

## 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

<b>TITLE (PROVISIONAL)</b>	Long-term trends and regional variations of hypertension incidence in China: a prospective cohort study from the China Health and Nutrition Survey, 1991-2015
<b>AUTHORS</b>	Luo, Yunmei; Xia, Fan; Yu, Xuexin; Li, Peiyi; Huang, Wenzhi; Zhang, Wei

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Cynthia Bell McGovern Medical School at UTHealth-Houston, USA
<b>REVIEW RETURNED</b>	19-Aug-2020

<b>GENERAL COMMENTS</b>	<p>The authors have reported the hypertension incidence in China over a 20+ year period based on a well designed dynamic cohort study. The findings show hypertension incidence has increased in recent years compared to the earlier cohort in the mid-1990s. In addition, the population in China has changed over these years and an adjusted analysis showed that some population characteristics are driving this increase in HTN incidence through independent associations with some of these changes (Eastern region, older age, higher BMI, more Han ethnicity, and higher education, specifically). The data is presented well but I suggest including more details about the analysis specifics and editing some wording to make the findings more clear.</p> <p>Strengths and limitations of the study, Page 2</p> <ul style="list-style-type: none"><li>- The first and second statement seem to be in conflict. Specifically the first statement says "individuals from diverse social and geographic contexts..." but the third statement says "did not include individuals from other provinces in China". Did you mean to say from "all provinces in China"?</li></ul> <p>Figure 1</p> <ul style="list-style-type: none"><li>- Wording on right column, 3rd bubble down is confusing. I understand the second part "had hypertension at first investigation" means excluding prevalent cases but I am unsure what "...individuals with <math>\leq 1</math> time of hypertension observation" means?</li><li>- Wording on right column, last bubble down also doesn't make any sense? Are these more prevalent cases? When are they "suffering from hypertension"? How are these prevalent cases different from the right column, 3rd bubble down.</li><li>- Please include an additional row in Figure 1 showing those that are missing covariate information as explained in the footnote of table 2</li></ul>
-------------------------	--

	<p>In general, the wording "suffer" from hypertension seems a bit extreme but this is a minor point.</p> <p>Statistical Analysis, page 8</p> <ul style="list-style-type: none"> <li>- person-years is defined for those who get hypertension but how is it defined for those who never get hypertension? What about those who die? How is death or lost-to-follow up dealt with and defined?</li> <li>- Please explain your age-standardization method more in line 41</li> <li>- line 46, what do you mean by "while considering geographic variations" with the cox proportional model? You considered many factors not just geographic regions, right? Was something special about how geographic region was included in the model?</li> </ul> <p>Patient and Public Involvement, page 9</p> <ul style="list-style-type: none"> <li>- You state members of the public were not involved but I am assuming that the subjects in the study recruited from the general public, right?</li> <li>- I like table 1, I wish more studies would show longitudinal data this clearly</li> </ul> <p>Results, page 10</p> <ul style="list-style-type: none"> <li>- In the text, you mention only a few differences in sample characteristics seen over the years but in reality there are many differences that reflect the social development and migration changes in China over the last 20+ years. I would add at least one more sentence that highlights that these findings are in line with the overall population changes in China, specifically migration towards to urban Eastern region, more Han population, more male, less smoking, and less activity (the last 2 likely related to the higher BMI as well as income levels).</li> <li>- Page 13 and Table 4 define time between "free from hypertension to hypertension" but, as mentioned in a previous comment, the time is also defined for those who never got hypertension, correct? The wording here is confusing and does not accurately represent the calculation of person-time or incidence.</li> <li>- I do not find Table 4 very helpful as there is confounding between amount of time followed and when a subject entered the cohort (and all the covariates confounded with cohort entry time - later cohort means older, higher BMI, and less time to follow up). So it means less to me that within 2 years 14% were hypertensive because likely the majority of this finding is driven by those in the later cohorts who were more likely to be hypertensive. Similar logic in opposite direction for the low proportion in those with longer follow up.</li> </ul> <p>Table 5</p> <ul style="list-style-type: none"> <li>- please explain in the methods what exactly is meant by the footnote on this table that says "**Estimated time effect of age, p&lt;0.001"?</li> </ul> <p>Page 20</p> <p>Last sentence makes no sense "In addition, we did not distinguish hypertension, and future research is necessary"</p>
--	---

