The Impact of Selective Survival on Sex Disparities in Stroke Incidence: Evaluating Bias using a Simulation Model

Xiaolei Lin
Fudan University

Leah Li (leah.li@ucl.ac.uk)
University College London Great Ormond Street Institute of Child Health

Research Article

Keywords: Selective survival, survivor bias, sex disparities, stroke incidence rate, risk factors for survival

Posted Date: January 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1163954/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Males tend to have higher stroke incidence than females. The difference attenuates (or even reverses) in late-life. Using a simulation study, we assessed how selective survival impacted on age attenuation of the male-female difference in stroke incidence.

Methods: We simulated a birth cohort of 50,000 males and 50,000 females. The distributions of survival and first stroke incident were calibrated from published data. We generated an unobserved construct U (i.e. risk factors for incident stroke and/or survival) and considered four causal scenarios based on the sex-specific association between U and survival. To assess the bias under each scenario, we compared male–female difference in stroke incidence from 45 to 95y with pre-specified value (5 per 1000 person-year).

Results: Under all causal scenarios, male-female difference in stroke incidence was close to pre-specified value from mid-adulthood but decreased (thus bias increased) at older ages. There was only a small increase in bias when U did not affect survival (Scenario A). When U had a direct effect on survival, the difference decreased (thus bias increased) rapidly, reversed from 95y when U affected survival with a similar magnitude for males and females (Scenario B: hazard ratio (HR)=1.5). The difference reversed earlier when the effect of U on survival was greater for males than females: at 85y (Scenario C: HR=1.5 vs 1.0) and 80y (Scenario D: HR=2.5 vs 1.5).

Conclusions: Males had higher stroke incidence overall, but the male-female difference attenuated with age when there existed unobserved common risk factors for stroke and survival. The survivor bias was greater when their association with survival differed by sex.

Introduction

Stroke is currently the second leading cause of death and is responsible for approximately 11% of the total deaths globally(1). During most of the adult life, stroke incidence is higher for males compared to females. The reasons for the sex disparities in stroke incidence are not fully understood. Differential distributions of stroke risk factors in males and females may partly account for the disparities. For example, high blood pressure, atrial fibrillation and type 2 diabetes have been found to be more common in males than in females(2). Smoking and heavy drinking are also more prevalent among males(3). Moreover, associations between risk factors and stroke may differ by sex(4, 5).

However, the magnitude of the male-female difference (or ratio) in stroke incidence attenuates in late life(6, 7). A meta-analysis of 59 studies worldwide showed that the age-adjusted stroke incidence rate was 33% higher in males than in females. The male-to-female incidence rate ratio peaked in the 65-74y age band, reduced between 75-84y, and attenuated further thereafter(8).

The sex difference in stroke may eventually reverse the direction at older ages. In a population-based study in Sweden, females had 60% lower stroke incidence rate between 55-65y, but 50% higher rate after 75y than males(9). In the UK Oxford Vascular Study, the incidence rate of ischemic stroke was lower between 55-75y, but higher after 85y in females compared to males in the same age groups(10). In the US REGARDS (Reasons for Geographic and Racial Differences in Stroke) cohort, stroke incidence was four times greater in males than in females aged from 45 to 54y. The male-to-female incidence rate ratio attenuated at older ages and was lower in males by 85y(11).

This attenuation (or reversal) of the sex difference in stroke incidence in increasing age could be due to the higher mortality rate for males than for females(12). Furthermore, people live into an old age stroke-free may represent a more advantageous population. Males who survive to old ages are healthier and have a lower risk of stroke than males have died. Consequently, stroke incidence rate is lower for males than females at older ages. This selective survival can manifest the sex–stroke relationship(13, 14), and thus, introduces survivor bias when some unobserved determinants of survival (or mortality) are also associated with stroke. However, existing studies of stroke incidence often restricted to individuals who were alive(9, 10, 15–17). The sex-stroke relationship found might not reflect the casual effect of sex on stroke and could be subject to selection bias, especially when the study sample consisted of old people. Therefore, the degree to which sex disparities in stroke are due to social or biological influences, as opposed to manifestation (an artifact) of selective survival is not well understood.

The objective of this investigation was to evaluate the role of selective survival in explaining the sex disparities in first stroke incidence from mid-life until very old age. We adopted a simulation approach to assess the magnitude of the potential survival bias under several causal scenarios.

