

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Distribution of Risk Factors Differ from Coronary Heart Disease and Stroke in China: A National Population Survey
AUTHORS	Li, Yapeng; Yin, Ling; Shen, Ying; Wang, Anran; Zhou, Lue; Gao, Yuan; Li, Yusheng; Lu, Jie; shi, songhe; Tian, Chuansheng; Xu, Yuming; Wang, Longde

VERSION 1 – REVIEW

REVIEWER	Lautsch, Dominik Merck & Co., Inc., Center for Observational and Real-World Evidence
REVIEW RETURNED	11-Jul-2022

GENERAL COMMENTS	<p>1) The authors tackle relevant research, given the rate of strokes is higher in China than in many other countries. The research question to identify risk factors that are more closely associated with stroke over others more closely associated with CHD is an important one. I congratulate the researchers on their noteworthy findings. The manuscript is overall good, but needs additional work.</p> <p>MAJOR 2) Kindly provide additional data on how the CNSSS is representative.</p> <p>MAJOR 3) The data stem from a survey - can you add on quality control (section 2.4.): were there clinical monitors who assessed the information entered in select offices or another way to confirm correctness of entered data?</p> <p>MAJOR 4) How were the risk factors selected? You could have ended up with far more variables given the size of the study and apparent completeness of the dataset. Stats overall need to be clearer.</p> <p>5) Implications: How can "more attention" be paid to prevention of either risk factor if both are very similar in prevention, maybe lower treatment targets in patients with enhanced risk of stroke? Is this what you are referring to?</p> <p>6) Given this is hypothesis generating research as there is no additional cohort to confirm your finding (which is often the case in novel research), could you publish your statistical methods including the variable selection, so that other researchers could apply your methods in a different dataset?</p> <p>7) Figures 2 and 3 AND abstract(!): Please indicate that the 1 line separates enhanced risk from CHD ($0 < 1$) over stroke (> 1). This is not clear and needs to be added.</p>
-------------------------	--

REVIEWER	Zhong, Chongke Soochow University
REVIEW RETURNED	15-Sep-2022

GENERAL COMMENTS	<p>This is a very interesting study. The authors explore the distribution differences of common risk factors between CHD and stroke in China. The advantage of the study is the large sample size, over 40 years old of study participants, and the risk factors collection as more as possible. The strength of this study relies on the large sample size and the representativeness of the data. Overall, it is a well-written paper, especially with the clear flow of the sampling. However, there are some comments that the authors need to consider.</p> <p>1. In the section 2.3, please mention clearly how were CHD and stroke defined, by confirmed with medical records (medical insurance system), or just confirmed by the subjects by their recall, i.e. saying Yes/No. So as other data of risk factors. Recall bias should be considered and discussed deeply.</p> <p>2. In the section 2.6, "Secondly, people with both CHD and stroke were excluded, as well as those neither with CHD nor with stroke, we just took the samples only with CHD and the samples only with stroke as our analysis population, then taking CVD as dependent variables (stroke was defined as 1, whereas CHD was defined as 0) and the common risk factors as independent variables, multivariate logistic regression analysis was carried out to study the distribution difference of the common risk factors between CHD and stroke." CVD was defined as cardiovascular disease in line 56, not cerebrovascular disease. So why is stroke defined as 1 and coronary heart disease defined as 0? Why did not include patients without CHD and stroke to state as a control group?</p> <p>3. In the section 3.1, "Of these, we excluded data from project areas with a response rate less than 85%, incomplete baseline information and abnormal data." This sentence is confusing. Response rate is the attribute of the whole survey, not the attribute of a single research object. The authors cannot exclude subjects based on the response rate.</p>
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Dominik Lautsch, Merck & Co., Inc.

Comments to the Author:

1) The authors tackle relevant research, given the rate of strokes is higher in China than in many other countries. The research question to identify risk factors that are more closely associates with stroke over others more closely associated with CHD is an important one. I congratulate the researchers on their noteworthy findings. The manuscript is overall good, but needs additional work. MAJOR 2) Kindly provide additional data on how the CNSSS is representative.

Response: We appreciate the reviewer's important comment. The CNSSS used a third-stage stratified cluster sampling method. In the first stage, 128 prefecture-level cities were selected from all 31 provinces in China mainland according to the different proportion of the Sixth National Population Census of China in 2010. In the second stage, one urban street and one rural town were selected from each prefecture-level city, respectively. In the third stage, an urban community and a rural village were selected from each urban street and rural town. All residents aged 40 years or older were surveyed during the primary screening, and the response rate of each place was required to be no

less than 85%. Ultimately, 828 764 subjects from 256 communities and villages participated in the survey. We have revised these contents in the section 2.1.

MAJOR 3) The data stem from a survey - can you add on quality control (section 2.4.): were there clinical monitors who assessed the information entered in select offices or another way to confirm correctness of entered data?

Response: We thank the reviewer for pointing this out. In the on-site survey, each sub-center had a trained staff, usually a neurologist, responsible for the review and reporting of data. Epidemiologists and statistical experts were organized by the project office to analyze the data reported by the sub-centers, and responsible for checking abnormal data and returning to the sub-centers. We have added these contents in the section 2.4.

