Importance of Sex and Age Factor in Assessing Family History of Stroke.

Tomohiro Saito 1, Seiichiro Nanri 2, Ikuo Saito 2, and Toshihito Furukawa 3

Incidence of stroke differs between men and women and it increases nearly exponentially with age. Therefore, assessment of family history of stroke disregarding sex and age of family members results in bias or misclassification. In this study the effects of sex and age on the positivity of the past history were analyzed numerically.

Sex- and age-specific proportion of a positive history of stroke among 24,007 family members was obtained from a questionnaire survey of 2,316 high school students. By analyzing the sex- and age-specific proportion with the logistic regression model odds ratios resulting from sex and age difference were estimated. The odds ratio for sex difference was 2.458 (95% confidence interval: 2.067-2.924) and odds ratio for age difference was 1.064 (95% confidence interval: 1.058-1.070). This indicated that a positive history of stroke was 2.458 times higher in male members than in female members of the same age and that a positive history increased by (1.064)^y, where y was age difference in years. Potential bias or misclassification resulting from disregarding sex and age can be substantial. Some measures to control for sex and age of family members are required in assessing the family history.

J Epidemiol, 2000 ; 10 : 328- 334

stroke, family, risk

INTRODUCTION

Family history was proposed as one of the risk factors for stroke 3. But its role in stroke is not unequivocal 5. In some studies a significant association was found between the family history and morbidity or mortality of stroke 3-10, while in others such an association was not found 13-17. The disagreement may derive from complexities of the genetic role and differences in the study design. The definition of a positive family history also differs: presence of a family member with the past history of stroke 3,5-7,9-11,14-17; presence of a family member who developed the disease under a certain age 3. The range of family members also differs, sometimes limited to parents 7,8,12 and sometimes extending to uncles and aunts 9,12. Analytical methods employed were a twin study approach 13, the use of the life-table method 6, a survival statistic called a modified Gehan statistic 8, and a comparison based on a positive family history 3,5-7,9-11,14-17. The last method has been used most frequently but it has a serious drawback: age of family members is not considered.

Sex and age are major risk factors to be considered in epidemiology 18. This notion is also pertinent to the assessment of family history as a risk factor. Since the incidence of stroke increases steeply, almost exponentially with age, and it differs between men and women 19, sex and age of family members need to be controlled in assessing the family history. Failure to do so leads to bias or misclassification in risk assessment and results in distortion of a true familial association. In many studies in which a positive family history was used as an index of familial clustering, neither sex nor age of family members was controlled 5,6,9,10,14,16. In some studies sex of family members was considered 3,7,11,17, but age of family members was controlled in few studies 12,18 apart from the aforementioned
studies in which some statistical techniques were employed.

Under these situations, the purpose of this study was threefold. First, to present sex- and age-specific proportion of a positive history of stroke, which has been rarely available. This proportion is different from prevalence of disease and more pertinent in family history studies in which information on deceased family members is included. Prevalence of disease does not include deceased persons in the denominator. The second purpose was to express, for the first time, the effects of sex and age on the positivity of past history numerically. The third purpose was to call attention to the potential bias or misclassification resulting from disregarding sex and age of family members. This was done through the quantitative analysis of the sex-and age-specific proportion of a positive history of stroke.

**SUBJECTS AND METHODS**

Sex- and age-specific proportion of a positive history of stroke was obtained from a questionnaire survey of high school students. The questionnaire contained information on the parents, grandparents and uncles and aunts; the collected data included the present age or age at death, and age at onset, by decade, of stroke which had been diagnosed or treated by physicians. The questionnaire was handed to 3,145 students at school and filled in at home by parents. The questionnaire was returned by 2,316 students (a return rate of 73.6%). Two uncle-aunt families were most frequent on both the paternal and maternal sides. Among the total of 25,139 family members listed in the returned questionnaires, 24,188 (96.2%) members had full information necessary in this study. Of these members 181 were below 30 years of present age or age at death and they were excluded from the following analyses. Thus, data on the remaining 24,007 family members were used in the calculation of the sex- and age-specific proportion of a positive history of stroke and in the logistic analysis.