Methods

We simulated a cohort of 100,000 (50,000 male and 50,000 female) individuals who were followed from birth to 95y (or death, whichever occurred first). The mortality rates for age bands 0-1y, 1-5y and every 5y from 5 to 95y were obtained from the US 1959-1961 birth cohort for white males and females(12).

We assumed the first-ever stroke occurred from age 45y. The marginal incidence rates for first stroke between 45-95y were calibrated using the 10y ischaemic stroke incidence rates from the Oxford Vascular Study(10). The rates were adjusted by an increase of ~50-100% for each sex and age band in our simulation, as the Oxford Vascular Study members were less deprived(18) and had a lower risk of stroke than the rest of England(19). Details about the adjustments are provided in the Supplementary materials (A1).

To quantify the survivor bias, we generated data under the null hypothesis that the male-female difference in first-ever stroke incidence remained 5 per 1000 person-year across age bands from 45 to 95y. Any deviation in the observed sex difference from 0.005 person-year reflected survivor bias. We allowed incident stroke to be associated with increased hazard of death within the same or subsequent age band(s), i.e. the effect of stroke history. We assumed that there existed an unobserved and time-invariant construct U that influenced the risk of stroke and/or survival (e.g. genetic variant or lifestyle factors), and no other
confounders or mediators in the sex–stroke causal pathways (Figure 1). We explored four causal scenarios according to the relations between U, incident stroke and survival for males and females to access sex disparities in stroke incidence in each situation considering the mortality rates at different life stages.

Causal scenarios

In all four causal scenarios (Figure 1), the mortality rate was higher in males than females, especially during early and middle life(12) consistent with the UK national statistics(20). We assumed that stroke incidence rates were also higher for males than females. In addition, the hazard ratios (HR) for mortality associated with instantaneous stroke and stroke history were both set to be the same across all causal scenarios.

In the simulated cohort, U was generated from a standard normal distribution for each individual. The effect of U on incident stroke was the same in all four scenarios (HR=1.5), while its effect on survival (or mortality) varied (Figure 1). Details about the simulation parameters are provided in Supplementary materials (A2).

In causal scenario A, U did not influence mortality (HR=1.0). In scenarios B to D, U had a direct effect on mortality. In scenario B, U influenced mortality with the same magnitude for males and females (HR=1.5). In scenario C, U influenced mortality for males (HR=1.5), but not females (HR=1.0). In scenario D, the influence of U on mortality was greater for males (HR=2.5) than females (HR=1.5). As U was positively associated with stroke and/or survival, individuals with larger values of U were more likely to die, resulting in a biased sample consisting of those with smaller values of U. Therefore, the distribution of U would provide insights into the magnitude of survivor bias.

Data simulation and analyses

Under each scenario, we generated 1000 datasets, each consisted of 100,000 individuals and were followed to 95y. In each simulated dataset, survival data were generated for each age band (0-1y, 1-5y and every 5y thereafter) according to a piecewise Cox proportional hazard model:

\[ H_{i,t}^{\text{survival}} = \frac{h_{i,t}^{\text{baseline}}}{\bar{h}_{i,t}^{\text{survival}}} \exp\left(1 \cdot gender_i + 2 \cdot U_i + 3 \cdot gender_i \cdot U_i + 4 \cdot \text{strokehistory}_{i,t}\right) \]

where \( H_{i,t}^{\text{survival}} \) denotes the survival hazard for individual \( i \) in band \( t \), \( h_{i,t}^{\text{baseline}} \) denotes the baseline survival hazard and \( 1, t \) denotes the sex difference in survival in band \( t \) (same for all individuals); 2 and 3 denote the effect of U and interaction between U and sex on survival respectively; 4 represents the influence of stroke history on survival.

We first generated the survival hazard \( H_{i,t}^{\text{survival}} \), and then survival time \( T_{i,t}^{\text{survival}} \) by an inverse variable transformation under the exponential failure time model(21). If \( H_{i,t}^{\text{survival}} \) was greater than the interval length, individual \( i \) survived band \( t \) and the simulation continued in the next band (\( t+1 \)), otherwise, individual \( i \) died in this band (\( t \)).