MAJOR 4) How were the risk factors selected? You could have ended up with far more variables given the size of the study and apparent completeness of the dataset. Stats overall need to be clearer.

Response: We appreciate the reviewer's concerns. Due to the limitation of project funds, the project office conducted full discussion when designing the questionnaire and selected common variables with appropriate cost for investigation. Some items with high cost, such as serum homocysteine and carotid artery ultrasound, were only tested in high-risk of CVD and were not included in this study.

5) Implications: How can "more attention" be paid to prevention of either risk factor if both are very similar in prevention, maybe lower treatment targets in patients with enhanced risk of stroke? Is this what you are referring to?

Response: We apologize for our obscurity. We agree with the reviewer's opinion that there should be different treatment targets of different CVD types, such as a lower LDL-C target value in patients with CHD. In addition, what we want to express is that more specific prevention and control measures should be formulated to prevent stroke and CHD. The heterogeneity in the association between particular risk factors and specific CVD types demonstrated in the present study could improve the selection of high-risk patients for population-based screening programs.

6) Given this is hypothesis generating research as there is no additional cohort to confirm your finding (which is often the case in novel research), could you publish your statistical methods including the variable selection, so that other researchers could apply your methods in a different dataset?

Response: We appreciated the reviewer's constructive comment. The statistical methods including the variable selection are available upon reasonable request from the corresponding authors. We look forward to the validation of our results in other datasets.

7) Figures 2 and 3 AND abstract(!): Please indicate that the 1 line separates enhanced risk from CHD ($0 < 1$) over stroke (> 1). This is not clear and needs to be added.

Response: We thank the reviewer for pointing this out. We have added these contents in the Figures 2 and 3.

Reviewer: 2

Dr. Chongke Zhong, Soochow University

Comments to the Author:

This is a very interesting study. The authors explore the distribution differences of common risk factors between CHD and stroke in China. The advantage of the study is the large sample size, over 40 years old of study participants, and the risk factors collection as more as possible. The strength of this study relies on the large sample size and the representativeness of the data. Overall, it is a well-written paper, especially with the clear flow of the sampling. However, there are some comments that the authors need to consider.

1. In the section 2.3, please mention clearly how were CHD and stroke defined, by confirmed with medical records (medical insurance system), or just confirmed by the subjects by their recall, i.e. saying Yes/No. So as other data of risk factors. Recall bias should be considered and discussed deeply.

Response: We appreciate the reviewer's concerns. We are very sorry that we have not described clearly in the methods section. The diagnosis of CHD and stroke were confirmed by professional doctors, based on self-reported history and medical records. The risk factors were diagnosed by combination of self-report history, current use of drugs, and results in the on-site measurement. We have revised these contents in the section 2.3.

2. In the section 2.6, "Secondly, people with both CHD and stroke were excluded, as well as those neither with CHD nor with stroke, we just took the samples only with CHD and the samples only with stroke as our analysis population, then taking CVD as dependent variables (stroke was defined as 1, whereas CHD was defined as 0) and the common risk factors as independent variables, multivariate logistic regression analysis was carried out to study the distribution difference of the common risk factors between CHD and stroke."

CVD was defined as cardiovascular disease in line 56, not cerebrovascular disease. So why is stroke defined as 1 and coronary heart disease defined as 0? Why did not include patients without CHD and stroke to state as a control group?

Response: We appreciate the reviewer's concerns. Due to the main objective of this study, which was to explore "the distribution differences of Risk Factors for Coronary Heart Disease and Stroke", the population only with CHD and the samples only with stroke were our target population. So in this section, after excluding those both with Stroke and CHD, as well as those neither with Stroke nor with CHD, we just took the samples only with CHD and the samples only with stroke as our analysis population, defined the disease classification (Stroke=1, CHD=0) as dependent variables, and use binary logistic regression analysis to detect the different association of the various influence factors between the Stroke and CHD.

3. In the section 3.1, "Of these, we excluded data from project areas with a response rate less than 85%, incomplete baseline information and abnormal data."

This sentence is confusing. Response rate is the attribute of the whole survey, not the attribute of a single research object. The authors cannot exclude subjects based on the response rate.

Response: We appreciate the reviewer's concerns. We are very sorry that we have not described clearly in this section. The CNSSS is a cluster sampling survey adopting multi-stage stratified sampling method, and the response rate of each project area was required to be no less than 85%. The response rate is a very important indicator of the quality of the survey. In the final analysis, we excluded data from project areas with a response rate less than 85%. Finally, 725 707 people were included in the study.

VERSION 2 – REVIEW

REVIEWER	Lautsch, Dominik Merck & Co., Inc., Center for Observational and Real-World Evidence
REVIEW RETURNED	09-Nov-2022

GENERAL COMMENTS	Excellent work - congrats for addressing this important topic.
-------------------------	--

REVIEWER	Zhong, Chongke Soochow University
REVIEW RETURNED	03-Nov-2022

GENERAL COMMENTS	The authors have well addressed my comments.
-------------------------	--