<b>REVIEWER</b>	Jiapeng Lu Fuwai hospital, China
<b>REVIEW RETURNED</b>	28-Aug-2020

<p><b>GENERAL COMMENTS</b></p>	<p>The authors analyzed data from a population-based cohort study and aimed to explore the trends of hypertension incidence and regional variations in China from 1991 to 2015. The major finding was that the incidence of hypertension showed increasing trends in Chinese adults from 1991 to 2015. I have a few comments the authors may wish to consider:</p> <p>Major comments:</p> <ol style="list-style-type: none"> <li>1. The age distributions were largely different in three study periods. More younger people were recruited in 1991-1997 and 2000-2009. It could result in higher incidence of hypertension in 2011-2015 compared with the first two study periods. As authors reported, the crude incidence in 2011-2015 was 48.6/1000 person-years, but only 31.3 in 1993-1997 and 36.4 in 2000-2009. It might influence on the comparability of the incidences in different study periods although authors reported the age-standardized rates.</li> <li>2. The definition of incident hypertension in this study was self-reported hypertension or systolic blood pressure <math>\geq 140</math> mmHg or the diastolic blood pressure <math>\geq 90</math> mmHg. Self-reported data may result in the underestimation of incidence. Why antihypertensive medication usage was not included in the definition of incident hypertension?</li> <li>3. The results of cox model suggested that the risk of hypertension was higher in people recruited in the 2000-2009 and 2011-2015 compared with those recruited in 1991-1997 in individual level, but it can't support the increasing trends of hypertension incidence in population-level.</li> <li>4. The authors mentioned that the blood pressure was detected in triplicate by professional health workers on the same day in the method section. The guidelines recommended that hypertension is diagnosed by using blood pressure values measured in different days. This limitation should be discussed in the manuscript.</li> </ol> <p>Minor comments:</p> <ol style="list-style-type: none"> <li>1. Blood pressure measurement should be described in detail, such as procedure of the measurement, devices used and values record and calculation, etc.</li> <li>2. Some descriptions are not suitable or clear, such as the setting in the abstract, page 21 lines 34-36: "we did not distinguish hypertension, and future research is necessary".</li> </ol>
--------------------------------	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer 1's comments to the Author

Comment#1:

Strengths and limitations of the study, Page 2

- The first and second statement seem to be in conflict. Specifically the first statement says "individuals from diverse social and geographic contexts..." but the third statement says "did not include individuals from other provinces in China". Did you mean to say from "all provinces in China"?  
 Author's response: Thank you for the meaningful suggestion on our manuscript. The third statement indeed means from "all provinces", to avoid ambiguity of the third statement, we revised it as follows: We did not employ a national-representative sample and did not include individuals from all provinces in China, which undermined the representation of our findings. As a community-based survey, CHNS

excluded institutionalized individuals, which further diminished the representation of our findings among Chinese. Please refer to pages 2 and 20.

Comment#2:

Figure 1

- Wording on right column, 3rd bubble down is confusing. I understand the second part "had hypertension at first investigation" means excluding prevalent cases but I am unsure what "...individuals with  $\leq 1$  time of hypertension observation" means?

Author's response: Thanks for your careful check. When calculating the hypertension incidence, we need at least two observations from each individual to identify the newly diagnosed hypertension case and the length of this period, i.e., person-year. This method was also adopted in the prior study [reference 25].

Reference:

Gao M, Kuang W, Qiu P, et al. The time trends of cognitive impairment incidence among older Chinese people in the community: based on the CLHLS cohorts from 1998 to 2014. *Age Ageing* 2017; 46: 787-93.