Similarly, the first incident stroke (after 45y) was generated according to the piecewise Cox proportional hazard model:

\[ H_{i,t}^{\text{stroke}} = \frac{h_{i,t}^{\text{baseline}}}{\bar{h}_{i,t}^{\text{stroke}}} \exp\left(1 \cdot gender_i + 2 \cdot U_i\right) \]

where \( H_{i,t}^{\text{stroke}} \) denotes the stroke hazard for individual \( i \) in band \( t \), \( h_{i,t}^{\text{baseline}} \) denotes the baseline stroke hazard in band \( t \) (fixed at log(0.005)) represents the male-female difference in stroke incidence (i.e., 5 per 1000 person-year for all age bands); 2 denotes the effect of U on first incident stroke. Similar to the survival model (1), we first generated the stroke hazard \( H_{i,t}^{\text{stroke}} \), and then the time to the first incident stroke \( T_{i,t}^{\text{stroke}} \) by an inverse variable transformation. Conditional on individual \( i \) surviving previous band (\( t-1 \)), if \( T_{i,t}^{\text{stroke}} \) was <5y, individual \( i \) experienced the first incident stroke, otherwise, he/she survived the band stroke-free. Details about the data generating process are provide in Supplementary materials (A3).

Once the survival and stroke data were simulated, we derived the cumulated death and survival rates and stroke incidence rates for males and females in each age band from 45y under four causal scenarios. The stroke incidence rate during a specific age band was calculated as the total number of events (i.e. first incident stroke) divided by the total person-year at risk. Individuals at risk refer to those who were alive and had not experienced the first incident stroke at the beginning of the age band.

All metrics were derived for each dataset, and then averaged over all datasets under each causal scenario. Data simulation and analyses were performed in R (version 3.6.1).

Results

In our simulated cohort, median survival time was 77.5y for females and 70y for males. For females, cumulative survival to ages 45, 55, 65, 75, and 85y differed little under four causal scenarios (Figure 2). Percentages of females surviving to all ages were only slightly lower under scenarios B and D (U influenced mortality, HR =1.5) compared to those under scenarios A and C (i.e. U did not influence mortality). For males, cumulative survival differed little under scenarios A, B and C (HR=1.0, 1.5 and 1.5 respectively) and was lower under scenario D (U had a greater effect on mortality than scenarios A-C, HR=2.5). Under scenario D, 86.4%, 74.9%, 55.0%, 29.8%, and 8.8% of males survived to 45, 55, 65, 75, and 85y respectively, compared to 90.5%, 81.3%, 63.9%, 38.4%, and 12.4% under scenario A. Similar patterns were seen for cumulative mortality rates by sex and scenario (Supplementary Figure A4.1 and Table A4.1).
Distribution of U

Figure 3 illustrates the mean value of U for individuals at risk for the first incident stroke (i.e. were alive and free of stroke at the beginning of each age band) under each causal scenario. For scenario A (with minimum bias), mean value of U remained 0 (pre-specified mean in simulation) until 50y. It declined slightly to below 0 thereafter as individuals with smaller U values were less likely to have a stroke and more likely to remain stroke free at older ages. The reduction was slightly greater in males due to their greater risk for stroke. For scenarios B and C, mean values for males dropped substantially after 45y since U was positively associated with both stroke and death; for females, mean value dropped more rapidly in scenario B (U was associated with both stroke and death) compared to scenario C (U was associated with stroke only). Under scenario D, mean value of U dropped to below 0 at an earlier age, more rapidly thereafter, especially for males due to their strong association with death (HR=2.5). Thus, survivor bias was introduced when U influenced both stroke and survival, was greater when the influence was greater for males. Males at risk for stroke at older ages had smaller U values than females, thus were less likely to have stroke (Figure 3). Sex-specific mean values of U at all ages under four causal scenarios are provided in Supplementary Table A4.3.

Stroke incidence rate

Figure 4 shows the incidence rate for the first incident stroke in 10-year age band from 45 to 95y. Under scenario A, the male-female difference in stroke incidence rate was close to the pre-specified 5 per 1000 person-year (no survivor bias) at most ages, became smaller and was biased downwards slightly after 80y. Although U did not directly affect mortality, there still existed small survivor bias at old age due to the indirect effect of U on death (via stroke). For example, stroke was more prevalent in males. Individuals with smaller U were less likely to have stroke, and more likely to survive (since stroke increases risk of death). Under scenario B, the association between U and mortality had the same magnitude in males and females. The male-female difference in stroke incidence was close the pre-specified value until age 75y, reduced to 0 (no sex difference) by 90-95y (Figure 4).

Survivor bias was greater for scenarios C and D when there was an interaction effect between U and sex on mortality. Under scenario C (U affected survival only in males), the difference declined slightly from 45 to 65y, rapidly from 75y, and reversed at age 80-85y. Under scenario D (greater effect of U on mortality for males than females: HR=2.5 vs 1.5), the male-female difference reduced throughout adult life, more rapidly after 65y, and reversed earlier at ~80y. For example, compared to females, stroke incidence rate in males was 5 per 1000 person-year greater at 45y, but was 9 per 1000 person-year lower at 95y. The difference dropped faster in scenario C than scenario D after 85y.

