PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Incremental medical cost of delirium in elderly patients with cognitive impairment: analysis of a nationwide administrative database in Japan
AUTHORS	Igarashi, Masakazu; Okuyama, Kotoba; Ueda, Naoya; Sano, Hideki; Takahashi, Kanae; P. Qureshi, Zaina; Tokita, Shigeru; Ogawa, Asao; Okumura, Yasuyuki; Okuda, Shoki

VERSION 1 – REVIEW

REVIEWER	Alba Sanchez University of Potsdam
REVIEW RETURNED	12-Jun-2022

GENERAL COMMENTS	The authors have conducted a cross-sectional study on the relevant topic of the incremental medical cost of delirium in hospitalized patients, with a large sample of patients. My suggestions for improvement and questions for the authors are as follows: - As a suggestion, in point 4, I would clarify that the medical costs associated with delirium are those referred to the period of hospitalization, for example: "The study demonstrates that delirium is associated with significantly higher medical cost during hospitalization"...Since other medium/long-term costs (for example those associated with nursing home placement or those derived from the acceleration of cognitive decline) have not been analyzed in this study. - Page 6, line 57: The authors state "The presence of cognitive decline has been defined as a diagnosis of dementia, prescription of anti-dementia medication, or presence of a low degree of independence in daily life". In the limitations section of the article it is explained that this algorithm is not yet validated, but I would like to know how the low degree of Independence in daily life has been defined. Has it been assessed with the Barthel index? Couldn't the low degree of independence in daily life be due to other reasons such as the presence of neuromuscular/musculoskeletal conditions? Please clarify this point. (Review Checklist '4' and '12') - Page 10, line 43: The authors state "Patients demographics were comparable between the two groups (Table 1)...". Although it is true that the differences between the two groups are not very large, the group with delirium showed more higher percentage of patients with dementia, PIMS, mental and behavioral disorders, emergency hospitalization, ADL dependence... For a better observation of the differences between the two groups, it would be good if you could include effect estimates: when comparing the difference in a continuous variable, please provide the mean or median (if not

	normally distributed) difference and the 95%CI of the difference; when comparing categorical variables, please also provide an effect estimate (e.g., risk ratio, relative risk) and the 95% CI of the effect estimate. (Review Checklist '7') - Page 12: For the duration of hospitalization please provide the median difference and the 95% CI of the difference. (Review Checklist '7') - Page 12/Fig S1: Figure S1 shows the medical costs categorized by patients characteristics, showing the medical costs for emergency hospitalization. What about the medical costs in the case of elective surgery? It would also be interesting to include them in the figure. - Given the differences observed in the literature between the postoperative delirium pattern in cardiac surgery patients with respect to other types of surgery, it would be interesting to know the type of surgery received by patients who have undergone surgery. Would it be possible to categorize the medical costs by type of surgery, for example cardiac vs. non-cardiac surgery?

REVIEWER	Ikaro Barreto Universidade Federal Rural de Pernambuco, Programa de Pos-Graduacao em Biometria e Estatistica Aplicada
REVIEW RETURNED	07-Jul-2022

GENERAL COMMENTS	Statistical Revision: Major revision: The major concern about statistical analysis relies on why the authors did not use random effects considering there are 30 million patients from over 400 acute care hospitals which indicates a clear clustering issue. Minor revisions: 1 – Regarding multicollinearity, it's well known that variance-covariance matrix of LS, GLM, and quasi-likelihood GLMs are not the same. Did the authors consider the use of specific VIF for GLM models? 2 – Maximum Quasi-likelihood estimation (MQLE) is quite powerful in many situations for overdispersed data, data which does not adhere to a member of the exponential family, and misspecification of dispersion matrix. However, MQLE is not a solution to clustering issue and a mixed model should be employed even if MQLE is applied too. 3 – Did the authors consider the use of mixed quantile regression, mixed GAMLSS, or any other model for heavy tail data? 4 – Did the authors considered the use of geometric mean instead of arithmetic mean? 5 – Why authors did not consider the use of multiple comparisons corrections such as FWER methods and FDR methods? 6 – I strongly recommend the use o size effects for a better understanding of how large are the observed differences in univariate results.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Dr. Alba Sanchez, University of Potsdam

Comments to the Author:

The authors have conducted a cross-sectional study on the relevant topic of the incremental medical cost of delirium in hospitalized patients, with a large sample of patients. My suggestions for improvement and questions for the authors are as follows:

Response:

We thank the reviewer for summarizing the key finding. We have now revised the manuscript as suggested by the reviewer.