Comment#3:

- Wording on right column, last bubble down also doesn't make any sense? Are these more prevalent cases? When are they "suffering from hypertension"? How are these prevalent cases different from the right column, 3rd bubble down.

Author's response: Thanks for your careful check. We have revised the content as "10 612 observations after developing hypertension". As we focused on the newly diagnosed hypertension cases, the observations after the identification/diagnosis would be excluded. For example, when someone, free of hypertension in 1991, was diagnosed with hypertension in 1993, the observations after 1993 would be excluded, as the individual's outcome has been observed.

Comment#4:

- Please include an additional row in Figure 1 showing those that are missing covariate information as explained in the footnote of table 2

Author's response: Thank you for the meaningful suggestion, we revised the title of figure 1 accordingly: Flowchart showing the selection of the subjects who were included in the final analysis of hypertension incidence in China, with covariate information missing rate of 9.78%. Because low missing rate of each covariate and no one miss much information at the same time, we did not exclude any individual, so we only revised the figure caption. Please refer to page 24.

In addition, because there was no hypertension data in wave 1989, and it would not contribute to the hypertension incidence and model, we exclude the data of 1989, so we revised the last column of figure 1 as: 12 952 individual, 53 703 observations.

Comment#5:

In general, the wording "suffer" from hypertension seems a bit extreme but this is a minor point.

Author's response: It's very kind of you to make this meaningful suggestion. We change the word "suffer from" to "develop", that maybe more suitable. Please refer to page 8.

Comment#6:

Statistical Analysis, page 8

- person-years is defined for those who get hypertension but how is it defined for those who never get hypertension? What about those who die? How is death or lost-to-follow up dealt with and defined?

Author's response: Thanks for your careful check. We calculate the incidence of hypertension as below [reference 25]:

individuals free of hypertension all the time would still contribute to the denominator. As individuals had at least two observations, they would stay in the cohort for the period of time, i.e., person-years, and contribute to a part of denominator, albeit they would die or be lost to follow-up. The observations would terminate after loss to follow-up or death, and individuals would not contribute to the denominator after that.

For the modeling analyses, because we do not have death data, we could not fit competing risk model to separate the effect of death on risk of hypertension. But Cox proportional hazard model has been used to deal with the attribution bias from loss to follow-up.

Comment#7:

- Please explain your age-standardization method more in line 41

Author's response: Thank you. We have added the footnote of table 3 to explain age-standardization method and provided more details in the method section.

Methods: "we conducted direct standardization to calculate the age-standardized incidence of hypertension by using the study sample from wave 2011 and wave 2015 as the standard population". Please refer to page 7.

Footnote: "Age-standardized incidence was calculated using the study sample in 2011-2015 as the standard population". Please refer to page 13.

Comment#8:

- line 46, what do you mean by "while considering geographic variations" with the cox proportional model? You considered many factors not just geographic regions, right? Was something special about how geographic region was included in the model?

Author's response: Thank you for pointing out this confusing description, we indeed considered many factors not just geographic regions, so we revised this sentence as: "To further evaluate the long-term trends and geographic variations of incident hypertension, we performed an extended Cox proportional hazard model while including all covariates to control for baseline variations." Please refer to page 8.

Comment#9:

Patient and Public Involvement, page 9

- You state members of the public were not involved but I am assuming that the subjects in the study recruited from the general public, right?

Author's response: Thank you for pointing this out. We derived data from CHNS, which is publicly available. The CHNS recruited participants from communities they selected. We rewritten the "Patient and public involvement" section: Not applicable. Data are derived from public domain. Please refer to page 8.

Comment#10:

- I like table 1, I wish more studies would show longitudinal data this clearly

Author's response: Thanks for your affirmation.