Figure 5 shows the male-female difference across age for four scenarios. Since any deviation from the pre-specified 5 per 1000 person-year can be depicted as survivor bias, scenario A had the minimum bias, while scenarios C and D had the highest bias. Our observation about the survivor bias of different magnitudes under scenarios A-D were consistent with the differential distributions of U (Figure 3). Detailed simulation results in terms of the sex-specific stroke incidence rate at all age bands under causal scenarios A-D are provided in Supplementary Table A4.2.

Additional Causal Scenarios

Simulations under additional causal scenarios E and F were conducted to validate our findings. In causal scenario E, the HR of U on death was 2.5 for males and 1.0 for females; while in causal scenario F, the respective HRs were 1.0 and 1.5. In scenario E, survivor bias was even greater than in scenarios C and D: the direction of the male-female difference in stroke incidence reversed at 75y (earlier than 80-85y in scenario C and 75-80y in scenario D); and the male-female difference reduced to -18 (per 1000 person-year) at 95y, compared to -12 in scenario C and -9 in scenario D. For causal scenario F, however, since U affected death in females only, the male–female difference in stroke incidence was biased upwards and increased from the pre-specified 5 (per 1000 person-year) at 45y to 18 (per 1000 person-year) at 95y (Supplementary Figures A5.1, A5.2 and A5.3).

Discussion

In our simulated cohort, males were observed to have a higher stroke incidence than females in most adult life, but the male-female difference attenuated in late life. The reduction was attributable to selective survival. Moderate bias was observed when there was an unobserved construct U (i.e. common risk factors for stroke and survival) and the effect of U on the hazard of mortality was the same for males and females. Large survivor bias was found when the magnitude of the effect differed by sex.

The sex disparities in stroke incidence have been reported in many studies. In the Oxford Vascular Study where 91,106 individuals were followed from 2002 to 2005(10), the relative risk (RR) of ischemic stroke for males (vs females) decreased from 1.45 before age 65y to 1.0 after 65y. In a population-based study of 15,739 individuals in Scotland during 1986-2005(21), the RR of first stroke hospitalization for males (vs females) decreased from 1.54 at 55-65y to 1.06 at age 85y after adjusting for years of admission and socioeconomic deprivation. In the Rotterdam study which followed 7721 participants who were free of stroke at baseline for an average of six years, males had higher incidence rates for hospitalized stroke than females between 55-85y, and lower rates after 85y(15). In addition, a national representative study in Spain found that males below 75y had higher rates of acute episodes of stroke, and lower rates after 75y compared to females(17). However, none of the above-mentioned studies considered the issue of selective survival and therefore, the results might be subject to survivor bias. The sex disparities in stroke incidence reported could be over-estimated, especially studies including individuals of advanced ages because of their high death rate.

Few studies have investigated the magnitude of survivor bias in stroke epidemiology. One study investigated the sizes of survivor bias in estimating the difference in stroke incidence between White and Black adults in the US using a simulation approach(22). Their findings suggest that selective survival explains the age attenuation of racial inequalities in stroke incidence. Our simulated cohort had a balanced male-to-female ratio. We explored comprehensive causal scenarios based on the unobserved construct U that affected survival of males and females at various magnitudes, and thus, provides an intuitive guide for assessing the size of survivor bias in the sex inequalities of stroke incidence.
The issue of survivor bias applies to other cardiovascular outcomes. For example, age-adjusted stroke mortality was reported to be 13% lower in black females than black males in the US according to the WONDER database and the sex-differences decreased with age after 55y(7). In a global assessment of aging during 1980-2010, mortality rates from cardiovascular diseases were higher in males than females, and this difference decreased with age(23). Similar patterns of differences in age-specific estimates have been reported for the associations of high BMI(24) and blood pressure(25) with mortality. The sex inequality was also reported in the case fatality, post-acute care and recovery of stroke(11, 26–33). Biologically, reasons for sex disparities in cardiovascular diseases are complex. It has been suggested that sex steroid hormones may partly explain the higher risk of stroke in males than females during most adulthood. However, limited human studies evaluating the effect of hormones on stroke, and studies assessed the impact of estrogen therapy on cardiovascular outcomes, including stroke among postmenopausal women revealed inconsistent findings(34–37). Therefore, understanding the true sex disparities in stroke epidemiology have important clinical and public health significance.

