - 1,235.8)	1,058.0 (1,035.5 - 1,084.1)	1,057.6 (1,034.8 - 1,082.1)

Supplementary Figure 1. Medians of estimated monthly risk differences of ischemic stroke and death comparing each of the comparison cohorts to the index cohort, along with bands covering the interquartile range, 10<sup>th</sup>-90<sup>th</sup>, 5<sup>th</sup>-95<sup>th</sup>, and 2.5<sup>th</sup>-97.5<sup>th</sup> percentiles.



Supplementary Figure 2. Box plots of log(hazard ratios) of ischemic stroke and death comparing the individual comparison cohort to the index cohort.



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

The table below contains additional data describing the successfulness of the sampling procedures in the heart failure study when sampling five comparators for each heart failure patient.

Supplementary Table 2. Numbers of comparators sampled to individual heart failure patients using the different sampling strategies.

	Comparison cohort sampled without replacement in random order	Comparison cohort sampled with replacement	Comparison cohort sampled without replacement in chronological order
Number of heart failure patients with 0 matches, N (%)	29,107 (8.37)	0 (0.00)	67,900 (19.5)
Number of heart failure patients with 1 matches, N (%)	565 (0.16)	0 (0.00)	988 (0.28)
Number of heart failure patients with 2 matches, N (%)	359 (0.10)	0 (0.00)	40 (0.01)
Number of heart failure patients with 3 matches, N (%)	295 (0.08)	4 (0.00)	22 (0.01)
Number of heart failure patients with 4 matches, N (%)	261 (0.08)	1 (0.00)	14 (0.00)
Number of heart failure patients with 5 matches, N (%)	317,295 (91.2)	347,877 (100)	278,918 (80.2)

Supplementary Table 3. Characteristics of the heart failure cohort and its comparison cohorts when sampling one to one.

	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	347,882	347,882	347,875
Male sex, N (%)	183,176 (52.7)	183,176 (52.7)	183,176 (52.7)	183,169 (52.7)
Median age (IQR), years	76.5 (68.0 - 83.3)	76.5 (68.0 - 83.3)	76.5 (68.0 - 83.3)	76.5 (68.0 - 83.3)
Unique persons, N (%)			311,537 (89.55)	
Number of comparators who developed heart failure during follow-up, N (%)		680 (0.2)	666 (0.2)	657 (0.2)
Follow-up, person-years	1,102,609	2,060,479	2,098,041	2,102,511
Median follow-up (IQR), person-years	2.0 (0.3 - 5.2)	6.0 (2.7 - 10.0)	6.2 (2.8 - 10.0)	6.2 (2.8 - 10.0)
Number of ischemic strokes during follow-up	21,554	29,825	29,454	29,587
Incidence rate of ischemic stroke per 1,000 person-years (95% CI)	17.6 (17.5 - 17.8)	10.5 (10.4 - 10.7)	10.1 (10.0 - 10.2)	10.1 (9.9 - 10.2)
Incidence rate ratio of ischemic stroke (95% CI)		1.67 (1.65 - 1.70)	1.74 (1.72 - 1.77)	1.75 (1.72 - 1.78)
Incidence rate ratio of ischemic stroke (95% CI) after restricting the analysis to patients with a match		1.67 (1.65 - 1.70)	1.74 (1.71 - 1.77)	1.75 (1.71 - 1.78)
Number of deaths during follow-up	251,408	155,337	150,639	150,304
Mortality rate per 1,000 person-years (95% CI)	228.0 (227.5 - 228.5)	75.4 (75.0 - 75.8)	71.8 (71.4 - 72.2)	71.5 (71.1 - 71.9)
Mortality rate ratio (95% CI)		3.02 (3.01 - 3.04)	3.18 (3.16 - 3.19)	3.19 (3.17 - 3.21)
Mortality rate ratio of ischemic stroke (95% CI) after restricting the analysis to patients with a match		3.02 (3.01 - 3.04)	3.18 (3.16 - 3.20)	3.19 (3.17 - 3.21)

Abbreviations: CI = confidence interval, IQR = interquartile range

Supplementary Figure 3. Cumulative incidence of ischemic stroke and death in the heart failure cohort and the comparison cohorts sampled one to one using different strategies.