First, to calculate the sex- and age-specific proportion of a positive history, the 24,007 members were stratified into 10-year age intervals in both sexes either by present age or age at death. Those whose present age or age at death was in their 30s and 40s were grouped together because of a small number of members who had developed stroke. As the representing age for this group, the weighted average by the numbers of family members in the two age groups (30s and 40s) were taken: 43 years of age for male and 44 years for female groups. For the same reason — that is small numbers of study subjects — the age intervals of 90 or more were combined with the interval of 80-89, and its representing age was 86 for the male and 85 for the female group. Second, in each age interval in both sexes the number of those with a positive history of stroke and the number of those without a past history of stroke were obtained. Thirdly, age-specific proportion of a positive history was calculated for both sexes by the following: [the number with a positive history] / (the number with a positive history + the number without a past history)]. Fourthly, the following logistic regression model was formulated:

\[
\log(p / (1-p)) = \text{intercept} + b(\text{sex}) + c(\text{age}),
\]

where \( p \) was the probability of a positive history, and the age was either present age or age at death. From the logistic model, the parameters \( b \) and \( c \) were estimated. Then, odds ratios were obtained by an exponent of the estimated \( b \) or \( c \) together with its 95% confidence interval by an exponent of \([b \pm 1.96 \times \text{Standard error of } b \text{ or } c]\). An interaction between sex and age was also examined, followed by further logistic analyses. The goodness of fit of the logistic regression model was assessed by the Hosmer and Lemeshow goodness of fit test.

The linearity of the explanatory variable was assessed by plotting the logit, \( \log(p / (1-p)) \), against the age scale. Also, the following logistic analysis was performed separately in both sexes to examine the log-linearity of the age-specific proportion of a positive history. The logistic model was the following:

\[
\log(p / (1-p)) = \text{intercept} + c_1(Dage_1) + c_2(Dage_2) + c_3(Dage_3) + c_4(Dage_4),
\]

where \( Dage_1, Dage_2, Dage_3, \) and \( Dage_4 \) were dummy variables taking the value of either 0 or 1 for age intervals of 30-49, 50-59, 60-69, 70-79 and 80-99.

The calculations were performed by the PC-SAS and the logistic analysis was carried out by the procedure LOGISTIC.

**RESULTS**

Figure 1 shows the sex- and age-specific proportion of a positive history of stroke in both sexes. The vertical axis for the proportion is in a log scale. In the male members the proportion is linear from the youngest age interval to the 60s, followed by a flat line afterwards. In the female members the proportion is linear until age 70 and after 70 it was linear with a gentler slope.

The effects of sex and age on a positive history estimated by the logistic analysis were the following. The odds ratio for sex difference in the whole age range was 2.458 with its 95% confidence interval of 2.067-2.924. This indicates that a positive history of stroke was 2.458 times higher in the male members than in the female members of the same age. The odds ratio for age difference was 1.064 with its 95% confidence interval of 1.058-1.070. This indicates that a positive history of stroke increased by 1.064 with an increase of age of family members by one year. Odds ratios resulting from age difference of y years were obtained by \((1.064)^y\). An odds ratio for a 5-year difference was 1.364 and for a 10-year age difference it was 1.860. The goodness of fit test showed a p value of less than 0.05. This was mainly due to age-by-age fluctuations in a positive history in younger family members because of small numbers of members with a positive history.
Figure 1. Sex- and age-specific proportion of a positive history of stroke.

The vertical axis for the proportion of a positive history is on the log scale. The mid-years are taken for the 10-year intervals from age 50 to 79, age 43 for male members and 44 for female members for the age group of less than 50, and age 86 for male and 85 for female members for the age group of 80 or more on the age axis (● male members; ○ female members).