Comment 1:

As a suggestion, in point 4, I would clarify that the medical costs associated with delirium are those referred to the period of hospitalization, for example: *"The study demonstrates that delirium is associated with significantly higher medical cost during hospitalization ..."* Since other medium/long-term costs (for example those associated with nursing home placement or those derived from the acceleration of cognitive decline) have not been analyzed in this study.

Response:

We thank the reviewer for their comment. We agree that specifying the period of investigation for the medical cost in this study would help readers better understand the study outcomes. We have now corrected point #4 in the Article Summary as follows (Page 3):

"The study demonstrates that delirium is associated with significantly higher medical costs during hospitalization, suggesting that prevention strategies may be critical to reducing the economic burden imposed by delirium."

We have also made this revision in the Methods section of the Abstract as follows (Page 2, paragraph 2):

"Total medical costs during hospitalization were compared between the groups using a generalized linear model."

Comment 2:

Page 6, line 57: The authors state *"The presence of cognitive decline has been defined as diagnosis of dementia, prescription of anti-dementia medication, or presence of a low degree of independence in daily life"*. In the limitations section of the article it is explained that this algorithm is not yet validated, but I would like to know how the low degree of Independence in daily life has been defined. Has it been assessed with the Barthel index? Couldn't the low degree of independence in daily life be due to other reasons such as the presence of neuromuscular/musculoskeletal conditions? Please clarify this point. (Review Checklist '4' and '12')

Response:

We thank the reviewer for their comment. We understand that "low degree of independence in daily life" may not be an accurate description. The database used in this study includes a category of the dementia scale called the "degree of independence in activities of daily living (ADL) related to dementia" (Takigawa Y., *J Public Health Pract* 1994;58:73–5, Sakata N., et al, *J Am Geriatr Soc* 2018;66:871–8), which we have described as "degree of independence in daily life" in the manuscript. "The degree of independence in ADL related to dementia" is an observer-rated scale from rank 0 to rank M, which has high sensitivity and specificity for dementia (Sakata N., et al, *J Am Geriatr Soc* 2018;66:871–8, Meguro K., et al, *Psychogeriatrics* 2012;12:226–34). The ranks are defined as follows: rank 0, individuals with no dementia; rank I, having dementia with no ADL limitation; rank II, having dementia with mild ADL limitations and able to live independently with assistance; rank III, having dementia with moderate ADL limitations and requiring caregiving; rank IV, having dementia with severe ADL limitations and requiring full-time caregiving; rank M, having dementia and requiring medical care due to severe mental, behavioral, or psychological symptoms or physical disabilities. In the present study, a rank of I or greater (I–IV and M) on this scale was defined as one of the criteria for cognitive impairment, based on the article by Sakata et al (*J Am Geriatr Soc* 2018;66:871–8).

We have revised the "Patient selection and characteristics" subsection in the Methods as follows (Page 6, paragraph 5):

"Cognitive impairment was defined as the presence of at least a diagnosis of dementia (ICD-10 codes F00–F03, F067, F107, G238), one prescription of an anti-dementia medication during hospitalization (donepezil, galantamine, memantine, or rivastigmine), or a low rank (I–IV and M) on the Dementia Scale - an observer-rated scale used to assess the degree of independence in activities of daily living (ADL) related to dementia (Table S1).[24]"

We have also revised the description "*Low degree of independence in daily life*" in Table S1 to "*Low degree of independence in activities of daily living related to dementia"* (Supplementary Information, Table S1).

