

## 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>	A quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study
<b>AUTHORS</b>	Nojima, Masanori; Tokunaga, Mutsumi; Nagamura, Fumitaka

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Anthony Atkinson London School of Economics, London UK
<b>REVIEW RETURNED</b>	11-Feb-2018

<b>GENERAL COMMENTS</b>	<p>Report for the authors on BMJ open submission bmjopen-2017-021129 \A quantitative investigation of inappropriate use of multivariate analysis and the importance of medical statistics experts in observational medical re-search: a cross-sectional study " by Nojima, Tokunaga and Nagamura</p> <p>I enjoyed reading your paper. I have no methodological comments. Your English is generally excellent, but in a few places it breaks down. I list those I have found below, so that they can be corrected for improved comprehensibility.</p> <p>P.2, l.16. \an inappropriate algorithm". I think it would be clearer to say \Results: The inappropriate algorithm of using only variables ... was estimated to occur ...".</p> <p>l.25. You need to introduces the idea of a nation-level analysis. \A further, nation-level, analysis showed that the involvement ... are associated with a correlation <math>R = -0.0652</math>."</p> <p>l.46. \by creating multiple" ! \by tting several".</p> <p>l.47. \country" ! \nation"? Perhaps better to have country above in l.25?</p> <p>l.50. \confounding factor adjustments".</p> <p>P.3, l.5. \... for variable selection in multivariate ..."</p> <p>l.8. \experts".</p> <p>l.27. \Even papers we classify under the undesirable outcome may not necessarily use an inappropriate form of multivariate analysis".</p> <p>l.30. \Our two outcomes should then be considered as \potentially inappropriate"/\desirable" use ...".</p> <p>l.48. \Consequently ... " I do not understand this sentence. Perhaps it could just be omitted.</p>
-------------------------	---

	<p>P.4, I.22. \(\textit{purposes})".</p> <p>I.28. \(\textit{consultants})".</p> <p>I.52. \(\textit{whether several models were fitted for the same outcome and selected factors})". Omit \(\textit{as an outcome})".</p> <p>P.5, I.50 \(\textit{were})" ! \(\textit{are})". I.55. \(\textit{journals})".</p> <p>P.6, I.14. \(\textit{contracted with})"? \(\textit{available through})" might be better.</p> <p>I.18. \(\textit{to choose})" ! \(\textit{from choosing})". I think this is what you mean.</p> <p>P.7, I.7. It's not clear to me who is validating what.</p> <p>I.9. They are not multiple models from the same outcome / factor relation. The relationship changes due to inclusion of only selected factors.</p> <p>I.19. \(\textit{The following were considered ...})".</p> <p>I.23. \(\textit{analyses})".</p> <p>I.26. \(\textit{in the methods section})".</p> <p>I.43. \(\textit{criterion})".</p> <p>P.8., I.19. \(\textit{estimating equation})".</p> <p>P.9, I.19. \(\textit{basically})" ! \(\textit{for model selection})".</p> <p>I.21. \(\textit{above results, when the authors included an expert, preferable analyses ...})".</p> <p>I.31. \(\textit{independently})".</p> <p>P.10, I.53/ \(\textit{following ideas})".</p> <p>I.54. \(\textit{this has become...})".</p> <p>P.11, I.6. \(\textit{announced})" ! \(\textit{issued})".</p> <p>I.10. rst word. What is \(\textit{it})" here? Should this read \(\textit{inclusion of a factor is totally meaningless unless ... included variables})".</p> <p>I.30. \(\textit{analyses})".</p> <p>I.35. \(\textit{sizes})".</p> <p>I.46. \(\textit{should only be})".</p> <p>I.50. \(\textit{models; however})".</p> <p>P.12, I.32. \(\textit{only a small})".</p> <p>P.13, I. 30. See comment to p.3, I.27</p> <p>P.14, II.17 and 18. \(\textit{critically})" to \(\textit{critical})".</p> <p>P.15, I.22. \(\textit{inversely correlated})". And p.22, caption to Figure 2.</p>
--	--

P.22, l.3 of Figure 2. \expressed as". Incidentally, what is size? Diameter, area?

Referee's general discussion of BMJ open submission bmjopen-2017-021129 \A quantitative investigation of inappropriate use of multivariate analysis and the importance of medical statistics experts in observational medical research: a cross-sectional study " by Nojima, Tokunaga and Nagamura

Anthony C. Atkinson, Department of Statistics, London School

of Economics, London WC2A 2AE, UK. E-mail: a.c.atkinson@lse.ac.uk

The problem is the method used in practice for the selection of variables to be included in logistic regression and Cox models in observational medical studies.

The motivation came from the authors' work as statistical consultants. Many medical researchers had the idea that only variables which were individually significant should be included in the fitted model. This is in contrast to the correct multivariate procedure in which the model should contain variables that are jointly significant. To find these models requires fitting several models and selecting the best, rather than fitting just one. An example is in Table 1 below.