The logits, log(p/(1-p)), plotted against the age scale in both sexes were very similar to the proportion curves shown in Figure 1. Like the curves in Figure 1, the logit curve for men was quite straight in the age range from age 40 to 70. After age 70 it became nearly flat. The logit curve for women was quite straight from 40 to 70 and after 70 it was also quite straight, though the slope became gentler.

The results of the logistic analysis employing the dummy variables for age intervals were the following. In men odds ratios for the age intervals of 50-59, 60-69, 70-79 and 80-99 against the age interval of 30-49 were 4.46, 16.43, 18.34 and 16.69 respectively. Thus, below age 70 the assumption of log-linearity on the proportion of a positive history seemed reasonable. In women the odds ratios for the same age intervals were 3.68, 9.31, 11.49 and 16.55 respectively. Again below age 70 the assumption of the log-linearity seemed acceptable.

Since the logit curves for both sexes, as can be speculated from the curves in Figure 1, were not completely parallel, an interaction between sex and age was evaluated by adding an interaction term in the logistic analysis. The interaction was not statistically significant (p>0.05) including unity in its 95% confidence interval. However, we did separate analysis for each sex and for each age interval to examine the data closely.

The results of separate analyses are presented in Table 1. The age intervals in the table were the intervals extending over two adjacent age intervals. Namely, the logistic regression analysis was done separately for the 4 intervals made up of two adjoining points in Figure 1. For each interval the logistic regression model including only age variable was carried out in the male and female groups separately and this results are shown in the rows “From age difference.” The logistic regression model including sex and age variable was also carried out and this results are shown in the rows “From sex difference.”

<table>
<thead>
<tr>
<th>Age interval</th>
<th>30-59</th>
<th>50-69</th>
<th>60-79</th>
<th>70-99</th>
</tr>
</thead>
<tbody>
<tr>
<td>From age difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.148</td>
<td>1.104</td>
<td>1.010</td>
<td>0.996</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.106-1.190</td>
<td>1.079-1.129</td>
<td>0.990-1.030</td>
<td>0.972-1.020</td>
</tr>
<tr>
<td>Female</td>
<td>1.167</td>
<td>1.058</td>
<td>1.027</td>
<td>1.027</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.107-1.229</td>
<td>1.022-1.095</td>
<td>0.993-1.062</td>
<td>0.988-1.068</td>
</tr>
<tr>
<td>From sex difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>1.415-2.936</td>
<td>2.290-3.994</td>
<td>2.337-3.568</td>
<td>2.052-3.262</td>
</tr>
<tr>
<td>Goodness of fit(p)</td>
<td>0.0547</td>
<td>0.0041</td>
<td>0.5285</td>
<td>0.4217</td>
</tr>
<tr>
<td>Interaction between sex and age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>0.972</td>
<td>1.044</td>
<td>0.983</td>
<td>0.970</td>
</tr>
</tbody>
</table>
| CI: Confidence interval.
The results of the Hosmer and Lemeshow goodness-of-fit test for the model including both sex and age are shown in the row “Goodness of fit.” Interaction between sex and age was assessed in each age interval and these results are shown in the last row “Interaction between sex and age.” In both sexes the odds ratios for age difference tended to decrease with the advancement of age and after the age interval of 50-69 they were not statistically significant (p>0.05) including unity in their 95% confidence intervals. In the age intervals of 30-59 and 50-69 the odds ratios were fairly large resulting in an increase of positive history at least 1.640 times in males and 1.326 times in females with an increase of age by 5 years. The odds ratios for sex difference were above 2.0 in all the age intervals and they were all statistically significant (p<0.05) with their 95% confidence intervals not including unity. In other words the proportion of positive history was more than twice in male members compared with that of female members all through the age.