Comment#11:

Results, page 10

- In the text, you mention only a few differences in sample characteristics seen over the years but in reality there are many differences that reflect the social development and migration changes in China over the last 20+ years. I would add at least one more sentence that highlights that these findings are in line with the overall population changes in China, specifically migration towards to urban Eastern region, more Han population, more male, less smoking, and less activity (the last 2 likely related to the higher BMI as well as income levels).

Author's response: Thank you for the meaningful suggestion on our manuscript. According to these comments, the paragraph 2 of "Study population" section has been rewritten as follow:

Table 2 presents the baseline characteristics of the study sample. Overall, variations existed in all baseline characteristics. Newly recruited individuals were older ( $P < 0.001$ ) and well-educated ( $P < 0.001$ ). They were more likely to be obese ( $P < 0.001$ ), Han ( $P < 0.001$ ), and male ( $P < 0.001$ ), and they were less likely to be smokers ( $P < 0.001$ ), employed ( $P < 0.001$ ), and physically active ( $P < 0.001$ ). Please refer to page 9.

Comment#12:

- Page 13 and Table 4 define time between "free from hypertension to hypertension" but, as mentioned in a previous comment, the time is also defined for those who never got hypertension, correct? The wording here is confusing and does not accurately represent the calculation of person-time or incidence.

Author's response: Thank you for pointing this out and we agree with you. In page 13, we provided a brief calculation of the duration from free of hypertension to hypertension identified in our cohort, the time here is particularly for those identified cases, excluding those never with hypertension in our cohort. We have revised the sentence as: "Among the identified cases, the duration from free of hypertension to incident hypertension ranged from 2 to 24 years, with a median of 9 years". Please refer to page 13.

In Table 4, the time represents the timing of entering the cohort, according to which all participants were clustered into three mutually inclusive group: 1991-1997, 2002-2009, and 2011-2015 (as we provided in the "exposures" of method section, page 6). We revised the terminology in Table 4 as "timing of entering the cohort" to improve the readability. Please refer to page 15.

Comment#13:

- I do not find Table 4 very helpful as there is confounding between amount of time followed and when a subject entered the cohort (and all the covariates confounded with cohort entry time - later cohort means older, higher BMI, and less time to follow up). So it means less to me that within 2 years 14% were hypertensive because likely the majority of this finding is driven by those in the later cohorts who were more likely to be hypertensive. Similar logic in opposite direction for the low proportion in those with longer follow up.

Author's response: Thank you for pointing this out and we agree with you. We have removed the details of table 4 and relevant description content. Please refer to pages 13-14.

Comment#14:

Table 5

- please explain in the methods what exactly is meant by the footnote on this table that says "\*\*Estimated time effect of age,  $p < 0.001$ "?

Author's response: We have performed statistical test to verify whether every character met hypothesis of proportional hazards, the results suggesting that age didn't not met this hypothesis, so we fit 'time-dependent cox regression model'. That "\*\*Estimated time effect of age,  $P < 0.001$ " means age as a time-dependent variable, these was elaborated in page 8, "statistical analysis" section.

Comment#15:

Page 20

Last sentence makes no sense "In addition, we did not distinguish hypertension, and future research is necessary"

Author's response: We are very sorry for our incorrect writing. We have changed this sentence as follows: "we did not distinguish the grade of hypertension, and future research is necessary". Please refer to pages 3 and 20.

Reviewer 2's comments to the Author

Comment#1:

The age distributions were largely different in three study periods. More younger people were recruited in 1991-1997 and 2000-2009. It could result in higher incidence of hypertension in 2011-2015 compared with the first two study periods. As authors reported, the crude incidence in 2011-2015 was 48.6/1000 person-years, but only 31.3 in 1993-1997 and 36.4 in 2000-2009. It might influence on the comparability of the incidences in different study periods although authors reported the age-standardized rates.