Comment 3:

Page 10, line 43: The authors state *"Patients demographics were comparable between the two groups (Table 1)..."*. Although it is true that the differences between the two groups are not very large, the group with delirium showed more higher percentage of patients with dementia, PIMS, mental and behavioral disorders, emergency hospitalization, ADL dependence... For a better observation of the differences between the two groups, it would be good if you could include effect estimates: when comparing the difference in a continuous variable, please provide the mean or median (if not normally distributed) difference and the 95%CI of the difference; when comparing categorical variables, please also provide an effect estimate (e.g., risk ratio, relative risk) and the 95% CI of the effect estimate. (Review Checklist '7')

Response:

We thank the reviewer for their comment. While we believe that there could be interesting findings when patient demographics are compared, the present study was conducted per a predefined study protocol and statistical analysis plan based on the STROBE statement, which quotes as follows: *"inferential measures such as standard errors and confidence intervals should not be used to describe the variability of characteristics, and significant tests should be avoided in descriptive tables*" (Vandenbroucke JP et al, *Int J Surg* 2014;12:1500–24). Therefore, between-group comparisons for these patient characteristics were not aligned with the objectives of this study.
We have also considered the potential effects of patient background characteristics on the cost comparison between the two groups. The following categories were considered as potential covariates of medical cost and adjusted for in the generalized linear model (GLM) in the present study: age, sex, ADL, Charlson comorbidities, emergency hospitalization, type and duration of anesthesia during surgery, number of PIMs, and ICU admission.

Comment 4:
Page 12: For the duration of hospitalization please provide the median difference and the 95% CI of the difference. (Review Checklist '7')

Response:
We thank the reviewer for their comment. In this study, the duration of hospitalization was not an outcome variable but a prognosis variable. As described previously, the present study was conducted per a predefined study protocol and statistical analysis plan based on the STROBE statement item no. 14(a) (Vandenbroucke JP et al, *Int J Surg* 2014;12:1500–24). Therefore, we believe that computing the median difference and the associated 95% confidence interval (CI) for the duration of hospitalization is not aligned with the objectives of this study.

Comment 5:
Page 12/Fig S1: Figure S1 shows the medical costs categorized by patients characteristics, showing the medical costs for emergency hospitalization. What about the medical costs in the case of elective surgery? It would also be interesting to include them in the figure.

Response:
We thank the reviewer for their comment. While we agree that it would be interesting to evaluate the medical cost associated with elective surgery, Figure S1 was created as supplementary information for the GLM. Therefore, we would like to present the cost breakdown results limited to the covariates used in the GLM. Thus, evaluating medical cost associated with elective surgery was beyond the scope of the research plan/objectives of this study, which were based on a prespecified statistical analysis plan.

Comment 6:

Given the differences observed in the literature between the postoperative delirium pattern in cardiac surgery patients with respect to other types of surgery, it would be interesting to know the type of surgery received by patients who have undergone surgery. Would it be possible to categorize the medical costs by type of surgery, for example cardiac vs. non-cardiac surgery?

Response:
We thank the reviewer for their comment. Although we agree that the type of surgery as a subcategory would be interesting, as described previously, evaluating medical costs by the type of surgery was beyond scope of the research plan of this study, which was based on a prespecified statistical analysis plan. As mentioned in our response to the previous comment, Figure S1 was created as supplementary information for the GLM to present the cost breakdown results limited to the covariates used in the GLM. Therefore, we would like to leave the original Figure S1 as it is, without adding data on the medical costs by the type of surgery. We appreciate your understanding.

Reviewer 2:
Dr. Ikaro Barreto, Universidade Federal Rural de Pernambuco

Comments to the Author:
Statistical Revision

Major Revision

Comment 1:
The major concern about statistical analysis relies on why the authors did not use random effects considering there are 30 million patients from over 400 acute care hospitals which indicates a clear clustering issue.

Response:
We thank the reviewer for their comment. We could not include the hospital identification variable as a random effect in the statistical model because the Medical Data Vision (MDV) database does not provide hospital identification data; these are masked by the database vendor. However, because the variability across sites was included in the variability of error in the model (i.e., we used a larger variability of error than that adjusted by the random effect), the current results are considered adequately conservative. We have now revised the Discussion to include this information as a limitation of the study as follows (Page 16, paragraph 3):
"Additionally, because the MDV database does not provide hospital identification data, we could not include the variability across hospitals as a random effect in the GLM. However, the variability across sites was included in the variability of error in the model (i.e., we used a larger variability of error than that adjusted by the random effect). Therefore, the current results are considered adequately conservative."