The paper presents the results of a survey in which the frequency of an incorrect method of variable selection was measured as a function of the assessed statistical expertise of the authors of the papers: first author, any other author or none. The expertise was based on the authors' departmental affiliations. It was found that the frequency of correct variable selection increased with the statistical qualifications of the authors. Clinical trials, as opposed to observational studies, were not included.

The authors also consider how the situation might be improved. A break-down of the results by country from papers in which the first author is not an expert shows North America and Northern Europe show relatively high expert involvement compared with East Asia, which have a lower involvement. Taiwan is an exception. In the authors' own country of Japan the education of biostatisticians is developing rapidly. However, it will take time to develop well-trained experts and this is only in one country. The authors suggest that data be made available and analyzed as part of the peer review process. Such suggestions have been made before, for example in power calculations in grant proposals. It would be excellent if some such system could be made to work. Unfortunately, statistical referees are busy and the standard of reviewing of statistical papers seems to be deteriorating not improving.

In their Supplementary Table 3 the authors present an example of a logistic regression analysis which illustrates the incorrect analysis using inference on one factor at a time and the results of the proper analysis. I summarise this analysis in Table 1.

There are three factors

A: adjuvant chemotherapy

L: Lymph node metastasis

B: Biomarker positive.

The three left-hand columns of the table indicate the factors included in the model and the right-hand columns indicate the significant variables. Seven models are fitted. The first three rows show the results of fitting the three one-factor models: L and B are individually significant, but not A. For the three two-factor models, L and B and A and L are both significant, but when A and B are fitted, only B is significant. However, as the last row of the table shows, when all three factors are included, all are significant.

<p>Table 1: Significance of factors when seven different models are fitted to logistic regression analysis of hypothetical data on recurrence of cancer after surgery</p>																																																							
	<table border="1"> <thead> <tr> <th colspan="3">Factors</th> <th colspan="3">Significance</th> </tr> <tr> <th>A</th> <th>L</th> <th>B</th> <th>A</th> <th>L</th> <th>B</th> </tr> </thead> <tbody> <tr> <td>X</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>X</td> <td></td> <td></td> <td>*</td> <td></td> </tr> <tr> <td></td> <td></td> <td>X</td> <td></td> <td></td> <td>*</td> </tr> <tr> <td></td> <td>X</td> <td>X</td> <td></td> <td>*</td> <td>*</td> </tr> <tr> <td>X</td> <td>X</td> <td></td> <td>*</td> <td>*</td> <td></td> </tr> <tr> <td>X</td> <td></td> <td>X</td> <td></td> <td></td> <td>*</td> </tr> <tr> <td>X</td> <td>X</td> <td>X</td> <td>*</td> <td>*</td> <td>*</td> </tr> </tbody> </table>	Factors			Significance			A	L	B	A	L	B	X							X			*				X			*		X	X		*	*	X	X		*	*		X		X			*	X	X	X	*	*	*
Factors			Significance																																																				
A	L	B	A	L	B																																																		
X																																																							
	X			*																																																			
		X			*																																																		
	X	X		*	*																																																		
X	X		*	*																																																			
X		X			*																																																		
X	X	X	*	*	*																																																		
<p>The example shows that all three variables need to be fitted in order to obtain the best model. In other examples only some of the factors may be required. But several models have to be fitted to determine which best describes the data. Such tables can be amplified by using one or more asterisks to express significance levels.</p> <p>In their discussion section the authors also mention problems that arise with few data and several factors. In this case again several models should be fitted, although it will not be possible to fit a full model like that in the last row of Table 1. However a similar table may be helpful in assessing the properties of the various fitted models, perhaps augmented by a measure of model adequacy such as the information criterion AIC.</p>																																																							

<b>REVIEWER</b>	Shinichi Nakagawa University of New South Wales, Australia
<b>REVIEW RETURNED</b>	19-Mar-2018

<b>GENERAL COMMENTS</b>	<p>This study is unique in a way that it attempts to quantify the usefulness of statistical experts in relation to inappropriate statistical use. They discover involvements by a statistical expert reduce inappropriate use of statistics fairly dramatically. Although this is not surprising at all, the study actually provides very numbers on the impact of the statisticians. This is an excellent contribution to the field. Also, the authors are very much aware of their study limitations, which I appreciated. Overall, I felt this study was well conceived and conducted and written up. So I do not have particular comments.</p> <p>One thing is the use of the term “multivariate”. I think that multivariate statistics are more often used to describe a set of statistical methods which have multiple response variables rather than predictable variables (e.g. PCA, DF). So I thought it was about PCA etc when I read the title. I wondered whether the authors want to use something like “..inappropriate regression model</p>
-------------------------	--

	construction.." at least in their title. But I leave this to the editor and the authors.
--	--

## VERSION 1 – AUTHOR RESPONSE

To the reviewers and the editor.