Author's response: It's common and undeniable that age distribution might vary across study periods in cohort studies. Therefore, our study as well as many prior ones employed standardization methods to cope with the variations, whereby we calculated age-standardized incidence by using the same standard population. In our case, we used the study sample in the last period to account for the variations. The method was also used in prior studies as below:

Reference:

- Gao M, Kuang W, Qiu P, et al. The time trends of cognitive impairment incidence among older Chinese people in the community: based on the CLHLS cohorts from 1998 to 2014. *Age Ageing* 2017; 46: 787-93.
- Geiss LS, Wang J, Cheng YJ, et al. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. *JAMA* 2014; 312: 1218-26.

Moreover, in our multivariate analyses, we also adjusted for other confounders such as BMI, race, educational attainment, which may also change over time. Results provide robust evidence of the increasing pattern.

Comment#2:

The definition of incident hypertension in this study was self-reported hypertension or systolic blood pressure  $\geq 140$  mmHg or the diastolic blood pressure  $\geq 90$  mmHg. Self-reported data may result in the underestimation of incidence. Why antihypertensive medication usage was not included in the definition of incident hypertension?

Author's response: Thank you for pointing this out. In China, the criteria to identify hypertension were  $\geq 140/90$  mmHg and (or) past history of hypertension, criteria studies focused on the national prevalence of hypertension in China in 1991, 2002, 2012, and 2015 was  $\geq 140/90$  mmHg and (or) taken drugs in the least two weeks, but we only have the data of this question: "Are you currently taking anti-hypertension drugs?" with high missing rate. Therefore, we choose the criteria of Guidelines for the Prevention and Treatment of Hypertension of China ( $\geq 140/90$  mmHg); at the same time, we adopted self-reported hypertension according to a prior study as below:

Reference: Li J, Shi L, Li S, et al. Urban-rural disparities in hypertension prevalence, detection, and medication use among Chinese Adults from 1993 to 2011. *Int J Equity Health* 2017, 16: 50.

Comment#3:

The results of cox model suggested that the risk of hypertension was higher in people recruited in the 2000-2009 and 2011-2015 compared with those recruited in 1991-1997 in individual level, but it can't support the increasing trends of hypertension incidence in population-level.

Author's response: In the Cox proportion hazard model, estimates were adjusted for the age, socioeconomic, and lifestyle attributes. The estimates of the timing of entering the cohort could be interpreted as: when individuals shared the same age, socioeconomic status, and lifestyle attributes, individuals that entered the cohort in later waves had higher risks of hypertension. In this case, individuals entering the cohort in later waves could be those born later, as they were conditional at the same age with those entering the cohort earlier. In addition, we also construct multi-level Poisson regression as sensitivity analyses (results are provided below), both two models indicate the

increased time trend of hypertension incidence. In Poisson analyses, time was identified as the time when each observation was recorded, which represents the actual timing of developing hypertension. However, as Cox model has dealt with the attribution bias, we provided the results in the primary analyses and mentioned the details about the Poisson sensitivity analyses in the Methods section. Please refer to page 8.

Table 1 Poisson analysis of hypertension incidence

Characteristic	aRR (95% CI)	P value
Time		
1993-1997	Ref.	
2000-2009	1.47 (1.32-1.63)	<0.001
2011-2015	1.50 (1.33-1.69)	<0.001
Geographic region		
Western	Ref.	
Central	1.21 (1.10-1.33)	<0.001
Northeastern	1.62 (1.46-1.80)	<0.001
Eastern	1.39 (1.26-1.54)	<0.001

Note: Time in this table represents the time each observation was recorded. All other covariates have been adjusted, while results are not provided here.

Moreover, we calculated age-standardized incidence in Table 2. Incidence is an indicator that reflect new cases in a population during a certain period. Therefore, incidence measures in Table 2 have demonstrated the increasing pattern of hypertension trajectories in the population level, and our multivariate analyses provided addition information after controlling for the changes in socioeconomic and life-style behaviors.

Comment#4:

The authors mentioned that the blood pressure was detected in triplicate by professional health workers on the same day in the method section. The guidelines recommended that hypertension is diagnosed by using blood pressure values measured in different days. This limitation should be discussed in the manuscript.