The goodness of fit test and the interaction between sex and age became statistically significant (p<0.05) only in the age intervals of 50-69. This was mainly due to yearly fluctuations in a positive history and a non-parallel increase in a positive history between male and female members. This results do not necessarily make the estimated odds ratios unjustifiable but make some reservations in the interpretation of the numerical figures.

What can be said from these results are the following. The effects of sex difference on a positive history was substantial all through the age. The effects of age on a positive history was fairly large before age 70 in both sexes, but after age 70 the effect was small particularly in male members. Although a close look at the sex- and age-specific proportion of a positive history of stroke and detailed analyses provided different odds ratios from sex difference and age difference, the odds ratio of 2.458 from sex difference and 1.064 from age difference for the whole age range stood as summary statistics.

**DISCUSSION**

The results of this study indicated that disregarding sex and age of family members in assessing family history of stroke would lead to a substantial misclassification or bias. Control for sex should be necessary for all family members and control for age should be required for family members below age 70. The difference in the proportion of a positive history between male and female members and the odds ratios from sex difference were not small. In assessing the family history as a risk factor this point needs to be considered. A risk for a person whose father, an uncle or grandfather has a positive history and a risk for a person whose mother, an aunt or grandmother has a positive history should be different. The latter should be given a higher risk. A woman who developed the disease at a certain age is regarded to carry a heavier risk for the disease than a man who developed the disease at the same age. This is a strong message led from the results of this study. A control for age within 5 years will be necessary when risk ratios of other factor are less than 1.5.

The actual degree of misclassification or bias depends on the study type and the genetic role working in this disease. The results would be biased in a case-control study if the sex and age of family members are disregarded. Suppose that a case's grandfather with a past history of stroke is older than a control's grandfather without a past history of stroke. This assessment of the family history as a risk factor may be biased if the control's grandfather later develops stroke at the same age with the case's grandfather. If a case's family member with a past history of stroke is male and all other family members in that case and control have no past history, a similar bias may enter into the assessment of risk. In a cohort study the results would be diluted towards no association due to misclassification. “Exposure” here is a positive history of stroke. Suppose that age of family members of a target person in the exposed group is older than age of family members of a counterpart in the non-exposed group. If a family member of the counterpart in the non-exposed group later develops stroke at the same age with the target person in the exposed group, the ascertainment of exposure may have been misclassified. However, under uncertainty about the genetic role in stroke the degree of actual bias or misclassification in a particular study cannot be directly estimated from the results of this study.

Nevertheless, the results will be helpful and useful in planning in what way and to what extent sex and age should be controlled. An age difference larger than 5 years or even 10 years will not be infrequent in family history studies, as conceived from the distributions of age of marriages and age of mothers at delivery 80. A similar misclassification may enter into the assessment of exposure, a risk factor, when sex of a family member with a positive history is disregarded. One of the possibilities for controversial results in the role of genetics in cerebrovascular disease may be due to the lack of control for sex and age among family members in such analyses.

Before discussing implications of this study, the validity of the family history data and its effects on the odds ratio need to be stated. Firstly, accuracy of the collected information on the family history: whether the information obtained through questionnaire reflects true evidence. Since the questionnaire was filled in at home by parents, present age and age at death of the study family members must be fairly reliable. It is unlikely that parents did not know the present age or age at death of their brothers, sisters and parents — uncles, aunts and grandparents in the questionnaire. It is also unlikely that the parents — mostly in their 40s and early 50s — did not know whether their brothers and sisters suffered from stroke and the onset by decade age if they did. The grandparents in this study were mostly in their 60s and 70s by present age or age at death. Their sons and daughters — parents in the questionnaire —
most likely knew whether their parents had developed such a serious disease as stroke and its onset by decade age. Overall, the sex- and age-specific proportion of a positive history of stroke used in this study should not be grossly inaccurate.