There are strengths and limitations associated with our simulation study. The survival and stroke rates by sex and age band were calibrated using data from real-world studies to make the hypothetical cohort more "realistic". According to clinical observations, we allowed both instantaneous and history of stroke to be associated with death. Therefore, survivor bias under each causal scenario can be compared with the true value. To better illustrate the effect of selective survival, we assumed that the case-fatality rate and the effect of stroke history on death were equal for males and females, which may not always be true in some clinical settings and for some types of stroke(26). However, this has little impact on our findings since the marginal all-cause mortality was calibrated with respect to the US 1959–61 life table.

In summary, males had higher stroke incidence overall, but the male-female difference attenuated with age (or even reversed) because of selection bias. While males have a higher mortality rate than females, males survive to older age may represent a more advantageous group with a lower risk for incident stroke compared to females. With the growing ageing population, understanding the extent to which the age attenuation of sex disparities represents the elimination of risk factors in males or survivor bias will inform appropriate stroke prevention, treatment and care for females and males. Our simulated birth cohort shows that the magnitude of survivor bias is determined by the unobserved common risk factors for stroke and survival, especially the effect of their interactions with sex on survival. Appropriate statistical analyses are crucial for disentangling the sex inequalities in stroke (as well as other clinical outcomes) from the artifact of selective survival.

Declarations

**Ethics approval and consent to participate:** For this simulation study, data were completely simulated, which did not require approval from the ethics committee or consent from participants. All methods were carried out in accordance with relevant guidelines and regulations. The mortality rates by age group used in our simulation were obtained from the US National Centre for Health Statistics and are publicly available from the website: https://www.cdc.gov/nchs/data/dvs/lewk3_2006.pdf.

**Consent for publication:** Not applicable.

**Availability of data and materials:** Methods of data simulation are included within the article. The R-code for simulation is provided in online supplementary materials. The simulated datasets analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests:** The authors declare that they have no competing interests.

**Funding:** XL is supported by the Shanghai Sailing Program (19YF1402900) and the General Projects of Shanghai Science and Technology Commission (21ZR1405000). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of Shanghai Commission of Science and Technology.

**Authors' contributions:** LL and XL conceptualized the study and designed the analytical strategy. XL performed the analysis. LL and XL wrote the manuscript. Both authors have read and approved the final manuscript.

**Acknowledgements:** The authors would like to thank Mingyue Gao for her help to produce some of the figures in this paper.

**Authors' information:** 1 School of Data Science, Fudan University, Shanghai, China. 2 University College London Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK. WC1N 1EH.

**References**


**Figures**

**Scenario A**

- U → Stroke → Death
- Sex

**Scenario B**

- U → Stroke → Death
- Male: HR=1.5
- Female: HR=1.5

**Scenario C**

- U (Female: HR=1.0) → Stroke → Death (Male: HR=1.5)
- Sex

**Scenario D**

- U (Female: HR=1.5) → Stroke → Death (Male: HR=2.5)
- Sex

**Figure 1**

Four causal scenarios in the simulation study†

† Scenario A: U is not associated with death; Scenario B: U is associated with death with a similar magnitude (HR=1.5) for males and females; Scenario C: U is associated with death for males, not females (HR=1.5 vs 1.0); Scenario D: association of U with death is greater for males than females (HR=2.5 vs 1.5). Associations between sex and death, sex and stroke, stroke and death, U and stroke remain the same across four scenarios.
Figure 2

Percentages of females and males surviving to each age in the simulated cohort under four causal scenarios, and differences in percentages between Scenarios B-D and Scenario A*

* Differences in percentages correspond to y-axis to the right. As values were similar for some scenarios, for females, red curve represents difference between scenarios C and A; blue curve represents differences between scenarios B/D and A. For males, red curve represents differences between B/C and A; blue curve represents difference between D and A.
Figure 3
Mean of U for those at risk for the first incident stroke, from birth to age 95y under causal scenarios A to D

![Figure 3](image)

Figure 4
Stroke incidence rate for males and females at 5y age bands in the simulated cohort under four causal scenarios

![Figure 4](image)

Figure 5

![Figure 5](image)
Difference in stroke incidence rate (per 1000 person-year) between males and females for 5y age bands in the simulated cohort under four causal scenarios. Pre-specified male – female difference is 5 per 1000 person-year and marked by the black dash line

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary28Dec21submitted.docx