Minor Revisions

Comment 1:

Regarding multicollinearity, it's well known that variance-covariance matrix of LS, GLM, and quasi-likelihood GLMs are not the same. Did the authors consider the use of specific VIF for GLM models?

Response:

We thank the reviewer for their comment. In addition to the correlation coefficients between each pair of variables, we estimated the variance inflation factor (VIF) in the multiple linear regression framework as the second-best procedure (SAS output below). There were no variables with a VIF value >10, suggesting the limited effect of multicollinearity.

Variables	**VIF**
Delirium (With or Without)	1.02418
Age category (65-74, 75-84, ≥85)	1.1334
Sex (Male, Female)	1.07876
ADL Score Category (0-59, 60-100)	1.34221
Charlson disease categories (Myocardial infarction Yes, No)	1.03348
Charlson disease categories (Congestive heart failure Yes, No)	1.06552
Charlson disease categories (Peripheral vascular disease Yes, No)	1.01353
Charlson disease categories (Cerebrovascular disease Yes, No)	1.06842
Charlson disease categories (Chronic pulmonary disease Yes, No)	1.02192
Charlson disease categories (Rheumatic disease Yes, No)	1.00812
Charlson disease categories (Peptic ulcer disease Yes, No)	1.00585
Charlson disease categories (Mild liver disease Yes, No)	1.05004
Charlson disease categories (Diabetes without chronic complication Yes, No)	1.01158
Charlson disease categories (Diabetes with chronic complication Yes, No)	1.02285
Charlson disease categories (Hemiplegia or paraplegia Yes, No)	1.02865
Charlson disease categories (Renal disease Yes, No)	1.03404
Charlson disease categories (Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin Yes, No)	1.16284
Charlson disease categories (Moderate or severe liver disease Yes, No)	1.04587
Charlson disease categories (Metastatic solid tumor Yes, No)	1.08807
Emergency hospitalization (Yes, No)	1.85507
Number of PIM drugs (0, 1, 2, 3, >4)	1.08849
ICU admission (immediately after the admission) (Yes, No)	1.05807
Type of surgery (No surgery, Surgery + no/local/light general anesthesia, surgery + general anesthesia [<2 hours], surgery + general anesthesia [≥2 hours])	1.6027

We have now revised the "Statistical analysis" subsection in the Methods to indicate this as follows (Page 8, paragraph 4):
"Multicollinearity was evaluated using pairwise correlation coefficients and variance inflation factors (VIFs) for the multivariable linear regression framework were calculated prior to a quasi-likelihood analysis."

Comment 2:
Maximum Quasi-likelihood estimation (MQLE) is quite powerful in many situations for overdispersed data, data which does not adhere to a member of the exponential family, and misspecification of dispersion matrix. However, MQLE is not a solution to clustering issue and a mixed model should be employed even if MQLE is applied too.

Response:
We thank the reviewer for their comment. As described in the response to Major Revision, Comment 1, we could not include the hospital identification variable as a random effect in the statistical model as this information is not available in the MDV database. However, the variability of site was included in the variability of error in the model (i.e., a variability of error greater than that adjusted by the random effect was used). Therefore, the current results may be considered adequately conservative. We have now indicated this as a limitation in the Discussion section as follows (Page 16, paragraph 3):
"Additionally, because the MDV database does not provide hospital identification data, we could not include the variability across hospitals as a random effect in the GLM. However, the variability across sites was included in the variability of error in the model (i.e., we used a larger variability of error than that adjusted by the random effect). Therefore, the current results are considered adequately conservative."

Comment 3:
Did the authors consider the use of mixed quantile regression, mixed GAMLSS, or any other model for heavy tail data?