In addition to the comments below, we made some minor changes as follows:

1. "Primary event"/"primary endpoint" was changed to "primary outcome" in section 2.3 and 3.2
2. The header row of Table 2: "event" was changed to "outcome"
3. The title of Table 4
4. The addition of an acknowledgment section

Also, we simplified the strengths and limitations as instructed by the editor. All revisions are marked by red letters.

To the reviewer 1: Dr. Anthony Atkinson

Dear Dr. Atkinson,

Thank you so much for your supportive comments and advanced advice. I was very encouraged by your comments. Also, thank you for organizing the issues raised by us in the general discussion.

Replies to the comments:

Please note that we have not made reply to most comments because we corrected our manuscript as advised (thank you for your detailed review). For other comments, we responded as follows.

- According to the comments for P2.I.46, P4.I.52, and P7.I.9, we changed the description for the third outcome to "Fitting several models for the same outcome and selected factors" thoroughly.
- Since the strengths and limitations needed to be simplified according to the editor instructions, some items have been deleted including P3.I.27-30.
- According to the comment on P7.I.7, we rewrote the sentence more simply.
- According to the comment on P11.I.46, we rewrote the sentence more simply based on the observation because the original sentence was somewhat redundant and a bit different from what we intended to say.

To the reviewer 2: Dr. Shinichi Nakagawa

Dear Dr. Nakagawa,

Thank you so much for your many supportive comments. I was very encouraged by your comments to report this research.

Replies to the comments:

This advice is related to a recent important topic. It is true that there is some controversy about using the term "multivariate analysis" for regression analysis as indicated in the introduction and many authors [\*]. However, this usage is now quite common in medical literature, and we believe that this is not necessarily misuse of the terminology as described in newly added Supplementary Discussion.

Together with the above observations and advice from the reviewer, we changed the title to emphasize academic universality. In addition, although the description in the text was basically left as "multivariate analysis", we enclosed this term at the beginning of the abstract and in the introduction with quotation marks to indicate nonstandard use of the terminology. In accordance with the term "regression model construction" in the new title, we inserted this term in the abstract's study design, in some parts in the introduction and discussion, and at the end of the conclusion.

**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Anthony Atkinson London School of Economics, London UK
<b>REVIEW RETURNED</b>	06-Apr-2018

<b>GENERAL COMMENTS</b>	<p>Report for the authors on BMJ open submission bmjopen-2017-021129.R1 \A quantitative investigation of inap-propriate model construction and the importance of medical statistics experts in observational medical re-search: a cross-sectional study " by Nojima, Tokunaga and Nagamura</p> <p>Thank you for the careful revision of your paper. I have a few comments on minor details in your revision. Unfortunately what I have been sent is a Word document, without line or page numbers and in two versions - what is on the screen does not show your corrections, unlike the printed version. I hope we can manage. Everything I mention is in the passages marked with a vertical line.</p> <p>p.2, Strengths. \This is unique research quantitatively investigating ...".</p> <p>p.3, Limitations, l.2. \complicated de nitions".</p> <p>Last para., l.1. \of the previous"</p> <p>p.4, last l. of 3rd complete paragraph. \for the same outcome and selected factors". Replace with \for the same outcome and sets of selected factors (\selected factors)". It is important that it is clear that several sets were considered. But you can continue to call this procedure \selected fac-tors".</p> <p>p.7, x2.4, l.2. \estimating equations".</p> <p>p.10, rst bullet point at bottom of p. \abuse of p values from the Amer-ican Statistical Association (ASA) was issued [7] in 2016.</p> <p>p.11, 2/3 down. \However, this step tended".</p> <p>p.13. x4.2, end. \See the Supplementary".</p> <p>p14, Acknowledgments. \We", not \I"!</p>
-------------------------	--

**VERSION 2 – AUTHOR RESPONSE**

To the reviewer 1: Dr. Anthony Atkinson

Dear Dr. Atkinson,

We really appreciate your detailed review again. Your comments has greatly improved our manuscript with high expertise.

Replies to the comments:

p.2, Strengths. "This is unique research quantitatively investigating ...".  
p.3, Limitations, l.2. "complicated definitions". Last para., l.1. "of the previous"  
-> We revised our manuscript as you advised.

p.4, last l. of 3rd complete paragraph. "for the same outcome and selected factors". Replace with "for the same outcome and sets of selected factors ("selected factors")". It is important that it is clear that several sets were considered. But you can continue to call this procedure "selected factors".  
-> We revised our manuscript as you advised. This comment is very helpful to describe this outcome properly. Thank you so much.

p.7, §2.4, l.2. "estimating equations".  
p.10, first bullet point at bottom of p. "abuse of p values from the American Statistical Association (ASA) was issued [7] in 2016."  
p.11, 2/3 down. "However, this step tended".  
p.13. §4.2, end. "See the Supplementary".  
-> We revised our manuscript as you advised.

p14, Acknowledgments. "We", not "I!"  
-> We revised it as you advised. (It is an embarrassing mistake...!)