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)	Comparison of incidence, rate and length of all-cause hospital admissions between adults with normoglycaemia, impaired fasting glucose and diabetes: a retrospective cohort study in Geelong, Australia
AUTHORS	Sajjad, Muhammad; Holloway, Kara; de Abreu, Lelia; Mohebbi, Mohammadreza; Kotowicz, Mark; Pedler, Daryl; Pasco, Julie

VERSION 1 – REVIEW

REVIEWER	Dr Duong Tran Centre for Big Data Research in Health, UNSW Sydney
REVIEW RETURNED	11-Nov-2017

GENERAL COMMENTS	Thank you for the opportunity to review the manuscript. This paper is well-written and potentially contributes to the field. My major comments are below 1. Description of outcomes: While the authors clearly described the glycaemic groups and covariates, description about the outcomes requires clarification. Firstly, the hospitalisation outcome was analysed as a Yes/No outcome variable, and as a count variable (number of admissions during follow-up) these were not described in the method section. Secondly, were planned admissions included? Thirdly, the length of stay (LOS) calculation: was it calculated based on dates of admission and discharge, was it the LOS for each admission or total LOS for each participant during the follow-up? Forth, reasons of admission: the categorisation of diagnoses seems arbitrary, what classification system was used to categorise these diagnoses (body systems?) 2. Statistical analyses: The background, statistical analyses, the presentation of results and the stated aims don't go hand-in-hand. I don't follow the Table 2 and the rationale of gender stratification (similar findings). If the effects of SES, demographic or lifestyle are also of interest, then that should be stated in the aims. I don't think logistic regression is the right approach given the varying length of follow-up. More description about the Poisson regression is necessary (as above mentioned, the outcome variable, how to deal with zero inflated data?). In terms of terminology, what is meant by tri-variate? Does it mean a model with 3 variables or a variable (e.g. glycaemic group) with 3 categories? Suggest to use the term: admission rate, admission rate ratio, rather than frequency rate, or incidence rate ratio. BMI, Smoking should also be adjusted for in the multivariable models, especially when the normal glycaemic group had greater proportion of admissions for respiratory conditions. 3. Data linkage: Although the authors cite the protocol of the main

	Osteoporosis Study, it would be useful to describe procedures used to identify admission and death records. Some sorts of reliable identifiable information are required to make sure that questionnaire data of a participant were linked to his/her hospital and death data. 4. The main rationale for this study is around IFG (background section, line 36-38), but established diabetes was the central point for the discussion and conclusion. Further modification of the discussion is worthwhile to strengthen the paper. Can the author comment on IFG progression to diabetes in this sample?

REVIEWER	Daniel Rubin Temple University
REVIEW RETURNED	22-Nov-2017

GENERAL COMMENTS	Summary: In this retrospective cohort study of almost 2,000 Australian community dwelling adults followed over 7+ years, participants with diabetes were twice as likely as normoglycemic participants to be hospitalized, had a higher incidence rate of hospitalization, and longer LOS. These outcomes were not different between IFG and normoglycemic participants. The study is overall valuable, but the presentation could be improved and some items should be clarified. 1) Design should be stated as retrospective cohort study 2) State the units for the LOS quartile and decile differences. Is it days? 3) The value of the tri-variate analyses is unclear. One cannot make conclusions about the effect of individual risk factors without controlling for the other confounders. This analysis and table should probably be deleted. 4) What is described as “frequency” or “frequency rate” of hospitalization appears to be what is usually called an incidence rate. Here, the # of hospitalizations/total person-years. 5) On page 11, why are the incidence rates of hospitalization not compared directly as opposed to using ORs? 6) It would be helpful to readers to have some evidence supporting the assertion that most of the hospitalizations in this cohort were captured. What proportion of the population is served by the smaller local hospitals? 7) Also, loss to follow up is an issue but was not reported. How stable is this population? Are there data on geographic and vital status at the end of followup?

VERSION 1 – AUTHOR RESPONSE

Responses to reviewers-Sajjad et al
Thank you for the opportunity to revise the manuscript. We have made amendments to the manuscript and documented our responses to reviewers’ comments as follows.

Editorial Requirements:
- Please revise your title to state the research question, study design, and setting (location). This is the preferred format for the journal.
Response: We have amended the title to “Comparison of incidence, rate and length of all-cause hospital admissions between adults with normoglycaemia, impaired fasting glucose and diabetes: a retrospective cohort study in Geelong, Australia.”