Author's response: Thanks for your valuable advice. We have added one statement in the strengths and limitations of this study section: Guidelines recommend to identify hypertension cases by using blood pressure values that are measured in different days, while individuals' blood pressure data in the CHNS were collected on the same day in the CHNS, leading to unavoidable bias. Please refer to pages 3 and 20.

Comment#5:

Blood pressure measurement should be described in detail, such as procedure of the measurement, devices used and values record and calculation, etc.

Author's response: CHNS was initiated in 1989, the information regarding survey design, data collection, and quality control could be retrieved from the literature that elaborate the cohort profile, so for the simplicity of the article, we only cited related reference of No.22 (last sentence of page 5, data source section). Please refer to page 5.



Comment#6:

Some descriptions are not suitable or clear, such as the setting in the abstract, page 21 lines 34-36: "we did not distinguish hypertension, and future research is necessary".

Author's response: Thank you for pointing this out. We have changed this sentence as follows: "we did not distinguish the grade of hypertension, and future research is necessary". Please refer to pages 3 and 20.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Cynthia Bell McGovern Medical School
<b>REVIEW RETURNED</b>	28-Oct-2020

<b>GENERAL COMMENTS</b>	<p>The authors have satisfactorily responded to most of my comments but a few issues still remain:</p> <ul style="list-style-type: none"><li>- I do not think the statement on "CHNS excluded institutionalized individuals" needs to be included on page 2 Strengths and limitations of the study. Most studies do not include institutionalized individuals so this lack of generalizability should not be a main focus. Stating not all provinces on page 2 is enough. But I would keep the full 2 sentences on page 20.</li><li>- Figure 1 statement could be further re-stated as "12681 individuals with <math>\leq 1</math> hypertension observation..."</li><li>- The authors have stated that "covariate information missing rate of 9.78%" but that "no one miss much information at the same time" in their reviewer responses. But this is not totally clear and needs to be explained clearly in the text. The analysis appears to use 2 datasets. I dataset of all individuals (n=12,952) to determine incidence rates of hypertension over time and then another complete case dataset of individuals with no missing data (n=11,685) in the covariate adjusted analysis. Again, I recommend another bubble on Figure 1 stating these 2 separate datasets used in the analysis.</li><li>- The authors responded to my comments on exactly how follow-up time was defined for subjects who did not have hypertension. Specifically stating that "The observations would terminate after loss to follow-up or death...". They should state this in the methods as well as noting that death data was not available so censoring could be for death or lost to follow up.</li></ul>
-------------------------	---

<b>REVIEWER</b>	Jiapeng Lu Fuwai hospital, China
<b>REVIEW RETURNED</b>	28-Oct-2020

<b>GENERAL COMMENTS</b>	No further comments.
-------------------------	----------------------

## VERSION 2 – AUTHOR RESPONSE

Reviewer 1's comments to the Author

Comment#1:

I do not think the statement on "CHNS excluded institutionalized individuals" needs to be included on page 2 Strengths and limitations of the study. Most studies do not include institutionalized individuals so this lack of generalizability should not be a main focus. Stating not all provinces on page 2 is enough. But I would keep the full 2 sentences on page 20.

Author's response: Thank you for the meaningful suggestion, we had removed the sentence: "As a community-based survey, CHNS excluded institutionalized individuals, which further diminished the representation of our findings among Chinese" in the "Strengths and limitations of this study" section but kept the full two sentences on page 20. Please refer to pages 2 and 20.