To obtain the assurance of the above speculation, we examined the precision of the family history questionnaire 20). Precision means agreement among repeated data collection. The questionnaire was administered twice among one-fourths of the study population with a one year interval. The proportion of contradicting answers — namely interchange between presence and absence of a positive history, and discrepancy in age at onset — between the two surveys was low. It was below 0.8% among the parents, uncles and aunts, and between 1.8% and 6.9% among the grandparents. Inaccuracy, namely discrepancy between answers and true evidence, is difficult to confirm. From these results, it can be said that misclassification in the past history would not have distorted the results of our study so greatly as to endanger the conclusion of the present study. This statement will be supported by the following simulation.

The effects of possible inaccuracy in the family history on the odds ratio were examined by simulation. If the inaccuracy in the age-specific proportion of a positive history occurred to the same degree in all the age groups, the estimated odds ratio would change little. If the inaccuracies occurred to a greater degree in young age groups, the estimated odds ratio would become smaller. But this would have been less likely. What was more likely was that the inaccuracies occurred to a greater degree in elder age groups, that is among grandparents. The inaccuracies would be under-reporting of a positive history of stroke, more likely among the deceased grandparents. The effects of this type of inaccuracy were evaluated assuming that the under-reporting in the age group of 70 or more was 20%. In other words, the odds ratio was estimated by increasing the proportion of a positive history of stroke in that age group by 20%. The change in the odds ratios was very small: less than 1% for the age difference and virtually no change for the sex difference. The estimated odds ratios in this study were fairly robust and would not be endangered by the inaccuracies of a possible degree.

Another question to be considered is the type of stroke. Differential diagnosis between ischemic, embolic or hemorrhagic stroke was not made clear in the questionnaire. In most studies on the family history of stroke 3, 5-7, 9, 11-16, this distinction was not made. A few studies made this distinction in the family history 4, 8, 10, 17. But if this distinction is to be made in the family history assessment, odds ratios need to be estimated separately for both types based on respective sex- and age-specific proportion of a positive history. Age-specific incidence of cerebral infarction increases steeply with age, whereas those of intracerebral hemorrhage and subarachnoid hemorrhage increase less steeply with age 20). A potential misclassification or bias from age difference would be larger in the ischemic type than in the hemorrhagic type.

The results of this study can be extrapolated to other countries where the incidence of stroke is higher or lower than that in Japan. In this connection the relation between incidence and the proportion of a positive history need to be stated. The proportion of a positive history is conceptually close to cumulative incidence, the proportion of a fixed population that becomes diseased in a stated period of time 20). A major difference between them is that the proportion of a positive history includes in the denominator persons who died from other causes, while dead persons are out of consideration in cumulative incidence. Cumulative incidence is premised on the assumption of no competing risks of death 20). The relation between cumulative incidence and incidence is not so simple algebraically but cumulative incidence approximates $\sum I_i \Delta t_i$, when $I_i$ is small 20). Here $I_i$ denotes age-specific incidence and $\Delta t_i$ denotes a time interval.

From the above two relations it can be said that in countries where the incidence or the prevalence of stroke is high, the proportion of a positive history among family members is also high. The reverse also holds for low rates. Since the proportion of a positive history is not readily available, cumulative incidence calculated from incidence may be used as a substitute. If the slope of the age-specific cumulative incidence is steeper, misclassification or bias from age difference becomes larger. If the slope is less steep, it is smaller. If male and female cumulative incidence curves are apart, misclassification or bias from sex difference becomes large.

In addition to space, time also counts. Since incidence of stroke has decreased in Japan 20), proportion of a positive history of stroke must have also decreased. The proportion among family members should have been higher a few decades ago and will be lower in the coming decades than that in this study. In this sense the results only apply to the urban Japan in 1990s. The odds ratios for age difference and for sex difference depend on the slope of the proportion against age and not on the magnitude of the proportion itself. This point is described already in “Discussion.” A comparison of odds ratios for age difference and for sex difference in different times is of interest. The approach employed in this study is very useful in the comparison as numerical figures are obtained. If data on a wide range of time are available, a logistic analysis including time as an explanatory variable is worth carrying out and will yield interesting results.