Reviewer#1
General comment: Thank you for the opportunity to review the manuscript. This paper is well-written and potentially contributes to the field.
Response: Thank you.

Description of outcomes:
1. While the authors clearly described the glycaemic groups and covariates, description about the outcomes requires clarification. Firstly, the hospitalisation outcome was analysed as a Yes/No outcome variable, and as a count variable (number of admissions during follow-up) these were not described in the method section.
Response: The following sentence has been added for clarity (Methods, Outcome measures, paragraph 1),
"Secondary outcomes included admission rate based on the total number of hospital admissions over the follow-up period and length of admission in days, calculated from the admission and discharge dates, considering each admission as a separate occasion."

2. Secondly, were planned admissions included?
Response: Yes. The following words have been added (Methods, Outcome measures, paragraph 1), "planned or unplanned".

3. Thirdly, the length of stay (LOS) calculation: was it calculated based on dates of admission and discharge, was it the LOS for each admission or total LOS for each participant during the follow-up?
Response: Length of stay was calculated based on dates of admission and discharge from the hospital, and each admission was treated as a separate occasion (please refer to comment 1). Furthermore, we have added the following sentence to the "Admission rate" sub-section of "Results", to highlight the proportions of those admitted more than once during follow-up, by glycaemic category, "Overall, 50.6% of the participants with diabetes were admitted more than once over the follow-up period, compared with 30.8% and 22.0% of those with IFG and normoglycaemia, respectively."

4. Forth, reasons of admission: the categorisation of diagnoses seems arbitrary, what classification system was used to categorise these diagnoses (body systems?)
Response: The classification was based on collapsing individual diagnoses into broader categories based on the ICD-10 codes. More specifically, the decimal point was removed and classification was limited to the description before the decimal point. This is described under "Methods", "Primary reasons for hospital admission",
"We classified primary diagnoses into broad categories by aggregating individual disease codes, for instance, primary ICD-10 diagnoses codes of I21.0 'acute transmural myocardial infarction (MI) of anterior wall', I21.1 'acute transmural MI of inferior wall', and I21.4 'acute sub-endocardial MI', were combined as a single category of I21 'acute MI'."

Statistical analyses:
5. (a) The background, statistical analyses, the presentation of results and the stated aims don't go hand-in-hand.
Response: We have revised the "Statistical analysis" sub-section under "Methods" to address reviewers' points. Please refer to highlighted text in the manuscript and individual responses below.

(b) I don't follow the Table 2…
Response: Table 2 explores the impact of potential confounders, one factor at a time, on the incidence, rate and length of hospital admission. The multivariate models may not have enough statistical power to evaluate all risk factors simultaneously, particularly in the 'IFG' sub-group with only 275 men and 159 women, so exploring each confounder at a time enables us to explore potential

impact of confounding with more statistical power. Tri-variate analyses similar to the approach presented in this manuscript are common and are well accepted as analytical methods in the fields of epidemiology and medical statistics (see for example Agresti and Hosmer et al (references below) for applications of tri-variate methods in binary outcomes).
We have described this in the "Statistical analysis" sub-section under "Methods",
"A set of tri-variate analyses (i.e. the outcome and glycaemic status as exposure of interest and one potential confounder) was performed to examine the impact of each potential risk factor above and beyond the glycaemic category association with the study outcomes. We used (i) tri-variate logistic regressions for admission incidence, (ii) tri-variate Poisson regressions for admission rate and (iii) two-way ANOVAs on rank of admission length. Odds ratios, risk ratios and partial eta squared effect size were used to illustrate the impact of potential risk factors, respectively. Partial eta squared values of 0.009, 0.0588 and 0.1379 were considered as benchmarks for small, medium and large effect sizes, respectively."

(c) …. and the rationale of gender stratification (similar findings). If the effects of SES, demographic or lifestyle are also of interest, then that should be stated in the aims.
Response: In response to this comment, we have added the following paragraph to the "Methods" section, under a new sub-section "Potential confounders".
"The risk of hospital admission in diabetes is reported to vary by age, sex, unhealthy weight, cigarette smoking, physical inactivity, and socioeconomic deprivation. In addition, high alcohol use may cause difficulties in management of diabetes, resulting in early onset of complications. Hence, we included these potential confounders in our analyses to investigate the relationship between glycaemic status and hospitalisation outcomes. Furthermore, due to previously reported differences in hospitalisation patterns between men and women with diabetes, we stratified our cohort by sex, in addition to reporting findings for the overall sample."