Comment#2:

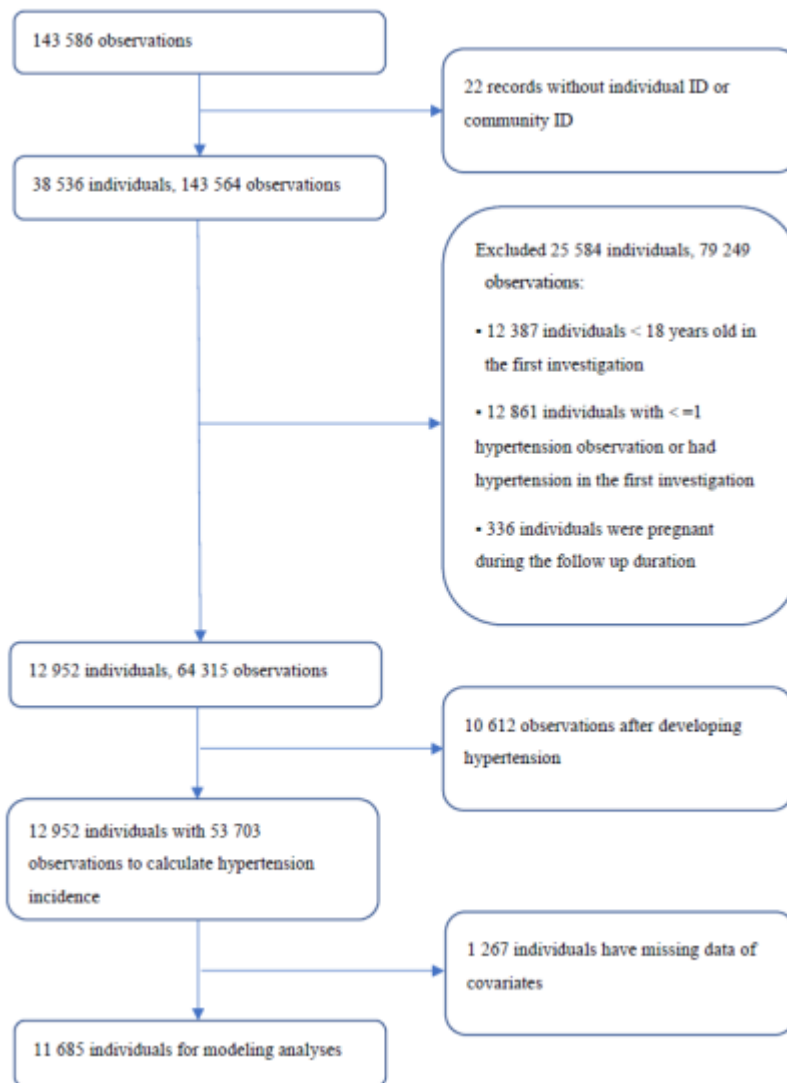
Figure 1 statement could be further re-stated as "12681 individuals with  $\leq 1$  hypertension observation..."

Author's response: Thank you for pointing this out and we agree with you. We changed the sentence "12 861 individuals with  $\leq 1$  time of hypertension observation or had hypertension in the first investigation" of the second right column to "12 861 individuals with  $\leq 1$  hypertension observation or had hypertension in the first investigation". Please refer to image attachment.

Comment#3:

The authors have stated that "covariate information missing rate of 9.78%" but that "no one miss much information at the same time" in their reviewer responses. But this is not totally clear and needs to be explained clearly in the text. The analysis appears to use 2 datasets. 1 dataset of all individuals (n=12,952) to determine incidence rates of hypertension over time and then another complete case dataset of individuals with no missing data (n= 11,685) in the covariate adjusted analysis. Again, I recommend another bubble on Figure 1 stating these 2 separate datasets used in the analysis.

Author's response: Thank you for pointing this out and we agree with you, according to your advice, we have added another two bubbles on Figure 1, as follows:



**Comment#4:**

The authors responded to my comments on exactly how follow-up time was defined for subjects who did not have hypertension. Specifically stating that "The observations would terminate after loss to follow-up or death...". They should state this in the methods as well as noting that death data was not available so censoring could be for death or lost to follow up.

Author's response: Thank you. We have added the censoring information in "Study design and exclusion criteria" section as follows: As death data were not available in CHNS, censoring could be for death or lost to follow up. Please refer to page 6.