When family history of stroke is assessed as a risk factor in epidemiologic studies or in clinical practice for a patient, the points raised in this study and the magnitude of a possible bias or misclassification need to be considered. In epidemiologic studies some measures to control sex and age of family members should be employed. The measures may be stratification, matching, multivariate analysis or the use of some risk indices. In stroke age was partially controlled by defining a positive history as having had stroke under a certain age 10. The life-
the estimate was for the offspring as a whole and not for a
probable risk of diabetes in the offspring was computed using
Ministry of Health and Welfare, Japan.

lishment of healthy lifestyle from childhood” from the
risk factor for a patient.

error must be carefully examined if family history is taken as a
family history is almost always taken. The possibility of such
cation need to borne in mind also in clinical practice, in which
ly history and the degree of the potential bias and misclassifi-
individual offspring in a particular family.

the Kaplan-Meier product limit estimate 33). But in that study
index instead of a positive history 32). Also in diabetes the
sex and age, and these adjusted levels were used as a genetic
the fasting glucose levels of family members were adjusted by
maximal age of 85 years. In another study of diabetes mellitus
empirical data of age at onset of the disease, under the postula-
tion that all the family members would reach the assumed
maximal age of 85 years. In another study of diabetes mellitus
the fasting glucose levels of family members were adjusted by
sex and age, and these adjusted levels were used as a genetic
index instead of a positive history 30). Also in diabetes the
probable risk of diabetes in the offspring was computed using
the Kaplan-Meier product limit estimate 30). But in that study
the estimate was for the offspring as a whole and not for an
individual offspring in a particular family.

The necessity for considering sex and age in assessing fami-
ly history and the degree of the potential bias and misclassifi-
cation need to borne in mind also in clinical practice, in which
family history is almost always taken. The possibility of such
error must be carefully examined if family history is taken as a
risk factor for a patient.

ACKNOWLEDGEMENTS

Supported in part by a grant-in-aid “Research on the estab-
lishment of healthy lifestyle from childhood” from the
Ministry of Health and Welfare, Japan.

REFERENCES

1. Dyken ML, Wolf PA, Barnett HJM, et al. Risk factors in
stroke: a statement for physicians by the Subcommittee
on Risk Factors and Stroke of the Stroke Council. Stroke,
2. Rastenyte D, Tuomilehto J, Sarti C. Genetics of stroke -
3. Gifford AJ. An epidemiological study of cerebrovascular
4. Marshall J. Familial incidence of cerebrovascular dis-
5. Alter M, Kluznik J. Genetics of cerebrovascular acci-
6. Khaw KT, Barrett-Connor E. Family history of stroke as an
independent predictor of ischemic heart disease in men
stroke victims associated with early cardiovascular mor-
9. Brass LM, Shaker LA. Family history in patients with trans-
risk factors for ischemic stroke in Italian men.
11. Graffagnino C, Gasecki AP, Doig GS, Hachinski VC.
The importance of family history in cerebrovascular dis-
and risk of stroke. A prospective follow-up of 14371
middle - aged men and women in Finland. Stroke, 1997;
28: 1361-1366.
13. de Faire U, Friberg L, Lundman T. Concordance for
mortality with special reference to ischaemic heart dis-
gease and cerebrovascular disease; a study on the Swedish
15. Diaz JF, Hachinski VC, Pederson LL, Donald A. Aggregation of multiple risk factors for stroke in siblings
of patients with brain infarction and transient ischemic
dence and risk factors for stroke in Copenhagen, Den-
Stroke, 1993; 24 : 1366-1371.
18. MacMahon B, Pugh TF. Epidemiology — Principles and
103-136.
disease in the community; results of a WHO collaborative