6. (a) I don't think logistic regression is the right approach given the varying length of follow-up.
Response: Hospital admissions have been investigated in three different ways to cover different aspects of the outcome: admission incidence, admission rate and admission length. For analysing admission incidence (a dichotomised yes/no outcome), logistic regression was used, and to examine admission length, we applied non-parametric ANOVA and linear regression.

(b) More description about the Poisson regression is necessary (as above mentioned, the outcome variable, how to deal with zero inflated data?).
Response: Sensitivity of the Poisson models against any deviations from model assumptions, including zero inflation, was examined by implementing negative binomial regression models and there were negligible changes in RRs, 95% CIs and p-values. This information has now been added to "Methods, Statistical analysis", "Results" and Table 3.

7. In terms of terminology, what is meant by tri-variate? Does it mean a model with 3 variables or a variable (e.g. glycaemic group) with 3 categories?
Response: The terminology refers to three components: (1) outcome (incidence, rate or length of admission), (2) glycaemic status as exposure of interest, and (3) one potential confounder at a time. Please also refer to comment 5(b).

8. Suggest to use the term: admission rate, admission rate ratio, rather than frequency rate, or incidence rate ratio.
Response: We have replaced the words "incidence of all-cause hospitalisation" with "incidence of all-cause hospital admission" or "admission incidence" throughout the document. The term "frequency rate" has been replaced by "admission rate" and "length of stay" with "admission length (days)" or "admission length", throughout the document.

9. BMI, Smoking should also be adjusted for in the multivariable models, especially when the normal glycaemic group had greater proportion of admissions for respiratory conditions.
Response: The impact of BMI and smoking as potential confounders was examined through tri-variate models first and whenever they met the criteria for inclusion in the multivariate models (i.e. when p-value from tri-variate model ≤0.1), they were included in the multivariate models. To elaborate this, the following text has been added ("Methods", "Statistical analysis")
"Multivariate logistic regression was performed to evaluate the association of admission incidence and glycaemic status after adjusting for potential confounders that were significant at 0.1 level in tri-variate analyses and two-way interactions of confounders and glycaemic status; model adjusted OR and 95% CI are reported"

Data linkage:
10. Although the authors cite the protocol of the main Osteoporosis Study, it would be useful to describe procedures used to identify admission and death records. Some sorts of reliable identifiable information are required to make sure that questionnaire data of a participant were linked to his/her hospital and death data.
Response: To identify hospital admission records, we used a unique identifier, referred to as Unit Record number or UR (This information has been added to "Outcome measures" in the "Methods" section)
To identify deaths, a combination of surname, first and second given names, address, date of birth, and date of last contact with the study were used. (This information has been added to the sub-section "Deaths" under "Methods")

11. (a) The main rationale for this study is around IFG (background section, line 36-38), but established diabetes was the central point for the discussion and conclusion. Further modification of the discussion is worthwhile to strengthen the paper.
Response: To our knowledge, no previous studies have explored the relationship between impaired fasting glucose (IFG) and hospital admission, thus comparable data are not available. We have added further information regarding IFG in the "Discussion" section, Paragraph 6, as follows:
"Current rates of IFG-to-diabetes progression are alarmingly high, with studies reporting development of diabetes in up to two-thirds of individuals with pre-diabetes. The authors of this study have previously reported that approximately one-third of Australian women have IFG, with a six-fold higher risk of progressing to diabetes over a decade if FPG ≥ 6.1 mmol/L."
and
"Our findings show that the incidence of hospital admission multiplies as IFG progresses to diabetes, which if used effectively in public health campaigns, could help reduce progression in the population."

(b) Can the author comment on IFG progression to diabetes in this sample?
Response: As our aim was to compare hospital admission outcomes (incidence, rate and length) with glycaemic status at cohort entry, we are unable to comment on IFG-to-diabetes progression. We have, however, previously reported on the predictors of progression for the female arm of the Geelong Osteoporosis Study, as described in the "Discussion" section, paragraph 6. In brief, the incidence of IFG-to-diabetes progression was 18.1 per 1,000 person-years over a decade of life.