Response:
We thank the reviewer for their comment. In this study, our purpose was to investigate whether each covariate could have an impact on the costs, rather than to estimate the cost curve in a precise manner as a functional form or shape. All three methods, including the quasi-likelihood method (QLM) and the two other suggested methods (mixed quantile regression and mixed GAMLSS), can be used to handle cost data, i.e., overdispersed data. The QLM is well accepted especially in the Health Economics and Outcomes Research (HEOR) literature ([Book] de Jong P, Heller GZ, Generalized Linear Models for Insurance Data*, Cambridge University Press*, 2008; Wang HJ, Zhou XH, *Biometrika*, 2010;97:147–58). In addition, the QLM was performed per a prespecified method

according to the statistical analysis plan documented prior to data review. Therefore, any of (mixed) quantile regression, (mixed) GAMLSS, or other methods were not conducted.

Comment 4:
Did the authors considered the use of geometric mean instead of arithmetic mean?

Response:
We thank the reviewer for their comment. We recognize that the terms for each group (i.e., "LS mean") and the group comparison (i.e., "Difference") are not optimal. We have now revised these terms to "Geometric LS mean" and "Geometric LS mean ratio," respectively, as follows:

Statistical analysis (Methods, Page 9, paragraph 1)
"The geometric least squares (LS) mean for total medical cost in each group, the geometric LS mean ratio between the two groups, and its 95% confidence interval were calculated."

Table 3 (Page 14)
"Geometric LS mean [JPY] (SE)"
"Geometric LS mean ratio [JPY]"
"†Geometric LS mean ratio, with delirium/without delirium"

Generalized linear model – exponent calculation (Supplementary Information, Page 1, paragraph 1)
"The geometric least squares (LS) mean, the geometric LS mean ratio, and 95% confidence intervals for the total medical cost in the two groups were calculated."

Comment 5:
Why authors did not consider the use of multiple comparisons corrections such as FWER methods and FDR methods?

Response:
We thank the reviewer for their comment. In the present study, there was only one primary comparison. On the other hand, subgroup analyses involved several comparisons. As subgroup analyses were conducted for exploratory purposes, no adjustment for multiple comparisons was made.

Comment 6:
I strongly recommend the use o size effects for a better understanding of how large are the observed differences in univariate results.

Response:

We thank the reviewer for their comment. We have now performed univariate analysis. The effects of almost all covariates in the multivariable analysis except for emergency hospitalization were generally consistent with those in the univariate analysis. The effect of emergency hospitalization in the univariate analysis (geometric LS mean ratio=1.26; Table S4) may be confounded as the subgroup without emergency hospitalization incurred more expenses than that with emergency hospitalization. This confounding might have been adjusted for in the multivariable analysis (geometric LS mean ratio=0.76). As the LS means were estimated at the combination of average levels of all covariates (e.g., if one codes gender as male=0 and female=1 then the LS mean for gender was estimated as 0.5. This means that we regard the percentages for both male and female as 50% even if the actual percentage of males was not the same as that of females) in the multivariable model, the LS mean in each subgroup could be different in the univariate and multivariable analyses.

We have now added the results of the univariate analysis as Supplementary Table S4 (Supplementary Information, Page 7). We have also briefly mentioned the results of the univariate and multivariable analyses in the Results section (Page 13, paragraph 1) as follows:

"The geometric LS mean ratios of the total medical costs from the univariate analysis were generally similar to those from the multivariable analysis, although only emergency hospitalization was adjusted for in the multivariable analysis (Table S4)."

In addition, we have added a description on univariate analysis to the "Statistical analysis" subsection in the Methods section as follows (Page 8, paragraph 3):

"Univariate analysis was performed with each covariate listed above."

VERSION 2 – REVIEW

REVIEWER	Alba Sanchez University of Potsdam
REVIEW RETURNED	21-Oct-2022

GENERAL COMMENTS	The article has improved with the corrections and the changes introduced. The authors have adequately explained the reasons why they did not want to follow some of the suggestions. I would recommend that a sentence be included in the manuscript explaining that the study was conducted per a predefined study protocol and statistical analysis plan based on the STROBE statement.

REVIEWER	Ikaro Barreto Universidade Federal Rural de Pernambuco, Programa de Pos-Graduacao em Biometria e Estatistica Aplicada
REVIEW RETURNED	20-Oct-2022

GENERAL COMMENTS	Thank you for the response. Most of my suggestions were accepted, and the justification is plausible for those that were not. I consider the manuscript statistically acceptable for publication.