Reviewer#2
Summary: In this retrospective cohort study of almost 2,000 Australian community dwelling adults followed over 7+ years, participants with diabetes were twice as likely as normoglycemic participants to be hospitalized, had a higher incidence rate of hospitalization, and longer LOS. These outcomes were not different between IFG and normoglycemic participants. The study is overall valuable, but the presentation could be improved and some items should be clarified.

1. Design should be stated as retrospective cohort study
Response: We have amended the text describing study design to "Retrospective cohort study", in the relevant sections.

2. State the units for the LOS quartile and decile differences. Is it days?
Response: Yes. This has been added to the text and Tables for clarity.

3. The value of the tri-variate analyses is unclear. One cannot make conclusions about the effect of individual risk factors without controlling for the other confounders. This analysis and table should probably be deleted.
Response: Please refer to comment 5(b)

4. What is described as "frequency" or "frequency rate" of hospitalization appears to be what is usually called an incidence rate. Here, the # of hospitalizations/total person-years.
Response: In response to this comment and a suggestion by another reviewer, we have replaced the words "incidence of all-cause hospitalisation" with "incidence of all-cause hospital admission" or "admission incidence" throughout the document. The term "frequency rate" has been replaced by "admission rate" and "length of stay" with "admission length (days)" or "admission length", throughout the document.

5. On page 11, why are the incidence rates of hospitalization not compared directly as opposed to using ORs?
Response: We have changed this to "risk difference" comparison of 'IFG' and 'diabetes' groups with 'normoglycaemia' for a more direct (absolute rather than relative) comparison. On page 11, the text under "Incidence of all-cause hospital admission (admission incidence)" has been amended to,
"Bivariate analyses showed that men with IFG had 10% more admission incidence (risk difference 0.10, 95% CI 0.02-0.17, p=0.006) and men with diabetes had almost 40% more admission incidence (risk difference 0.28, 95% CI 0.17-0.39, p<0.001), compared to men with normoglycaemia. Similarly, women with IFG and diabetes were also more likely to be admitted as compared to normoglycaemia, (risk difference 0.10, 95% CI 0.01-0.18, p=0.024) and (risk difference 0.28, 95% CI 0.16-0.39, p<0.001), respectively."

6. It would be helpful to readers to have some evidence supporting the assertion that most of the hospitalizations in this cohort were captured. What proportion of the population is served by the smaller local hospitals?
Response: This study is based in regional Victoria (Australia), and the University Hospital Geelong (UHG) was the only public tertiary hospital (principal referral hospital) in the region during the study period. This hospital had the only 24-hour Emergency Department over the follow-up period. However, there is a private sector hospital and other smaller hospitals and we have acknowledged as a limitation ("Discussion", paragraph 7, line 12), that some hospital admissions might have been missed.

7. Also, loss to follow up is an issue but was not reported. How stable is this population? Are there data on geographic and vital status at the end of follow-up?
Response: Loss to follow-up was not an issue because we captured information on hospital admissions and deaths using unique identifiers detailed in the "Methods" section. Therefore, even if we lost contact with the participants during follow-up period, we were still able to obtain information regarding their hospital admissions and mortality. Thus, the following has been added to the "Discussion", paragraph 7,
"Finally, we used unique identifiers to capture hospital admissions and mortality data, hence, we were able to obtain this information even if we lost contact with participants over the study period."

However, we have acknowledged the possibility of missing some hospital admissions as a result of participants moving outside of the study region (Discussion, paragraph 7),
"Second, although our study region (BSD) is considered to have a stable population, it is still possible that some of the participants might have moved intercity or interstate during the follow-up period."

References

Agresti, A., 2003. Building and applying logistic regression models. Categorical Data Analysis, Second Edition, pp.211-266.
Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. A comparison of logistic regression and stratified analysis. Applied logistic regression (Vol. 398). John Wiley & Sons.

VERSION 2 – REVIEW

REVIEWER	Daniel Rubin Temple University, USA
REVIEW RETURNED	28-Dec-2017

GENERAL COMMENTS	No further comment.

REVIEWER	Duong Tran UNSW, Australia
REVIEW RETURNED	02-Jan-2018

GENERAL COMMENTS	Thank you for a revision and response to the reviewers comments.